FROM SURFACE EVENTS IN DEAP-3600 TO THE RADIAL ACCELERATION RELATION IN SPIRAL GALAXIES

by

Connor Stone

A thesis submitted to the
Department of Physics, Engineering Physics, & Astronomy
in conformity with the requirements for
the degree of Master of Science

Queen's University
Kingston, Ontario, Canada
August 2018

Abstract

One of the most outstanding mysteries in the Universe is the nature of dark matter. In this thesis, two unique projects that focus on understanding dark matter are presented. First, Machine Learning is used to detect background signals in the DEAP-3600 direct detection dark matter experiment. Machine Learning algorithms are found to achieve a near order of magnitude improvement on discrimination of surface backgrounds ove a standard position reconstruction algorithm. To ward against over-fitting, the algorithms are fit to real data before being tested on simulated data. Validation is done by modifying the detector simulation and recording the change in performance. It is found that the Machine Learning algorithms are robust to small changes in the optics of the detector and so would be stable for use on the real data. The second project is a study of the intrinsic scatter of a scaling relation and its potential tension with the dark matter hypothesis. The Radial Acceleration Relation (RAR) relates the observed radial acceleration in a galaxy to that expected from a purely baryonic mass distribution. Lelli et al. (2017) have suggested that the RAR may have zero intrinsic scatter. If true, this would be inconsistent with the Λ Cold Dark Matter (ΛCDM) paradigm, but a natural prediction of Modified Newtonian Dynamics (MOND). A large sample of galaxies is collated from six independent samples, representing over an order of magnitude increase in the amount of data available

to examine this scaling relation. To test the zero scatter hypothesis, a Monte-Carlo simulation with a zero intrinsic scatter assumption is constructed; observational uncertainties are then re-introduced to the data. This technique shows observational uncertainties are too small to account for the observed scatter in the relation. It is found that the RAR is consistent with standard ΛCDM models and that it has an intrinsic scatter of order 0.13 dex ($\sim 50\,\%$ of observed). Further tests show that these results cannot be explained with non-circular motions or correlations with other galaxy parameters.

Acknowledgements

From the bottom of Creighton mine to the top of Maunakea I would like to thank my supervisors, Tony Noble and Stéphane Courteau for their support and guidance over the past two (+) years. Tony, your thoughtful insights led me to think more deeply at every turn. Stéphane, your endless enthusiasm for science and learning is inspiring, I would never have learned so much without you.

I would like to thank the DEAP-3600 collaboration, among whom I count every one as a friend. I would especially like to thank Chris Jillings for bringing me on to the collaboration so many years ago when my career was just getting started. Special thanks to Tina Pollmann, who was way too busy to help me, but always did anyway.

Nikhil Arora, Joe McLaughlin, Daniel Durnford, Christine Hall, and Ryan Groome have all made this experience truly exceptional in their own way.

Queen's University has provided me with many opportunities for which I am grateful. I would like to especially thank Loanne Meldrum, Peggy Hauschild, Kyra Funk, Tammie Kerr, and Melissa Balson for all your help.

Thank you to my family: Mom, Dad, and Bryan (+ many more) whose belief in me has always been unshakable. To Jessica Smith, in all cases you have been the support that I needed, either a push to strive for better, or a source of comfort and direction. To all of the above and so many others thank you, the future looks bright.

Contents

Abstract i

Acknowledgements iii

Contents iv

List of Figures viii

Glossary x

Chapter 1: Introduction 1
1.1 Motivation . . . 1
1.2 Dark Matter . . . 4
1.2.1 Observational Signatures . . . 4
1.2.2 Massive Astrophysical Compact Halo Objects . . . 6
1.2.3 Modified Newtonian Dynamics . . . 6
1.2.4 Particle Dark Matter . . . 9
1.2.5 Dark Matter Detection . . . 10
1.2.6 The ΛCDM Paradigm . . . 12

I MACHINE LEARNING TO BOOST SURFACE EVENT DISCRIMINATION IN THE DEAP-3600 DETECTOR 14

Chapter 2: Outline 15

Chapter 3: Detector Description and Analysis Overview 17
3.1 DEAP-3600 . . . 17
3.1.1 Design . . . 17
3.1.2 Stages of an Event . . . 19
3.1.3 Background Management . . . 21
3.1.4 Fiducial Volume by Position Reconstruction . . . 23

3.2 Machine Learning . . . 27
3.2.1 Classification . . . 27
3.2.2 Preprocessing . . . 28
3.2.3 Hyper-Parameters . . . 30
3.2.4 Logistic Regression (LR) . . . 31
3.2.5 Naive Bayes (NB) . . . 32
3.2.6 Support Vector Machine (SVM) . . . 33
3.2.7 Neural Network (NN) . . . 34
3.2.8 K-Nearest Neighbours (KNN) . . . 36

Chapter 4: Training with Simulated Data 37
4.1 Procedure . . . 37
4.2 Simulated ^{40}Ar Events . . . 38
4.3 Region of Interest (ROI) . . . 40
4.4 MBLikelihood on the Simulated Data . . . 42
4.5 Training . . . 43
4.6 Individual Performance . . . 46
4.7 Boosted Position Reconstruction . . . 48

Chapter 5: Trained on Real Data 50
5.1 ^{39}Ar Beta Decay Events . . . 50
5.2 ^{39}Ar Event Selection . . . 51
5.3 Training with Real Data . . . 52
5.4 Individual Performance . . . 53
5.5 Boosted Performance . . . 55

Chapter 6: Tests with Alternate Datasets 57
6.1 Modified Optical Model . . . 57
6.1.1 TPB Thickness . . . 58
6.1.2 Fill Level . . . 59
6.1.3 Scattering . . . 60
6.1.4 Absorption Fraction . . . 61
6.2 Americium Beryllium Neutron Events . . . 64
6.3 ^{210}Po Alpha Events . . . 65

Chapter 7: Further Characterization 69
7.1 Energy Dependence . . . 70
7.2 Differences in Simulated and Real Data . . . 70
7.3 Fiducial Volume Size . . . 73
7.4 Multi-Boosting . . . 74
7.5 Runtime . . . 76

Chapter 8: Summary and Conclusions **79**
8.1 Summary 79
8.2 Future Work 80
8.3 Conclusion 82

II AN EXTENSIVE STUDY OF THE SCATTER IN THE RADIAL ACCELERATION RELATION IN SPIRAL GALAXIES **84**

Chapter 9: Outline **85**

Chapter 10: Background **87**
10.1 Scaling Relations 87
10.2 Galaxy Measurements 89
10.2.1 Surface Brightness Profiles 89
10.2.2 Rotation Curves 90
10.2.3 Size Measurements 92
10.3 Galaxy Structural Parameters 93
10.3.1 Inclination 93
10.3.2 Colours and Mass to Light Ratios 94
10.3.3 Distances 97
10.4 The Radial Acceleration Relation (RAR) 98
10.4.1 What is the RAR? 98
10.4.2 Calculating the Observed Acceleration (g_{obs}) 100
10.4.3 Calculating the Baryonic Acceleration (g_{bar}) 100
10.4.4 The RAR as an Alternative to the Baryonic Tully-Fisher Relation 102
10.5 Statistical Techniques 103
10.5.1 Bootstrap Sampling 103
10.5.2 Regression Analysis 105
10.5.3 Monte-Carlo Sampling 107

Chapter 11: Datasets **109**
11.1 Mathewson 1992 (M92) 109
11.2 Mathewson 1996 (M96) 110
11.3 Courteau 1997 (C97) 111
11.4 ShellFlow (SF) 111
11.5 SPARC 112
11.6 SHIVir 113
11.7 Data Quality Selections 113

Chapter 12: The Scatter in the RAR **116**
12.1 The Scatter in the Observed RAR 116
12.1.1 Fitting the RAR 116
12.1.2 Testing for Dependencies of the Residuals 122
12.2 A Simulated RAR with Scatter 123
12.2.1 A Simulated Galaxy on the RAR 123
12.2.2 Simulated RAR Under Zero Scatter Assumption 126

Chapter 13: Summary and Conclusions **134**
13.1 Summary 134
13.2 Future Work 135
13.3 Conclusions 137

Bibliography **160**

Appendices **161**

Appendix A: Comparison of Dataset Measured Quantities **161**

Appendix B: Scaling Relations for Each Dataset **171**

Appendix C: The Residuals in the RAR Against Individual Galaxy Parameters **178**

List of Figures

3.1 DEAP-3600 design representation 19

4.1 Simulated data reconstructed radius 43
4.2 Charge vector normalization 45
4.3 Individual leakage when trained with simulation data 47
4.4 Boosted leakage when trained with simulation data 49

5.1 Training sample taken from real data using Centroid 53
5.2 Individual leakage when trained with real ^{39}Ar data 54
5.3 Boosted leakage when trained with real ^{39}Ar data 56

6.1 Surface event leakage as a function of TPB thickness in the optical model 59
6.2 Surface event leakage as a function of fill level in the detector 60
6.3 Surface event leakage as a function of scattering length in the optical model 62
6.4 Surface event leakage as a function of absorption length in the optical model 63
6.5 Acceptance of real neutron events using different algorithms 65

7.1 Energy dependence of surface event leakage 71
7.2 Comparison of output parameters for real and simulated data 73
7.3 Fiducial Volume dependence of chosen surface event leakage 75
7.4 Output parameter comparison for each algorithm 77
7.5 Runtime for each algorithm 78

10.1 Demonstration of several linear regression algorithms 106

12.1 The observed Radial Acceleration Relation 120
12.2 Observed RAR fit residuals histogram 121
12.3 Example simulated galaxies on the RAR 126
12.4 Simulated Radial Acceleration Relation 129
12.5 Simulated RAR fit residuals histogram 130

A.1 Histograms of the morphological type distribution for each dataset. . 162

A.2 Histograms of the luminosity distribution for each dataset. 163
A.3 Histograms of the V_e distribution for each dataset. 164
A.4 Histograms of the central surface brightness distribution for each dataset. 165
A.5 Histograms of the dynamical mass distribution for each dataset. . . . 166
A.6 Histograms of the stellar mass distribution for each dataset. 167
A.7 Histograms of the effective radius distribution for each dataset. 168
A.8 Histograms of the $R_{23.5}$ distribution for each dataset. 169
A.9 Location of all studied galaxies on the sky 170

B.1 Plot matrix for SV data 172
B.2 Plot matrix for SP data 173
B.3 Plot matrix for SF data 174
B.4 Plot matrix for C97 data 175
B.5 Plot matrix for M92 data 176
B.6 Plot matrix for M96 data 177

C.1 RAR Residuals versus Dynamical Mass 179
C.2 RAR Residuals versus Rotational Velocity 180
C.3 RAR Residuals versus Inclination 181
C.4 RAR Residuals versus Morphological Type 182
C.5 RAR Residuals versus Luminosity 183
C.6 RAR Residuals versus Effective Radius 184
C.7 RAR Residuals versus Isophotal Radius 185
C.8 RAR Residuals versus Distance 186
C.9 RAR Residuals versus Central Surface Brightness 187

Glossary

Symbols | A | B | C | D | E | F | H | I | K | L | M | N | P | R | S | W

Symbols

1,1,4,4-tetraphenyl-1,3-butadiene (TPB) Organic wavelength shifter applied to the inner surface of the DEAP-3600 Acrylic Vessel (AV).

A

Active Volume The material in a detector that is sensitive to interactions with the target (dark matter) particle. In DEAP-3600 the active volume is the liquid argon volume.

AV Acrylic Vessel.

Axis Ratio In an ellipse the axis ratio is the ratio b/a of the semi-minor (b) to the semi-major (a) axis.

B

BTFR Baryonic Tully-Fisher Relation.

C

Centroid Algorithm for performing position reconstruction by taking a weighted mean position of the Photo-Multiplier Tube (PMT) charges in DEAP-3600.

CMB Cosmic Microwave Background.

Curve of Growth The integrated total enclosed magnitude as a function of projected radius in an extended astrophysical object. See magnitude.

D

DEAP-3600 Dark matter Experiment using Argon Pulse shape discrimination, with 3600 kg of liquid argon.

Decision Boundary Divides an N dimensional space into two sections using an $N-1$ dimensional hypersurface that can be shaped arbitrarily as long as it divides the space. Elements on one side of the decision boundary are all given the same classification, a different classification for events on the other side.

Dex A unit in $\log_{10}$ space, for example 1 dex indicates a factor of 10.

E

Effective Radius The projected major axis radius for an isophote enclosing half of the total light in a galaxy.

F

FB Filler Block.

Fiducial Volume Argon sub-volume within which events are accepted for Weakly Interacting Massive Particle (WIMP) search. The remaining argon volume is used as shielding from background events.

Fprompt Fraction of early light (150 ns) in a waveform (15 µs).

H

Hα The first Balmer emission line in hydrogen with a wavelength of approximately 656 nm. It is a useful probe for rotation curves in spiral galaxies as it is mostly emitted in star forming regions which occur at or near the plane of the disk.

HI A neutral hydrogen emission line associated with a spin flip of the electron (from aligned to anti-aligned with the proton). With a wavelength of approximately 21 cm it is a useful probe in galactic astronomy as it mostly samples the plane of the disk (in spiral galaxies).

HII Indicates ionized hydrogen, which generally occurs in star forming regions of galaxies.

Hyper-Parameter Parameter that controls some aspect of training or evaluation of a machine learning algorithm.

Hyperplane In N dimensional space, a linear boundary of $N-1$ dimensions. Defined by a vector normal to the surface and an intercept value.

I

Initial Mass Function Probability distribution function of masses for stars formed in a typical collapsing cloud. Often assumed to be universal, and thus all stars are drawn from the same distribution.

Isophotal $\mathbf{R}_{23.5}$ The projected major axis radius for the isophote with a surface brightness of 23.5 mag arcsec^{-2}.

Isophote A contour of constant brightness, almost always required to take the shape of an ellipse.

K

KNN K-Nearest Neighbours.

L

Λ Cold Dark Matter (ΛCDM) Standard dark matter paradigm in which there exists particle dark matter that is kinematically cold.

Leakage Fraction of events in a data set that are accepted by a cut. Usually only used when the events being accepted are (undesirable) background events, in the case of WIMP leakage into the Region Of Interest (ROI) it is generally called acceptance.

LG Light Guide.

LR Logistic Regression.

M

Machine Learning An approach to data analysis that focuses on inference problems using algorithmic predictions based on previous samples. In general, the goal is to learn a mapping from an input vector X to an output vector Y using a finite number of samples X for which Y are known.

MACHO Massive Astrophysical Compact Halo Object.

Magnitude A measure of light intensity in logarithmic units, from a specified wavelength range. For a source with intensity I_s (Wm^{-2}) it will have a magnitude of $m_s = m_{ref} - 2.5\log_{10}(I_s/I_{ref})$ compared to a well measured reference source m_{ref}. An apparent magnitude is the light intensity as observed, an absolute magnitude is corrected to a distance of 10 pc using $M_{abs} = m_{app} - 5\log_{10}(d/10)$.

Malmquist Bias The bias causing a flux limited sample to include more intrinsically bright objects than dim ones. This is due to the fact that intrinsically bright objects can be seen out to a greater distance, and that it is more likely for them to randomly scatter into the sample than out of it.

Mass-to-Light Ratio Conversion factor between the integrated light in a specified photometric band and total, baryonic, or stellar mass generally represented as Υ (see Section 10.3.2).

MBLikelihood Algorithm for performing position reconstruction using a likelihood maximization routine on the PMT charges in DEAP-3600.

MOND Modified Newtonian Dynamics.

Morphological Type Morphological classification of a galaxy indicating the strength of features such as spiral arms and central bulges. The numerical values represent the T-type with negative meaning elliptical, positive is spiral, and ≥ 10 implying dwarf or irregular morphologies.

N

NB Naive Bayes.

NED NASA/IPAC Extragalactic Database.

NN Neural Network.

Number of Photo-electrons Estimated number of photo-electrons registered by a PMT (or the sum for all PMTs) for a single event..

P

PDF Probability Density Function.

PMT Photo-Multiplier Tube.

Preprocessing Transformations made to data before it is used as input to a machine learning algorithm.

R

Radial Acceleration Relation (RAR) A scaling law which pairs the observed radial acceleration in a galaxy (V^2/R), with that which would be expected from baryonic matter alone (GM_*/R^2).

Reactor Analysis Tool (RAT) Code framework in which both simulation and analysis for DEAP-3600 is conducted.

Region Of Interest (ROI) Parameter space in which events are considered candidate WIMP events. May also refer to the values of a single parameter in which events are considered candidate WIMP events.

S

Scaling Relation A relationship (generally linear) between two parameters that describe a galaxy. Often represented in log space as they span several orders of magnitude, scaling relations are common tools for understanding the structure and behaviour of galaxies.

SDSS Sloan Digital Sky Survey.

Simple Stellar Population A model for a single population of stars generated simultaneously and evolved over time.

Spectral Energy Distribution (SED) A function giving the intensity as a function of wavelength (or frequency) for the light emitted by an object.

SPS Stellar Population Synthesis.

Star Formation History A description of the rate at which stars were formed as a function of time in the history of a galaxy. May also account for stars acquired by merging with other galaxies.

Surface Brightness The flux per unit of angular area in arcseconds^2. For a given magnitude m measured within and angular area A the surface brightness is $\mu = m + 2.5\log_{10}(A)$ and is indicated by the units mag·arcsec^{-2} (μ) or in units $L_\odot pc^{-2}$ (I).

Surface Brightness Profile The surface brightness measured as a function of projected radius in an extended astrophysical object (see surface brightness).

SVM Support Vector Machine.

W

WIMP Weakly Interacting Massive Particle.

Chapter 1

Introduction

1.1 Motivation

One of the most pressing goals of modern astrophysics and astroparticle physics is the identification and characterization of dark matter. Many phenomena indicate the presence of an as yet unexplained source of gravitational force at large scales. Galaxies, Galaxy Clusters, Cosmic Microwave Background (CMB) fluctuations, and the structure of the universe itself can only be understood in a context with dark matter (Freeman 2017). Despite its ubiquity in the universe, dark matter has eluded more direct study. Attempts to produce or directly detect dark matter have thus far been unsuccessful (CMS Collaboration 2017; Aprile et al. 2018). With gravity as the only method of observing it, even the rough distribution of dark matter within a galaxy remains challenging to determine (Bovy and Tremaine 2012). This thesis presents two distinct studies on dark matter projects, with a focus on using techniques which are generally applicable in their area of analysis. In the first project, Machine Learning is used to tackle a common problem in direct detection dark matter projects. The second project is heavily dependent on accurate measures of uncertainty, and so the

analysis focuses on general principles for computing reliable measures of uncertainty through multiple levels of calculation.

A common type of background that direct detection dark matter experiments must contend with is surface contamination, in Part I, a powerful technique for addressing this is explored. The active volume of the detector will always have some boundary that interfaces with the rest of the detector and it is common for small amounts of radioactive materials to get deposited at this boundary during construction (Cai et al. 2011; Jillings and DEAP Collaboration 2016; Lehnert 2018). To eliminate false dark matter signals, a fiducial volume is selected out of the active volume of the detector using position information from the event signal to eliminate any signal that is too close to the active volume boundary (thus central parts are kept) (Pollmann 2012). The process of selecting events from the fiducial volume is often complex and requires a unique position determination algorithm for each detector (Olsen 2009; Akerib et al. 2018). The performance of the algorithm can have a large impact on the sensitivity of an experiment and so great effort is put into them. In this thesis the use of Machine Learning is explored for the purpose of improving fiducial volume calculations in the specific case of the DEAP-3600 dark matter search experiment. Similar steps can easily be adopted in other experiments as the Machine Learning algorithms are generally applicable. By employing general purpose algorithms, analysis teams can save time in development and devote more resources to validation of the algorithm; as is done in this thesis. This work is timely: as the level of sophistication of dark matter search experiments increases, it becomes advantageous to exploit more subtle aspects of the detector, rather than making simplifying assumptions. With conventional methods, tailoring an algorithm to a specific experiment can be tedious and time

consuming, in Machine Learning it is done automatically.

Scaling relations are one of the primary methods of studying galaxy formation and evolution. The parameters that describe a galaxy can range over several orders of magnitude. Size, Mass, Rotational Velocity, Luminosity, Colour, Concentration, and Scale Height are just a few of many parameters that are used to describe a galaxy (Bland-Hawthorn and Gerhard 2016). These can be paired into scaling relations, such as a match between luminosity and the rotational velocity that is commonly used as a distance indicator (Tully and Fisher 1977; Mathewson and Ford 1996; Courteau 1997). A recently proposed scaling relation is the Radial Acceleration Relation (RAR) which compares the radial acceleration between that from observations and what would be expected from visible matter alone (Lelli et al. 2017). When presented, it was claimed that this relation had too little scatter to be described by the standard Λ Cold Dark Matter (ΛCDM) paradigm. In this thesis, the Lelli et al. (2017) claim is examined using more advanced uncertainty calculations and over and order of magnitude more data. By more carefully propagating uncertainty from the original measurements to the final calculated quantities, it is found that some scatter cannot be explained by observational uncertainties alone. Thus it is found that with a more careful analysis, ΛCDM is consistent with the observations, and the intrinsic scatter is found to be on the order of 0.13 dex.

Both of these projects are designed to lead towards a common goal of an exciting PhD program. Recent advances in the fields of Machine Learning and data mining are now enabling applications to sub-fields of physical science that rely on complex data analysis. As we enter an era where large datasets exceed a size that can be examined manually (Ivezić et al. 2012; Abolfathi et al. 2018; Gaia Collaboration et al. 2018),

one can search for subtle interconnections and high dimensional patterns by taking advantage of these new Machine Learning techniques. The upcoming PhD project will focus on discovering these subtle patterns in galaxy structure and dynamics, taking advantage of the large amount of data now available and preparing for future datasets from large surveys (Ivezić et al. 2008; Racca et al. 2016). Part I is clearly well suited to develop experience with Machine Learning algorithms. These will be central to the PhD analysis as the primary tools for discovering patterns in high dimensional data. Part II has two major aspects which are relevant. To test the RAR, several datasets have been compiled, these data will be the focus of analysis for the PhD. Also, the tests performed in Part II primarily focused on uncertainty in the measurements. The techniques used to make these uncertainty measurements are generalized enough that they can be applied to Machine Learning algorithms as well. While Machine Learning has been applied to many problems in physics, it is still in its infancy and often uncertainties are omitted as they are challenging to compute for such complex algorithms. The ability to compute meaningful uncertainties for these algorithms will be a major component of the future PhD work. While the techniques in both Part I and Part II are not novel, their application to the search for new understanding of galaxy structures and dynamics will be.

1.2 Dark Matter

1.2.1 Observational Signatures

In the 1930s, Fritz Zwicky used the virial theorem to demonstrate that galaxy clusters had more mass than was expected from photometric observations (Zwicky 1933; van den Bergh 1999). By observing the light from each galaxy in a cluster, it is

possible to predict how much mass is in the cluster. This is because gas and stars produce light in predictable distributions and so a mapping can be created between light and mass (Conroy 2013). It is also possible to measure the line of sight velocity of a galaxy from the Doppler shift of its light. The virial theorem can then determine the amount of gravitational mass needed to bind objects moving at those speeds (Heisler et al. 1985; Binney and Tremaine 2008). Since Zwicky's original observations, other measurements have demonstrated a discrepancy between gravitational mass and visible mass. In the 1970s, Vera Rubin observed the rotational velocity of the Andromeda galaxy and saw that the rotation curve did not decrease as $R^{-1/2}$ as expected, but instead remained flat (Rubin and Ford 1970). A rotation curve is a measure of the average rotational velocity as a function of radius in the disk of a galaxy, explained further in Section 10.2.2. Using the luminosity profile of a galaxy, it is possible to determine an estimate of the surface mass density and, consequently, the rotation curve of that galaxy (Toomre 1963; Binney and Tremaine 2008). While Vera Rubin used Hα measurements of HII regions, the results were later confirmed with observations at radio wavelengths of the neutral hydrogen 21 cm emission line. Just like with Zwicky's galaxy clusters, the visible mass did not match with the Newtonian mechanics gravitational mass predictions unless large mass-to-light ratios were invoked (Bosma and van der Kruit 1979). Many other observations have been made that show a discrepancy between visible mass and gravitational mass (Arnaud 2005; Planck Collaboration et al. 2016a; Freeman 2017) leading to the prediction of some form of extra gravitation that does not interact with light (Faber and Gallagher 1979). There are many proposed models that could describe this phenomenon, in broad terms they fall into three categories described in Sections 1.2.2 - 1.2.4.

1.2.2 Massive Astrophysical Compact Halo Objects

It was possible that some regular matter had simply been missed. The most popular version of this is called a Massive Astrophysical Compact Halo Object (MACHO), and they were postulated to be asteroids, dwarf planets, and black holes among others. Efforts to find and characterize MACHOs determined that they could not account for the mass discrepancy (Patrignani 2016). Using microlensing, it would be possible to detect flickering in luminous objects as they are obstructed by a MACHO (Paczynski 1986). Gravitational microlensing would temporarily brighten objects in a characteristic signature that has thus far not been observed to the degree necessary to account for dark matter. There is still parameter space available for black hole MACHOs and searches continue (Chapline and Frampton 2016; Wang et al. 2018; Carr and Silk 2018). However, most focus has moved to other possibilities due to the unfavourable experimental results so far and lack of clear motivation. Models of the early universe indicate a large non-baryonic component to the universe which for a MACHO could only be black holes.

1.2.3 Modified Newtonian Dynamics

A second possibility to describe the discrepancy was that the laws of physics are simply different at the large scales where these measurements are made. Just as General Relativity replaced Newton's Law of Universal Gravition for a larger scale of mass, distance, and time, it is possible that at the galactic scale a new physical understanding is needed. Modified Newtonian Dynamics (MOND), first proposed by Milgrom (1983) is an alteration to Newtonian dynamics that is able to recover

rotation curves as observed in galaxies. It does this by introducing a somewhat ad-hoc modification of Newton's second law at low acceleration in roughly the regime where galaxy rotation curves depart from predictions. Alternatively, one can modify Newton's Law of Universal Gravity directly. This modification takes the traditional Newtonian $\vec{F} = m\vec{a}$ and adds a term to give: $\vec{F} = m\mu(a/a_0)\vec{a}$. The function $\mu(x)$ has the properties $\mu(x \gg 1) \approx 1$ and $\mu(x \ll 1) \approx x$; the scale constant would have a magnitude on the order of $a_0 \approx 10^{-10}\,\mathrm{ms}^{-2}$. The modification at Newtonian scales is assumed to be recoverable from a modification of Relativity, there are several directions this could take but as yet it is not clear how this would be done (Milgrom 2015).

An immediate consequence of MOND is that galaxy rotation curves should be flat after a certain radius. As demonstrated in (Milgrom 1983) (and reproduced here for continuity) at large radii one can assume mostly circular orbits and a constant enclosed mass. This means the standard gravitational acceleration is $g = GMr^{-2}$ and the standard acceleration is $a = V^2 r^{-1}$. At large r the acceleration is low and so the modified acceleration becomes $a \to a^2 a_0^{-1}$. The speed of a test particle is then $V^4 = MGa_0$, which has no radial dependence. This consequence was the inspiration for the form of MOND and it thus not a prediction, but a fit to the data. It also follows from this, and an assumption of minimal variation in mass-to-light ratio, that a relationship between luminosity and rotational velocity should be observed. This is indeed the case, known as the Tully-Fisher relation it is one of the most useful scaling relations in galaxy physics (Tully and Fisher 1977).

Another consequence, which will be explored more deeply in Part II is the strength of the link between gravitational acceleration and baryonic matter. In MOND there

is no dark matter (or very little) and so all of the gravitational force is generated by visible matter, albeit through a non-Newtonian relation. This implies there must be a perfect relation between the two, whereas dark matter explanations include hidden matter which can undergo its own dynamics and be distributed in a non-trivial way. While this would seem a clear avenue for distinguishing MOND from other models, it has traditionally been very challenging to accurately measure the baryonic mass in a galaxy. That is still the case, but newer studies (Roediger and Courteau 2015; Zhang et al. 2017) provide careful analyses of uncertainties in these mass measurements, coupled with large datasets it is possible to measure these variations. Indeed, the impetus for Part II are exactly these kinds of measurements.

Rotation curves are not the only evidence of dark matter. Gravitational lensing has been observed for many systems to determine their mass distribution and has revealed special cases such as the Bullet Cluster (Clowe et al. 2006). This is actually two clusters of galaxies that are in the early stages of a collision, a process that will eventually merge them into one larger cluster. Using gravitational lensing it is observed that the majority of the mass passed through the collision unobstructed, leaving a large cloud of gas between the two clusters that can be observed using x-ray band telescopes (Clowe et al. 2004). Gas from the two clusters interacted very strongly and so was unable to pass through the collision along with the rest of the mass. A significant amount of the baryonic mass in a cluster is known to be in the form of gas (Mathews 1978; Kravtsov and Borgani 2012), so it is very challenging to explain in a MOND framework while simultaneously matching other dark matter signals such as rotation curves. In modern astronomy and cosmology, one of the strongest pieces of evidence for particle dark matter (and thus against MOND) is

the Baryon Acoustic Oscillations (Bond and Efstathiou 1984). By observing the CMB it is possible to observe characteristic length scales of density waves travelling through the plasma of the primordial universe (Planck Collaboration et al. 2016a). It is challenging for MOND to model these length scales in a satisfactory way; the CMB seems to indicate the existence of a form of matter that interacts graviationally, but not electromagnetically. However, MOND is not entirely inconsistent with the CMB, and so there are some who continue to investigate it (Sanders and McGaugh 2002; Kiselev and Timofeev 2012; McGaugh 2016; Lelli et al. 2017).

1.2.4 Particle Dark Matter

The third possibility, then, is a new particle beyond the standard model, a particle with mass and a small cross section with the other forces. One initial possibility was that the dark matter could be in the form of neutrinos. Because of their high speed, neutrinos could not clump together at the scales observed in our universe to create the galaxies and other large scale structures (White et al. 1983). Along with many others, a leading candidate for particle dark matter is the Weakly Interacting Massive Particle (WIMP) (Feng 2010). The reason it is called "weakly interacting" is that the hypothesized particle has an interaction strength with standard model particles that is similar to that of the weak nuclear force (Patrignani 2016). After exhausting other possibilities it was realized that a WIMP would have to be a new particle outside of the standard model, such as (but not limited to) the ones predicted by Super Symmetric models (Pagels and Primack 1982; Patrignani 2016).

Supersymmetry represents a family of possible extensions to the standard model that all introduce a new symmetry between fermions and bosons (Cooper et al. 1995).

A compelling part of supersymmetry is that it was developed to solve a number of other problems in physics such as the hierarchy problem. Only afterwards was it discovered that supersymmetry naturally produces WIMPs without tuning (Patrignani 2016). This is called the *WIMP Miracle* as the hypothesis naturally generates WIMPs with roughly the correct primordial abundances when weak scale interaction strengths are assumed. Current experiments have begun to probe the parameter space of possible supersymmetric dark matter particles without success (Aprile et al. 2018).

Another leading candidate today is the Axion (Peccei 2008; Sikivie 2010). Similar to Supersymmetry, an Axion comes from the introduction and breaking of a new symmetry in the standard model (Weinberg 1978). These particles would be extremely light, on the order of $10^{-5}\,\mathrm{eV/c^2}$. This means that they would have unusual properties, for example forming Bose-Einstein condensates in galaxies (Sikivie and Yang 2009). Thus far, experiments have not detected this potential form of dark matter (Asztalos et al. 2010).

1.2.5 Dark Matter Detection

WIMP search experiments can be placed in three categories: direct detection, indirect detection, and production. Direct detection experiments focus on observing an interaction between dark matter and standard model particles in a controlled environment. Indirect detection experiments look for signals of dark matter annihilations, such as gamma ray emission lines predicted by some dark matter models. In accelerator experiments, the hope is to produce dark matter (which would immediately

exit the system undetected) and deduce the existence of a particle from missing momentum that it would carry away from the interaction. The Large Hadron Collider is an accelerator experiment that has begun to exclude some of the simplest supersymmetric models (Baer et al. 2013; Patrignani 2016). In the near future there will be direct detection experiments that test some of the most promising parameter space for WIMPs predicted by supersymmetry (Cushman et al. 2013). This is a very exciting time for the field, with the potential to discover a supersymmetric particle or tightly constrain the model.

A WIMP is very challenging to observe in a direct detection experiment. First, a target nucleus must be chosen, this choice determines what phase space of WIMP masses and interactions the experiment will be sensitive to. The highest sensitivity will occur for WIMPs that are of similar mass to the target nucleus. The interaction can also be spin-dependent or spin-independent, meaning that the cross section is impacted by the spin of the target or the atomic number respectively (it is of course easier to find targets with high atomic number than high spin). Interactions between WIMPs and regular matter should occur very rarely and will deposit energy via interactions with the nucleus of an atom causing a nuclear recoil (Boulay and Hime 2006). This is also how neutrons interact, making them a source of background signal that many direct detection dark matter experiments must address. It is necessary to shield any detector from cosmic rays as best as possible, otherwise any signal would be indistinguishable from the high rate of background. This is generally accomplished by going underground and using the Earth as a shield. For example, the muon particle flux at the surface of earth is roughly $15 \times 10^6\,\mathrm{m}^{-2}\mathrm{day}^{-1}$ as opposed to $0.27\,\mathrm{m}^{-2}\mathrm{day}^{-1}$ in the SNOLAB facility (Duncan et al. 2010). For most other cosmic rays, such as

protons and gamma rays, the expected flux from the surface is zero over the lifetime of the experiment. WIMPs would be essentially unobstructed by the earth, allowing them to enter the detector. It is also important that a dark matter detector is built from materials with low concentrations of radioactive material, as it would negate the benefit of going underground if there was still a high rate of background events. For what little activity remains, projects must go to great lengths to characterize the backgrounds to reliably distinguish them from a dark matter signal.

1.2.6 The Λ Cold Dark Matter (ΛCDM) Paradigm

As the culmination of many observational signatures, the Λ Cold Dark Matter (ΛCDM) paradigm represents the standard model of cosmology. Its name combines the observations of accelerating expansion of space and characterizes how dark matter acts on large scale structure. At the largest scales, this model reproduces observations from the local universe to the CMB. It combines a general relativistic understanding of the universe using the Friedman-Lemaître-Robinson-Walker (FLRW) metric (Dodelson and Efstathiou 2004) with a prediction of a new particle (Feng 2010). These come together to simultaneously describe a universe which is expanding at an accelerating rate, and in which local structure grows with time as minute quantum fluctuations from the early universe are amplified.

The accelerating expansion of space was first discovered with type Ia supernova distance measurements (Riess et al. 1998; Perlmutter et al. 1999). This could be achieved in general relativity with a constant Λ which contributes to Einstein's field equation (Zel'dovich 1968; Frieman et al. 2008; Planck Collaboration et al. 2016b). Λ acts as an energy density which does not decrease as the universe expands; meanwhile

the matter energy density decreases as a^{-3} (i.e. with volume) where a is the FLRW scale, and radiation energy density decreases as a^{-4} (i.e. with volume and redshift).

It is observed that at large scales matter forms clusters, filaments, and voids with specific statistical properties (Peacock et al. 2001). Yet at early times, the universe was extremely smooth, with temperature fluctuations in the CMB being of order a few hundred μK (Planck Collaboration et al. 2016a). Thus, some mechanism must have caused these structures to evolve. Cold dark matter represents a family of particle models in which the particle is kinematically cold (has non-relativistic velocities at the time when major structures were formed) and so can condense into gravitational potential wells to form the halos in which we observe galaxies and clusters today. In certain parametrizations, CDM is capable of generating the observed large scale structure in simulations (Springel et al. 2006).

In ΛCDM, galaxies and clusters form hierarchically, with smaller structures forming first and combining to form larger structures with time (Schneider 2006). Merging structures form larger ones in a chaotic process that can rearrange the matter in a halo and even combine with different fractions of dark matter to baryonic matter (Reddick et al. 2013). In ΛCDM the distribution of baryonic matter and dark matter within a galaxy is quite different although gravity can act between the two in several ways (Dutton et al. 2007). The complexity of phenomena described by ΛCDM is remarkable given that it has so few parameters (Λ and one particle).

Part I

MACHINE LEARNING TO BOOST SURFACE EVENT DISCRIMINATION IN THE DEAP-3600 DETECTOR

Chapter 2

Outline

Part I explores the use of Machine Learning algorithms for the purpose of improving the background suppression in DEAP-3600. Surface background events occur at the boundary of the active volume in the detector and must be identified by position reconstruction. To begin, Chapter 3 lays out useful information for the project. Section 3.1 describes the design of the detector and the background management analysis. Section 3.2 introduces primary Machine Learning concepts, with basic descriptions of each algorithm used in the thesis. The analysis begins in Chapter 4 by calibrating/training and testing all algorithms in simulation, where the truth information is known completely. Here it is found that the Machine Learning algorithms are able to work alongside a standard position reconstruction tool (MBLikelihood) to reduce the background rate by nearly an order of magnitude. However, using only the simulator raises concerns that the Machine Learning algorithms might overfit specifics of the simulator and not properly represent the real detector. To address this, Chapter 5 explains how it is possible to train the Machine Learning algorithms with data directly from the detector. It is not possible to adequately test the algorithms on detector data, but it is shown that the algorithms perform just as well

at discriminating simulated surface events even when trained solely on real events. With the Machine Learning algorithms trained on real data, Chapter 6 tests them on simulations with modifications to the model of the detector. By observing how their performance changes, given reasonable changes to the detector model, it is possible to get a sense of how robust the algorithms are. In all cases, the Machine Learning algorithms roughly match the trends observed in MBLikelihood, which was the benchmark for comparison as it was specially designed for DEAP-3600. Finally, Chapter 7 tries to find other aspects of the algorithms that may be behaving pathologically. In Section 7.1 it is seen that lower energy events are, as expected, more challenging to classify, and in Section 7.4 it is shown that using more than one Machine Learning in tandem with MBLikelihood provides no extra benefit. Section 7.2 shows that the simulator is indeed different from the real data from the perspective of position reconstruction as there are clear differences in how events are distributed in the two cases, with the Machine Learning algorithms being the most sensitive to the differences. With improved position reconstruction it is possible to choose a larger fiducial volume, Section 7.3 demonstrates how much can be gained in this direction. A critical aspect of live analysis is the processing speed, Section 7.5 shows that all but one of the Machine Learning algorithms can potentially be applied to the live analysis. Chapter 8 summarizes the results and considers avenues for further exploration to improve performance or check that the algorithms are behaving as expected.

Chapter 3

Detector Description and Analysis Overview

3.1 DEAP-3600

Using a liquid argon target volume, DEAP-3600 is projected to be sensitive to a cross section of $10^{-46}\,\mathrm{cm}^2$ for a spin-independent WIMP-nucleon interaction with a WIMP mass of $100\,\mathrm{GeV/c}^2$ (DEAP-3600 Collaboration et al. 2017). A significant factor involved in reaching that sensitivity is a detailed understanding of all backgrounds in the detector, as well as robust mechanisms for discriminating them from a potential WIMP signal. However, this projected sensitivity is only possible thanks to the extreme care that was taken in the design and construction of the detector, which is discussed in this section.

3.1.1 Design

DEAP-3600 is located in the SNOLAB facility, located 2 km underground in Sudbury Ontario, which provides shielding from cosmic rays (Duncan et al. 2010). The detector is contained in a steel shell that is submerged in a large water tank. Located inside the steel shell, an 85 cm radius Acrylic Vessel (AV) is designed to hold 3600 kg of

liquid argon in a sphere. A nominal central fiducial tonne of liquid argon has been defined as the target volume for dark matter searches. The fiducial volume is chosen to be central, away from the surface where higher background rates are expected. Currently the detector is operated in a partial fill configuration with approximately 3300 kg of liquid argon total. This is due to an engineering design issue that could lead to a seal failure and radioactive contamination being introduced into the active volume if the detector was filled completely. Bonded to the AV are 255 cylindrical acrylic Light Guides (LGs) which optically couple the argon volume to an array of 255 Photo-Multiplier Tubes (PMTs). The LGs attached to the AV provide 76 % coverage facing the argon volume, but leave gaps that are filled by blocks of stacked high density polyethylene and insulating layers called Filler Blocks (FBs). All of this is shown in Figure 3.1. The LGs and FBs provide neutron shielding for the argon volume, which is very important as neutrons can mimic a WIMP signal. PMT glass is a significant source of neutrons and was one of the primary reasons for including the LGs and FBs to distance the PMTs from the active volume. Also, a thermal gradient is maintained between the liquid argon at $-200\,\mathrm{C}^\circ$ and the PMTs at $-20\,\mathrm{C}^\circ$ by the LGs and FBs, this is an ideal temperature range for the PMTs. The acrylic of the LGs and AV is opaque to the scintillation light produced within the liquid argon, thus a thin layer of wavelength shifter 1,1,4,4-tetraphenyl-1,3-butadiene (TPB) was deposited on the inner surface of the acrylic sphere. A more detailed description of the design can be found in (Jillings and DEAP Collaboration 2016).

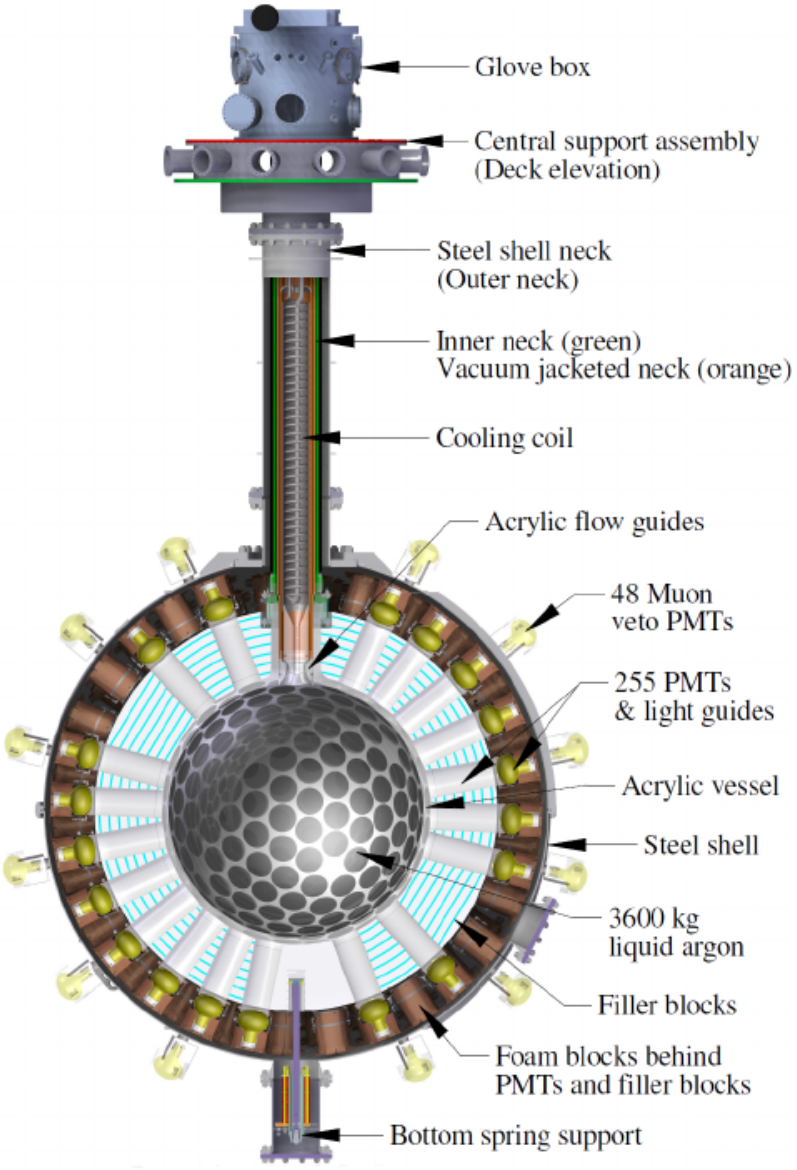

Figure 3.1: Design of the DEAP-3600 detector sliced through the meridian. Not shown is the TPB layer on the inner surface of the AV.

3.1.2 Stages of an Event

An event begins when a WIMP, alpha, neutron, etc. interacts with an ^{40}Ar atom causing it to ionize or excite, and recoil. The recoiling nucleus may ionize or excite

more ^{40}Ar atoms until it slows down. Excited ^{40}Ar atoms form dimers that can exist in two spin configurations; a singlet state with spin 0, or a triplet state with spin 0 or 1 (1 for two members of the triplet and 0 for the other). When the dimers decay they emit scintillation light, but the timescale of that delay into a photon is suppressed for the triplet because of its spin. Thus, there are two characteristic lifetimes for this decay: 7 ns and 1.6 µs, depending on the spin state (Hitachi et al. 1983). Nuclear recoils are more likely excite the singlet state, while the different excitation path of electronic recoils is more likely to excite the triplet state.

The scintillation light is emitted isotropically with a wavelength of 128 nm; this is below the excitation energy of an individual ^{40}Ar atom meaning that the light can travel mostly unobstructed through the argon volume. When a photon reaches the TPB surface it is absorbed; the TPB then emits a photon with a direction chosen from an isotropic distribution, the photon wavelength is drawn from a distribution peaked at 420 nm (Pollmann et al. 2011). As it can now travel freely through both liquid argon and acrylic, the photon is directed up a light guide by total internal reflection until it reaches a PMT and may be detected. The PMTs are 8 inch, high quantum efficiency Hamamatsu R5912 tubes with a quantum efficiency of approximately 25 %. Each PMT outputs a charge that is recorded and integrated for each event, which can be measured for an individual PMT or as a total for all 255 PMTs. The integrated charge is converted to a measure of the number of photo-electrons liberated within each PMT, called QPE (Q for charge and PE for photo-electrons). The QPE measurement only approximately represents the number of photo-electrons, as it may count other effects such as afterpulsing. Afterpulsing is a process by which back diffusing ions from the cascade process can create extra pulses later in the waveform. These extra pulses

may erroneously be identified as photoelectrons (Lindner 2015).

3.1.3 Background Management

The first stage of background management is simply the location of the detector at the SNOLAB facility. By going far underground most cosmic rays are blocked. The most notable surviving cosmic radiation is the muon flux, which is only $0.27\ \mathrm{m}^{-2}\mathrm{day}^{-1}$ (Duncan et al. 2010). The few residual muon events can be vetoed using an array of PMTs outside the steel shell, but inside the water tank; these detect Cherenkov radiation as a muon travels through the water. The water tank itself shields from radioactive sources in the lab or the rock of the mine. SNOLAB is a class 2000 clean room, reducing the amount of radioactive material that the detector was exposed to during construction. Most notably, radon in the air can be deposited on the AV during construction, so it is critical to have a clean facility. Some amount of contamination during construction was inevitable, so the inner surface of the AV was sanded away in a pure nitrogen environment immediately before liquid argon filling. Another primary way that the background is managed is by selecting materials for low radioactivity, especially the acrylic which was specially manufactured to be as clean as possible. The argon used is very pure, and an on site purification system cleans the argon immediately before it enters the detector. The purification system is specifically designed to remove radon, nitrogen, and oxygen as these are radioactive or can decrease the scintillation efficiency. It was not possible to choose a sufficiently low radioactivity material for the PMTs and so part of the job of the LGs and FBs is to separate them from the active volume. These along with countless other design decisions have lead to an exceptionally low background budget (Jillings and DEAP Collaboration 2013,

2016; Lehnert 2018).

With extreme care taken to mitigate sources of radioactivity, only trace amounts of a few radioactive elements are left. The final stage of background management comes from a detailed understanding and strong discrimination of these sources. ^{39}Ar occurs naturally at about 10^{-16} $\frac{\text{g } ^{39}\text{Ar}}{\text{g Ar}}$ in liquid argon (Benetti et al. 2007). At approximately 3600 Hz, DEAP-3600 observes ^{39}Ar beta decays. These ^{39}Ar betas are discriminated from WIMP-like events using their pulse shape in time (see Section 3.1.2), which has a characteristically high fraction of light later in the pulse; that fraction is called fprompt. Specifically, fprompt is calculated as the ratio of integrals $\int_{-50}^{150} V(t)dt$ and $\int_{-50}^{15000} V(t)dt$ where t is in nanoseconds and $V(t)$ is the amplified voltage output from the PMTs. The discrimination power of fprompt is of order 10^{10} in the energy region of interest (DEAP-3600 Collaboration et al. 2017). Pulse shape discrimination does not work for nuclear recoil events such as those which are generated by ^{210}Po alpha decays. ^{210}Po events occur at the surface of the AV, both the recoiling nucleus and the produced alpha have a very short mean free path, meaning that all light is produced from a point in or near the acrylic surface. For this reason it is desirable to only select events which occur in the bulk of the argon volume, away from the acrylic, for further analysis. The volume in which events are accepted is called the fiducial volume. Several position determination algorithms were created for use in DEAP-3600, among them is the MBLikelihood algorithm. With a 3D position computed for each event it is possible to select a geometry within which events are accepted for further analysis (for example, a centred sphere). The purpose of Part I will be to assist MBLikelihood (or any other position reconstruction algorithm) with properly characterizing events as background or signal, hence optimizing the the fiducial volume, and the sensitivity

of the experiment.

3.1.4 Fiducial Volume by Position Reconstruction

The focus of this work is on using the QPE charge information from the array of 255 PMTs to recover the position of an event in the detector. Due to the use of TPB and the size of the detector, it is challenging to use timing information from the PMTs to determine the location of an event. TPB re-emits the wavelength shifted photons on a timescale of 11 ns (Pollmann et al. 2011), with this uncertainty in the timing of light arriving at a given PMT the uncertainty in position would be on the same order as the radius of the detector. Instead, the relative charge in each PMT is analyzed to extract the position of an event. An event depositing energy in the argon will emit light isotropically. If that event occurs in the centre of the detector then, on average, light should be evenly distributed in the PMTs. For an event which occurs near the surface of the detector, a small number of PMTs will subtend a large solid angle and collect a disproportionately large fraction of the total light. The Region Of Interest (ROI) includes events that will induce as few as 120 photo-electrons in the PMTs. This is smaller than the number of PMTs surrounding the detector, meaning that Poisson noise is a dominant source of uncertainty. This makes the task of position reconstruction difficult, but it is still possible to extract position information from the relative charges. Position reconstruction is further complicated by the TPB layer at the argon/acrylic interface, which wavelength shifts the light and re-emits it isotropically (Pollmann et al. 2011). A great deal of effort has been put into the position reconstruction algorithms for DEAP-3600 to address these difficulties.

There are only a few deviations from spherical symmetry in DEAP-3600 (Jillings

and DEAP Collaboration 2016). One of the largest deviations being the partial fill geometry which can cause asymmetries by total internal reflection at the liquid-gas interface. For an event that occurs in the liquid argon, reflections at the liquid-gas interface will reduce the amount of light seen by the PMTs above the interface. For an algorithm that assumes spherical symmetry, this effect will cause the reconstructed position to be biased downwards as the algorithm is presented with reduced light in the top PMTs. Another deviation from spherical symmetry is the neck of the detector. Light that travels up the neck will be lost to the array of 255 PMTs and so will not be accounted for in position reconstruction, again biasing the reconstructed position to be lower in the detector. A less impactful source of asymmetry comes from the placement of the PMTs. While the spacing is generally regular, there are 11 locations where an extra large gap occurs (often called a *pentagon* as there are 5 PMTs adjacent to it). If an event occurs on the surface in front of one of these gaps then it is more likely for light to be reflected to the opposite side of the detector and cause the reconstructed position to move towards the centre of the detector. These sources of asymmetry must be accounted for during position reconstruction to achieve an optimal resolution.

The MBLikelihood algorithm performs a likelihood maximization routine on the position of an event. MBLikelihood is trained/calibrated using simulated data at various locations in the detector, which are chosen to give a representative sample of radial and angular positions. Using the training data, a table is generated which gives a Probability Density Function (PDF) of the charge in each PMT. PDFs are generated as a smooth function of r and θ, excluding ϕ as the detector is effectively symmetric to rotations about the z-axis. It is then possible to determine the 3D

position of an event by searching for the maximum likelihood PDF (with ϕ chosen afterwards to maximize likelihood for a given r and θ).

Using tests in a Monte-Carlo simulation it has been shown that MBLikelihood performs very well at reconstructing the location of each event, with a radial random uncertainty in position of order 8 cm near the surface and slightly worse towards the center (Rethmeier 2017). Position uncertainty can also determined using Bismuth-Polonium (BiPo) tagging in the detector (Xu et al. 2017). Trace amounts of ^{222}Rn in the liquid argon has a decay chain resulting in ^{214}Bi which undergoes a beta decay to produce ^{214}Po, this quickly undergoes alpha decay ($\sim 160\,\mu$s) to produce ^{210}Pb which has a longer lifetime. By scanning through the data for alpha events preceded by a high energy beta event, it is possible to find pairs of events that occur at effectively identical locations. A comparison of the reconstructed locations of the two events can give the random uncertainty in MBLikelihood position determination, which roughly agrees with simulations. This does not, however, identify systematic biases in the position reconstruction.

Using the reconstructed MBLikelihood position a fiducial volume is selected to contain one tonne of liquid argon. The fiducial volume of one tonne is chosen for consistency throughout the thesis in order to make meaningful comparisons between algorithms. While the DEAP-3600 collaboration may choose to use a different size for the fiducial volume, it is expected that a value of one tonne is close enough to provide a relevant validation of the method. In Chapter 7 the effect of choosing a different fiducial volume size is explored briefly. For simplicity, this thesis will always consider a spherical geometry for the fiducial volume centred at the centre of the AV. The DEAP-3600 collaboration may choose to use a non-spherical geometry for

the fiducial volume, but a sphere will likely be close to the final geometry. The fraction of events that occur at the surface of the detector, but reconstruct inside the fiducial volume will be called the surface event leakage or leakage for brevity and when the context is clear. The fraction of events which occur in the liquid argon and are reconstructed within the fiducial volume will be called the *Acceptance* and is generally treated as a free parameter to be fixed such that the fiducial volume is one tonne.

In the context of Machine Learning, the problem can be written simply as: given a vector of 255 PMT charges, output a single value which can be used as a cut to distinguish surface events from fiducial events. One could simply output a Boolean value, however it is advantageous to instead output a single continuous value. To get a binary classification, a threshold value is chosen above which all events are rejected. In analogy with MBLikelihood, if the fiducial volume is a sphere then the reconstructed radius of an event is a single continuous value for which a threshold can be chosen. Any value above the threshold is rejected.

Having 3D information about each event is still useful for other analyses and cuts. For example, with the partial fill configuration of the detector it is useful to have a z-axis (parallel to the neck) cut to remove events which occur at or above the liquid-gas interface. The intention in this thesis is to improve the rejection of surface events, which can be validated by maintaining the same effective fiducial volume and reducing the surface event leakage. This has been achieved by allowing MBLikelihood to accept a larger volume (higher radius threshold); then the machine learning algorithm is used to remove events until the same effective volume is reached as the original fiducial volume. An effective volume can be defined based on the uniformly distributed ^{39}Ar,

a tonne of fiducial mass is defined by a sphere of radius 55.46 cm and would contain 1 kHz of beta events, a tonne of effective fiducial mass also has 1 kHz of beta events but the geometry containing that volume is no longer well defined. By requiring that the boosted algorithm have the same acceptance as MBLikelihood meaningful comparisons can be made about the surface event leakage.

3.2 Machine Learning

This section provides a brief background on Machine Learning as is relevant for this thesis. A description of classification problems is discussed and why fiducial volume selection can be interpreted as this type of problem. Preprocessing is a critical step in Machine Learning, and some of the basic methods are discussed here. Finally, a brief description of each Machine Learning algorithm is provided with reference to more detailed material. The algorithms used in this analysis were: Logistic Regression, Support Vector Machine, Neural Network, Naive Bayes, and K-Nearest Neighbours. For more detailed descriptions, see Shai Shalev-Shwartz (2014) and accompanying references.

3.2.1 Classification

Machine Learning focuses on the goal of generating a mapping from inputs $\vec{x}$ to outputs $\vec{y}$, where the mapping is learned from sample data and is intended to represent an underlying distribution. The outputs $\vec{y}$ can be discrete, in which case the mapping is called a labelling or a classification problem. If the output has come from a continuous distribution, then the mapping is a regression problem. A large number of tasks can be easily formulated as classification problems. In many classification

tasks, the decision is binary. In this thesis the output will be a scalar with low values representing fiducial like events and high values representing surface like events (in analogy with the MBLikelihood radius). Once a threshold value is chosen the algorithm outputs a binary decision. However future analyses may be able to use the scalar value to give a confidence in the prediction on a per-event level. The charges in each PMT are recorded in a 255 dimensional vector $\vec{x}$. The task is then to find a decision boundary in that 255D space that divides fiducial from surface events as well as possible. Fundamentally the goal is the same as defining a fiducial volume in 3D space, but this formulation provides access to a host of powerful algorithms that have been designed to solve this problem in general. Instead of investing time designing a custom algorithm, one can quickly generate discriminators and test their performance. Most of the analysis work in Machine Learning is focused on validation of the algorithm instead of development.

3.2.2 Preprocessing

It is common in Machine Learning to do some form of preprocessing on data before fitting. This often takes two forms. One is to modify the data with scaling and translating so that all features are considered equally. For example, if one wanted to predict the probability of a customer purchasing a house, it might be relevant to know the price of the house and the number of bedrooms. Two houses to compare could have price/bedroom input vectors of $(400000, 4)$ and $(300000, 2)$ respectively. Any algorithm that relied on Euclidean distances (which is most of them) would give no weight to the number of bedrooms as $\sqrt{(400000 - 300000)^2 + (4 - 2)^2}$ is dominated by the first term. This can be resolved by converting all values into a more consistent

range, such as $(0, 1)$ or $(-1, 1)$ as appropriate. In this case it would make sense to divide all house prices by 1 million and all bedroom numbers by 6 (for example). A more extreme version of this is called whitening, which uses principal component analysis to eliminate correlation between each variable by reorienting basis vectors in the input space. Whitening ensures that each input to the algorithm provides distinct information (Ball and Brunner 2010).

The second form of preprocessing involves feature selection or dimensionality reduction. This is the process of deciding what information is the most useful for an algorithm in its analysis. The simplest form of this is just dropping some of the data, for example in the house buying predictor it is likely not helpful to know the height of the previous owner. This can speed up the training of the algorithm as it does not have to search through as many possibilities. Often called *the curse of dimensionality*, the difficulty (training time) of determining a decision boundary increases exponentially with the number of input dimensions. Most Machine Learning algorithms are designed to increase computation time with the number of dimensions as $\mathcal{O}(d^n)$ where d is the number of dimensions and n is an integer; this is a considerable gain compared exponential $\mathcal{O}(e^d)$ time increases, but can still easily be prohibative if d is large. Still, large gains can be made by disposing of useless information.

More complicated preprocessing can involve doing some calculations before the algorithm sees the data. For example, in the house buying case it might be useful for the algorithm to know the ratio of bedrooms to bathrooms instead of, or in addition to, knowing each parameter individually. This way the focus of the algorithm is on the most important aspects of the data.

3.2.3 Hyper-Parameters

The training process in Machine Learning involves optimizing a set of parameters that define the decision boundary for that algorithm. However, there are some parameters which define the behaviour of the training process itself, or represent attributes that should not be altered during the training process. For example, when performing gradient descent to minimize some function L (for example the squared difference between predicted value and true value) one updates the parameters as follows:

$$x_{i+1} = x_i - \alpha \frac{\partial L}{\partial x_i},$$

the updated x_{i+1} parameter value is modified by $\alpha \frac{\partial L}{\partial x_i}$ in the direction of the minimum. The value α in this procedure is a hyper-parameter as it determines the behaviour of the training routine. A high value of α will move quickly towards the minimum, but may overshoot or be unstable. A low value of α may take far too long to reach the minimum and would waste computation time. It is also possible to alter α adaptively while fitting, however this just means that one's choice of technique is a form of hyper-parameter. As the shape of the parameter space being optimized is not known ahead of time, it is necessary to try several values of α in order to select one which behaves well for a specific problem.

The values chosen for the hyper-parameters can have a large impact on the performance of a Machine Learning algorithm. It is often not clear how to choose the best value for each hyper-parameter and so a grid search is generally used, but there are techniques to converge faster (Bergstra and Bengio 2012). Some algorithms have a large number of hyper-parameters making the training process computationally

expensive as multiple versions of the algorithm must be trained to find the best performing one. No Machine Learning algorithm will perform well on all problems (called a *no free lunch theorem* in Wolpert and Macready (1997)), but having more hyper-parameters can make an algorithm more versatile. Neural Networks are notorious for having large numbers of hyper-parameters, including multiple types of regularization and several ways to arrange the nodes in the network (Glorot and Bengio 2010). The trade off for having so many hyper-parameters is that Neural Networks can have exceptionally good performance once trained. The key features of a number of Machine Learning algorithms are described in the following sections.

3.2.4 Logistic Regression (LR)

The Logistic Regression classifier defines a hyperplane that separates two classes in the N dimensional space defined by the domain of the input vector $\vec{x}$ (Ng and Jordan 2002). The training process involves computing a weight vector $\vec{w}$ and an offset b so that for any given input $\vec{x}$ the function:

$$f(\vec{x}) = \frac{1}{1 + e^{-(\vec{x}\cdot\vec{w}-b)}},$$

will return values close to 1 for one class, and 0 for another. The form of the function restricts the value to the range $(0, 1)$ and in some situations can be interpreted as a probability. This classifier is very quick to train and run, as its largest computation is a single dot product. It has become a standard classifier to apply first on any problem to see if a simple linear decision boundary is sufficient before investing time and energy in more complicated algorithms.

3.2.5 Naive Bayes (NB)

A Naive Bayes algorithm analyzes individual inputs separately (Ng and Jordan 2002). It assumes that each variable in the input is independently distributed and then uses Bayes theorem to compute the posterior probability that a new event belongs to a class based on the sample data provided during training. The ultimate goal of any Machine Learning problem would be to compute the probability $P(\vec{y}|\vec{x})$ or in other words: the probability of a label being correct for a given input. That probability is not feasible to compute directly, so the Naive Bayes algorithm makes the assumption that every dimension of $\vec{x}$ is independent. Explicitly, the assumption made is: $P(x_i|\vec{y}, x_0, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = P(x_i|\vec{y})$ where the form of $P(x_i|\vec{y})$ can be chosen as seen fit for a given problem. For this thesis, the chosen form of $P(x_i|\vec{y})$ is the common Gaussian function:

$$P(x_i|\vec{y}) = \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}},$$

where μ_y is the mean and σ_y is the standard deviation. The final probability assigned to a given label is determined from this naive interpretation of Bayes theorem as: $P(\vec{y}|\vec{x}) = \frac{P(\vec{y})\Pi_i^n P(x_i|\vec{y})}{P(\vec{x})}$. To determine the class for a given input, the label with the highest probability is chosen. The factor $P(\vec{y})$ is chosen from the relative frequency of the labels given in the training set, or can be chosen to be uniform if the proportions in the training set are thought to be contrived. The factor $P(\vec{x})$ is not computed as it will be the same for every label, and so does not contribute to the classification. (Pedregosa et al. 2011)

Take, for example, spam detection in email, which amounts to assuming that the probability of each word appearing is independent, but that some words are

more commonly associated with spam. Thus, the conditional probability of the email being spam given that the word "free" is in the email might be high, but a name from your contact list would have a low probability. Each word will have a probability of appearing in any email (its frequency) and a probability of being associated with spam, which can be used together to provide a probability that an email is spam. Taking into account many words will give a good impression as to whether or not a given email should be classified as spam. It is obviously incorrect to assume that the appearance of any word is independent of the others, but this approach performs very well at some tasks (Domingos and Pazzani 1997).

3.2.6 Support Vector Machine (SVM)

In its most basic form, a Support Vector Machine creates a linear hyperplane that best separates two classes, similar to the Logistic Regression algorithm. If the sample data is linearly separable then the Support Vector Machine training algorithm places its decision boundary half way between the populations (Cortes and Vapnik 1995; Suykens and Vandewalle 1999; Tong and Koller 2001). Noise in the data may make the two classes cross, in which case a penalty in the optimization routine can be used to handle misclassified sample data. For a new event $\vec{x}$ one can compute the distance to the plane using an inner product, a threshold value (often 0) of the result decides which class that event belongs to. However, not all data is linearly separable and the real power of Support Vector Machines comes from how they manage this. Using a kernel function ψ, defined by the user, which represents some metric of distance between any two events, a Support Vector Machine maps data into a higher dimensional space where it is linearly separable. The resulting high dimensional

decision boundary can be represented by a subset of the training data chosen during training. The decision boundary is then defined by scalar weights a_i associated with each training data sample $\vec{t_i}$ so that a new event can be labelled using this formula:

$$d(\vec{x}) = b + \sum_i a_i \psi(\vec{x}, \vec{t_i}),$$

where b is a scalar offset just as with a linear function. To choose a kernel function some measure of similarity is needed between events; many of them are based on Euclidean distance, but in some environments there may be a more suitable distance metric such as the taxicab metric (distance only moving along grid-lines) when movement is confined to a grid (Thompson 2011). A commonly used kernel function is called the *radial basis function*, which takes the form: $\psi(\vec{x}, \vec{y}) = e^{-\gamma||\vec{x}-\vec{y}||^2}$ where γ is a hyper-parameter (Pedregosa et al. 2011). The radial basis function is simply a Gaussian evaluated on the Euclidean distance between two events. Thus, the label of an event $\vec{x}$ is determined by the weights a_i of the training samples that are nearest to it in Euclidean distance. Another common kernel is the *polynomial kernel* which takes the form: $\psi(\vec{x}, \vec{y}) = (\vec{x} \cdot \vec{y} + c)^d$, where c and d are hyper-parameters; when $d = 1$, one has a linear kernel. In this thesis the Linear Support Vector Machine (identified as LSVM) and the Radial Basis Function Support Vector Machine (identified as SVM) kernels were used.

3.2.7 Neural Network (NN)

A Neural Network combines several simple classifiers into a larger, more powerful classifier. The simple classifiers are called nodes, where each node is a linear decision boundary, just like the Logistic Regression algorithm. A node performs a dot product

with a vector of weights $\vec{w}$ and an input vector $\vec{x}$ the result of which is passed to a response function ψ to produce a single valued output. The result is: $\psi(\vec{w}\cdot\vec{x})$ which is a scalar value that depends on the response function. Common choices for the response function are the sigmoid (also called logistic function): $\psi_{sigmoid}(a) = 1/(1+e^{-a})$, or the rectified linear unit: $\psi_{relu}(a) = \max(0, a)$ (Levin and Fleisher 1988). These nodes are combined in layers to give a vector of outputs instead of a single value as: $o_i = \psi(\vec{w_i} \cdot \vec{x})$. The weight vectors for each node can be combined into a single weight matrix, making the calculation for the layer simplify to: $\vec{o} = \psi(W\vec{x})$. With multiple layers each computation is composed as follows: $\vec{o_2} = \psi(W_2\vec{o_1}) = \psi(W_2\psi(W_1\vec{x}))$, which can continue for an arbitrary number of layers (Krizhevsky et al. 2012).

A feed forward Neural Network is designed with at least 3 layers. The lowest layer is simply the vector of inputs to the network, where one node corresponds to each variable. The middle layers can be composed of any number of nodes that each perform a dot product with the output from the layer below and a unique set of weights. After each node does the dot product it evaluates a response function and outputs the result for the next layer to use. The final layer will have a number of nodes equal to the number of classes to be labelled and so classification is determined by the maximally activated node. By normalizing the sum of the output node values to be 1, it is possible to get a measure of confidence in a classification as well. If two or more top nodes have nearly identical activation then the confidence would be low.

There are a great number of node arrangements that allow Neural Networks to operate in different contexts (Goodfellow et al. 2014; Szegedy et al. 2015). For example, convolutional layers are often used in image analysis (Krizhevsky et al. 2012). Convolutional layers use nodes that only receive input from a small patch of the image

at a time, encoding the assumption that local information is most important in an image. Another example is a recursive network, in which the network scans through the signal only looking at a small segment at a time (Graves et al. 2013). Using some form of persistence during the scan a recursive network is useful for analyzing signals in which the direction of time is important. For the purposes of surface event classification in DEAP-3600 a feed forward network was deemed to be the most appropriate choice of architecture.

3.2.8 K-Nearest Neighbours (KNN)

K-Nearest Neighbours belongs to a group of classifiers known as *lazy learners*. There is no real training step for K-Nearest Neighbours except to save the training data and labels in a useful format. When classifying an event, it simply finds the k training samples that are most similar to it (using the Euclidean distance) and determines the average class for the k neighbours (Weinberger and Saul 2009). It is also possible to weight the k neighbours based on distance; this is because training samples that are more similar to the new event should take priority in determining the class. The weighting is especially relevant when the parameter space is not densely filled, in which case a "neighbour" may be very different from the sample just because there aren't many nearby. If computation time and amount of sample data is not an issue, then K-Nearest Neighbours is an excellent algorithm that will always reach an accurate reflection of the sample distribution (Weinberger and Saul 2009). The computation time is explored briefly in Section 7.5, K-Nearest Neighbours ended up being quite slow as it needed to compute euclidean distances for many thousands of samples each time it made a prediction.

Chapter 4

Training with Simulated Data

Before moving on to train the Machine Learning algorithms with real detector data, it will be useful to train them using simulated data. All the truth information is known for simulated events, meaning that the performance can be reliably measured. Those reliable measurements, however, are only meaningful to the degree to which the Monte-Carlo simulation is trusted. By first using the simulated data, a baseline can be established to determine what sort of performance should be expected later. Deviations from the baseline could imply over fitting, a poor Monte-Carlo model, or a faulty assumption in the training procedure with real data.

4.1 Procedure

Training a Machine Learning algorithm requires samples of data that have true labels already assigned to each event. The accuracy of the "true" labels is ideally perfect (as in the case of a simulation), but some algorithms are able to handle some noise in the labelling. This requirement lends itself to simulated data as the true values for the position of an event can be known perfectly and even controlled to get samples best suited for training. The goal of Part I is to train algorithms to identify WIMP

events that occur centrally in the detector against events which occur near the TPB surface. Thus, two samples will be simulated to represent argon volume events and surface events. The training procedure will involve separating as many of the argon volume events from the surface events as possible.

Any true surface contamination should be embedded in the acrylic or TPB and, when it decays, may deposit energy with daughter particles travelling towards the argon volume. There are computational challenges to simulating enough of these events in the energy region of interest for training. Also, it is expected that this would only make the surface signal stronger as more light is directed towards a few PMTs for events which deposit energy directly into the acrylic. A small sample was used for validation of this hypothesis and is discussed later (Section 6.3).

It will be useful to characterize how MBLikelihood performs on these data before training the Machine Learning algorithms to get a baseline for comparison. After the algorithms have been trained and the best hyper-parameters selected, their individual performance will be compared to MBLikelihood. Finally, each algorithm will be used to boost the performance of MBLikelihood and compared to how any of them would have performed on their own. With this procedure established, a more robust analysis will be done in Chapter 5 using real data from the detector.

4.2 Simulated ^{40}Ar Events

The Reactor Analysis Tool (RAT) is used by the DEAP-3600 collaboration as a synthesized code base for simulation and analysis of data. This makes it easy to use the same analysis flow on simulated data and measured data. Physics simulation is performed by GEANT4 (Agostinelli et al. 2003) in a detector geometry defined within

RAT (Caldwell 2014). This geometry as well as other properties such as the optical model of the detector materials can be modified with simple commands given in a macro that RAT reads. This system is capable of simulating a wide array of event types at any location in the detector or around it.

The hypothesized WIMP interactions and alpha backgrounds deposit energy in the detector via nuclear recoils. ^{40}Ar nuclear recoils were simulated to use as training and testing data for the Machine Learning algorithms. There was no attempt made to have the energy distribution of the ^{40}Ar nuclear recoils accurately reflect the WIMP interactions or an alpha source (such as ^{210}Po); instead the energy of the simulated events was chosen such that they produced a uniform distribution in the energy ROI. This was primarily for two reasons. First is that the energy distributions of WIMPs and alpha backgrounds have not been fully characterized and so any choice would only be an approximation. Second, the algorithms were intended to take advantage of position dependent information only and so should perform as well as possible at all energies in the ROI. Any non-uniform energy distribution would necessarily provide the algorithm with less training data in some regions, thus decreasing its performance. Adding energy dependent information would be its own study and can be applied independently of the position information (or any of the other cuts). If one were to evaluate events with a probability of being a WIMP, the energy information could be thought of as a prior probability distribution applied to the event before assessing position information. Later in the thesis there will be a discussion of how the algorithms perform as a function of energy within the ROI (Section 7.1). The WIMPs and alpha events are simulated in two regions, the first is evenly distributed throughout the argon volume, the second is evenly distributed in a shell that is 0.5 cm

thick, going from a radius of 84.5 cm to 85 cm. For both regions the events are only simulated where there is liquid argon, and so none occur above a z-axis value of 55 cm (30 cm below the neck).

Many events were simulated to test the fiducial volume analysis. In total, over 10 million events were simulated to perform various tests on the algorithms. To train the algorithms, 200,000 argon volume events and 200,000 surface events were simulated. Two more datasets were generated that are similar to the training data, they were used for validation and testing. The validation data is used to choose the best performing hyper-parameters for each algorithm and the testing data is used solely to measure performance.

Only ^{40}Ar recoils were simulated for this analysis. While there are many expected sources of background events (Cherenkov, ^{39}Ar beta, gamma rays, etc.) at varying rates, the sole focus of this analysis is position information. On real data, the fiducial volume analysis is done late in a chain of analysis routines intended to eliminate these other sources of background. It is thus only interesting for this analysis to measure the performance of the algorithms on nuclear recoil events at different positions in the detector. Other studies are done that test a full signal and background model with a full analysis pipeline.

4.3 Region of Interest (ROI)

Due to the cuts that were applied to define a representative ROI, only about 120,000 out of each 200,000 event set were usable. Only 50,000 of each class are needed for training, but all surviving events in the testing sets are used to ensure that the noise in leakage measurements is minimal. The need for 50,000 events was determined

by training the algorithms on increasing an number of events until their performance stabilized, which was found to be approximately 30,000 events, then 50,000 was chosen as conservatively above that without increasing the computation time dramatically. The majority of the lost events were outside of the energy ROI as the energy resolution is not perfect. Events with a flat simulated energy distribution will not have a flat measured energy distribution as the response of the detector is convolved with the input, thus many events had to be cut. Other cuts had a small effect on the total. The calibrated trigger time represents how well RAT was able to identify the start of an event and was used to remove simulated errors in the same way as is used for real data; in which it often removes pile-up events or other pathological distributions. A low level flag called *calcut* is used to identify events that did not meet certain standards in the data acquisition system, for instance, if the event rate is too high or if the trigger and digitizer timings are out of synchronization. This cut was used on the simulated data in the same manner as with real data, however for simulated data it only removes a tiny fraction of the events. While several other cuts were applied to get as close as possible to the real data analysis program, the nature of simulated data means that they were not truly needed and so did not remove many (if any) of the simulated events. To eliminate Cherenkov events that can occur in the LGs a parameter called *fmaxpe* is used; it records the fraction of light in the brightest PMT. This is because light produced in a LG will illuminate the corresponding PMT much more than the others. An fmaxpe cut was used to ensure consistency, however almost no simulated events were cut this way.

4.4 MBLikelihood on the Simulated Data

Before training and testing the Machine Learning algorithms on the simulated data, it will be useful to establish a baseline. MBLikelihood can reconstruct a full 3D position for each event. In this analysis the primary focus is on rejecting events from the surface, which can be done based on reconstructed radius. Figure 4.1 shows the reconstructed radius for both simulated argon volume events and surface events. The argon volume events uniformly fill the volume. The radial histogram for MBLikelihood, however, is not uniform in r^3 as would be expected for events uniformly distributed in the detector; this is due to radial biases in the algorithm. The most striking of these biases is the peak at $r = 850\,\text{mm}$ as this is where MBLikelihood places events that fail its likelihood maximization routine. A vertical line is included in the plot which represents the radius below which there is 1 tonne of liquid argon which is the nominal value. This threshold value is not placed based on a geometric volume, but instead by the total number of argon volume events that are accepted. In the partial fill geometry there are roughly 3300 kg of liquid argon and so to accept 1000 kg a fraction of $\frac{1000}{3300}$ liquid argon events should be accepted into the fiducial volume. This method of computing a fiducial volume will be useful later with algorithms that do not admit a geometric interpretation. With a radius of 60.5 cm the fraction of simulated surface events that pass into the fiducial volume is 0.002. This is not a measurement of the surface event leakage for MBLikelihood in the real detector, as noted earlier the simulated surface events are not intended to accurately model surface contamination in the acrylic or TPB. Simulating those events is more computationally expensive and would require more careful consideration as to the energy spectrum instead of using a uniform value. The surface leakage of 0.002 is still a good

value against which improvements in position reconstruction can be measured. If the Machine Learning analysis can reduce that value then it is very likely the true surface event leakage is lowered as well.

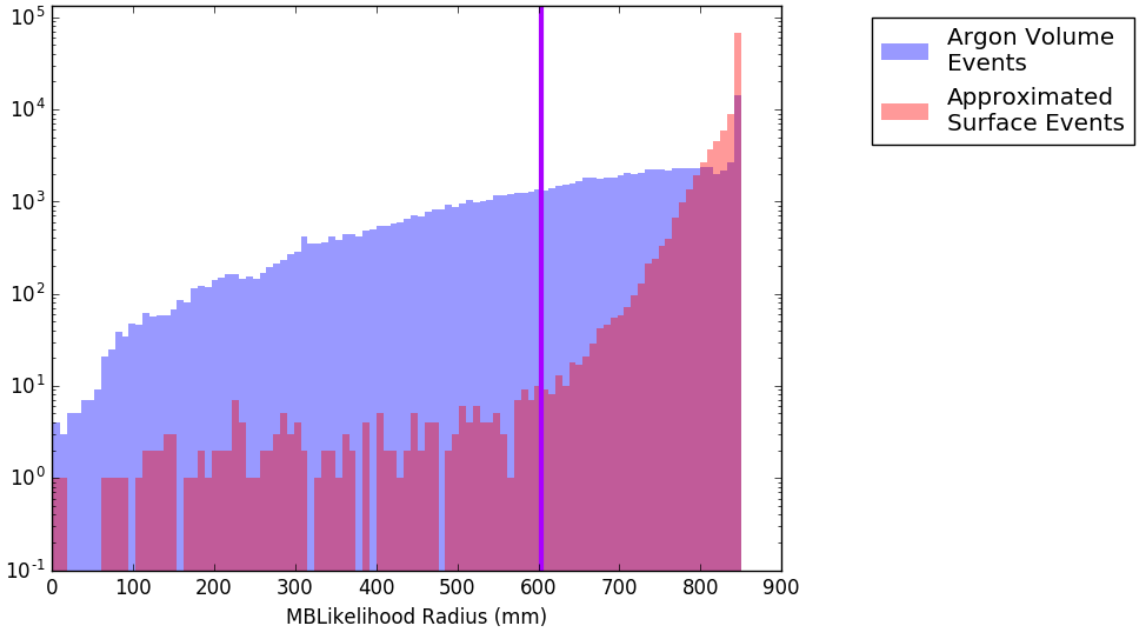

Figure 4.1: Histogram of MBLikelihood reconstructed radius for two populations of simulated data. One is uniformly distributed in the detector (blue), and the other is distributed in a shell 0.5 cm from the argon-acrylic surface (red).

4.5 Training

For training, each algorithm was fit to 50,000 events classified as surface, and 50,000 classified as fiducial. The optimization for each model was performed using the *sci-kit learn* Python package (Pedregosa et al. 2011). It has numerous classification models, some of which are described in Section 3.2. As a pre-processing step, the vector of 255 PMT charges were normalized. This removes most information about the total

charge of the event, but preserves topological information. There are three main reasons for this.

One is that it allows linear classifiers such as the logistic regression and linear kernel Support Vector Machine to have a chance of separating the two classes. To see why this is the case, imagine two PMTs at opposite sides of the detector. An event in the centre of the detector will generally evenly distribute the charge between the PMTs, while an event that is close to the surface will disproportionately deposit light in one PMT. As can be seen in Figure 4.2, the surface events cannot be linearly separated from the liquid argon events until the charge vectors are normalized to length 1. Based on this conceptualization of the normalization in 255D it is expected that the linear Machine Learning algorithms will be able to define a fiducial volume. The 2D example in Figure 4.2 may not be a perfect conceptualization for the 255 dimensions of the real detector and so non-linear models have also been fit to the data to see if they can improve performance by considering a larger set of decision boundaries.

The second reason to normalize the charge vector is to allow the algorithms to meaningfully make predictions across a wider range of charges without the need to simulate events at each energy regime. Topologically, the difference between a surface event and a fiducial event should take the same form at all energy scales. This is only true as long as the response of the PMTs is linear, which should be the case well beyond the energy ROI. Another effect which could cause a topological difference in energy is the signal to noise ratio, which is significant in the energy ROI. The energy dependence of each algorithm is explored briefly in Chapter 7.

The third reason is that it is desirable to restrict the position reconstruction

analysis to only eliminate events based on position information. The DEAP-3600 collaboration may want to consider different ranges of energy ROIs, which would not be possible if the position reconstruction analysis also eliminates events based on energy. While position reconstruction accuracy should be taken into account when determining the energy ROI, it is desirable to have control of this cut separately.

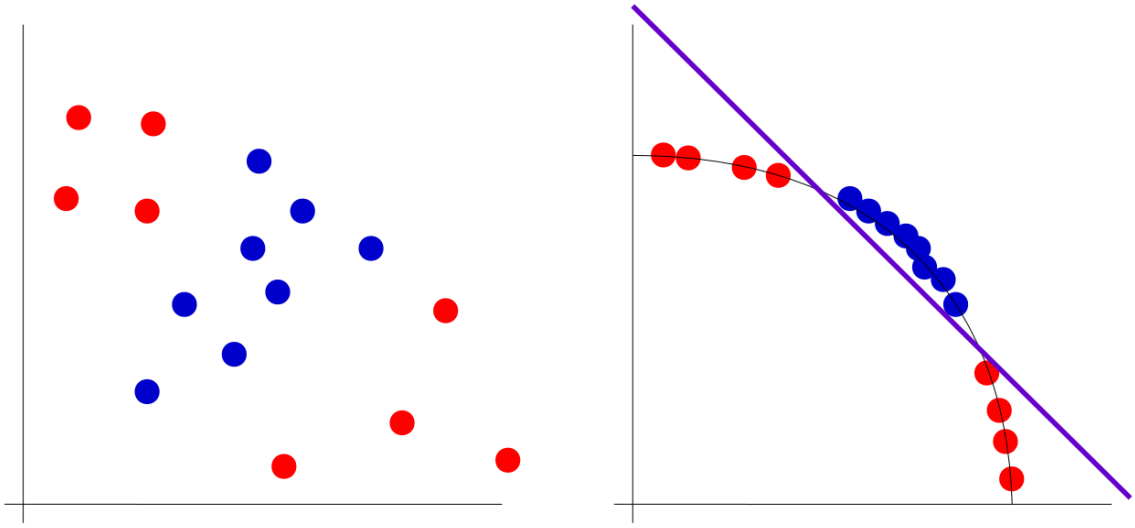

Figure 4.2: Cartoon representation of the reasoning behind normalization of PMT charge vector. The two axes represent the charge recorded in PMTs opposite to each other. Surface events are red, fiducial are blue. They are not linearly separable until after normalization.

Nmaxpe is computed as the fraction of light in the N brightest PMTs, up to 5. It is stored as a vector of length 5, with each element in the range $(0, 1)$. As a second preprocessing step, the *Nmaxpe* vector was included along with the 255 normalized PMT charges. Events that occur near the surface should have higher Nmaxpe values than events from the fiducial volume as more charge is deposited in a cluster of PMTs. As discussed in Section 3.2.2 this preprocessing step does not give any new information to the Machine Learning algorithms, but it does present them with the

information in a useful format.

4.6 Individual Performance

After training each algorithm with different hyper-parameters, they are tested on a new dataset. This validation set is used to select the best performing combination of hyper-parameters. Selecting hyper-parameters is simply another form of fitting, although a somewhat coarse version of it. A final set of new data is used, called the testing set, to measure the performance of the best version of each algorithm. The metric used to select the best algorithm is surface event leakage, such that the argon volume acceptance is 1 tonne. This is consistent with how MBLikelihood was assessed in Section 4.4. The performance of the best version of each Machine Learning algorithm is plotted in Figure 4.3 with MBLikelihood (MBR) included for reference.

The Naive Bayes learner performed the worst by a large margin which indicates that the information from the PMTs is highly correlated. As discussed in Section 3.2.5, the core assumption of the Naive Bayes algorithm is that each variable is independent. It should be expected that this would be the case as a single PMT charge does not provide much information, it is only the relative charges that are useful. The other algorithms perform nearly identically. This is unexpected due to the large difference in power of expression between some of the algorithms. The strong performance of linear models like the logistic regression and linear Support Vector Machine indicate that the PMT charge normalization does indeed structure the information in a useful way. The best performing model is MBLikelihood, which has just over half the surface leakage of the best Machine Learning algorithm. It is impressive that a linear model using only 261 coefficients is capable of coming so close to the

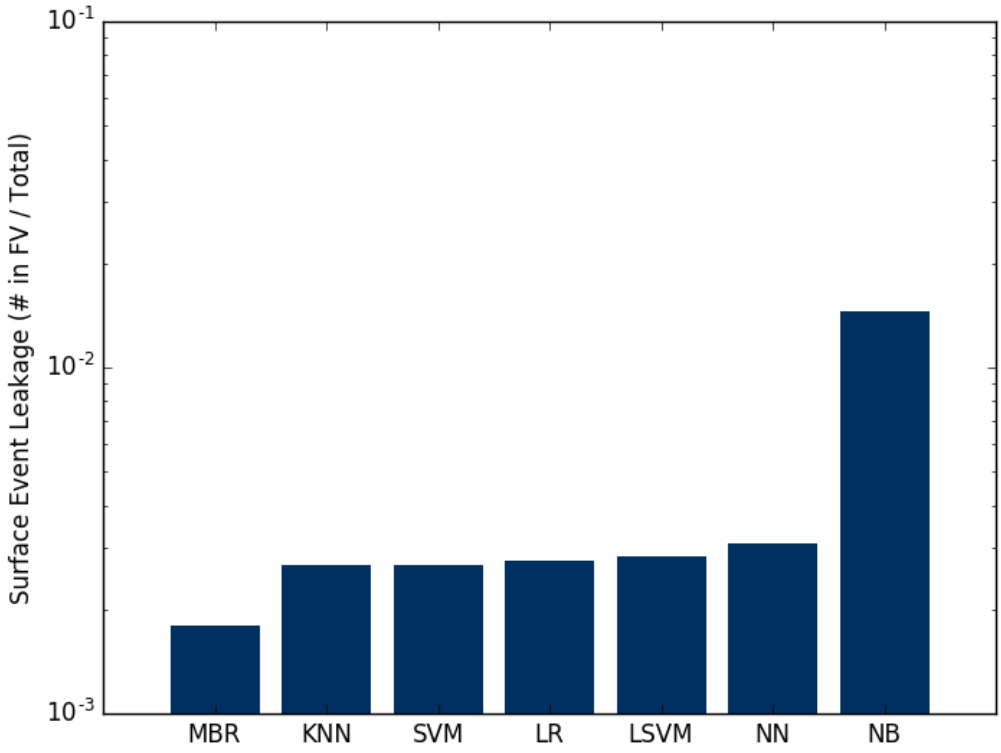

Figure 4.3: Leakage of each algorithm independently. There is a strong dependence on performance based on which algorithm is chosen.

performance of the much more complicated MBLikelihood algorithm. With enough data the K-Nearest Neighbours algorithm will always be one of the best performing models; the fact that it does so poorly suggests that it could take advantage of more data. However, with a total of 100,000 sample events used for training it is by far the slowest algorithm and already not feasible to run in live analysis (discussed in Section 7.5).

4.7 Boosted Position Reconstruction

Based on the results from Figure 4.3 MBLikelihood is still the best choice for surface event classification. However, it is possible that the Machine Learning algorithms are not identifying the same aspects of surface events that MBLikelihood does, and working together they might be able to do better than either one individually. This is accomplished by cutting events using an MBLikelihood threshold and a Machine Learning algorithm threshold. With two cuts there are an infinite number of combinations that will retain a 1 tonne fiducial volume, but by sampling the combinations it is easy to choose the combination that has the lowest surface event leakage. Figure 4.4 shows that the leakage is reduced by combining the algorithms.

All of the algorithms were able to reduce the MBLikelihood surface event leakage. Interestingly, the Naive Bayes algorithm was also able to boost MBLikelihood despite doing very poorly on its own. For the best performing models this amounted to an order of magnitude reduction in the number of surface events entering the fiducial volume. Such a substantial improvement is very surprising, it is reasonable to be skeptical. The goal of this analysis is to reduce the surface event leakage in the real detector. If the Machine Learning algorithms have simply learned some peculiarity of the simulator then they will not be effective at the goal. In Chapter 5 the same algorithms will be trained using real data to see if they can still achieve the same performance.

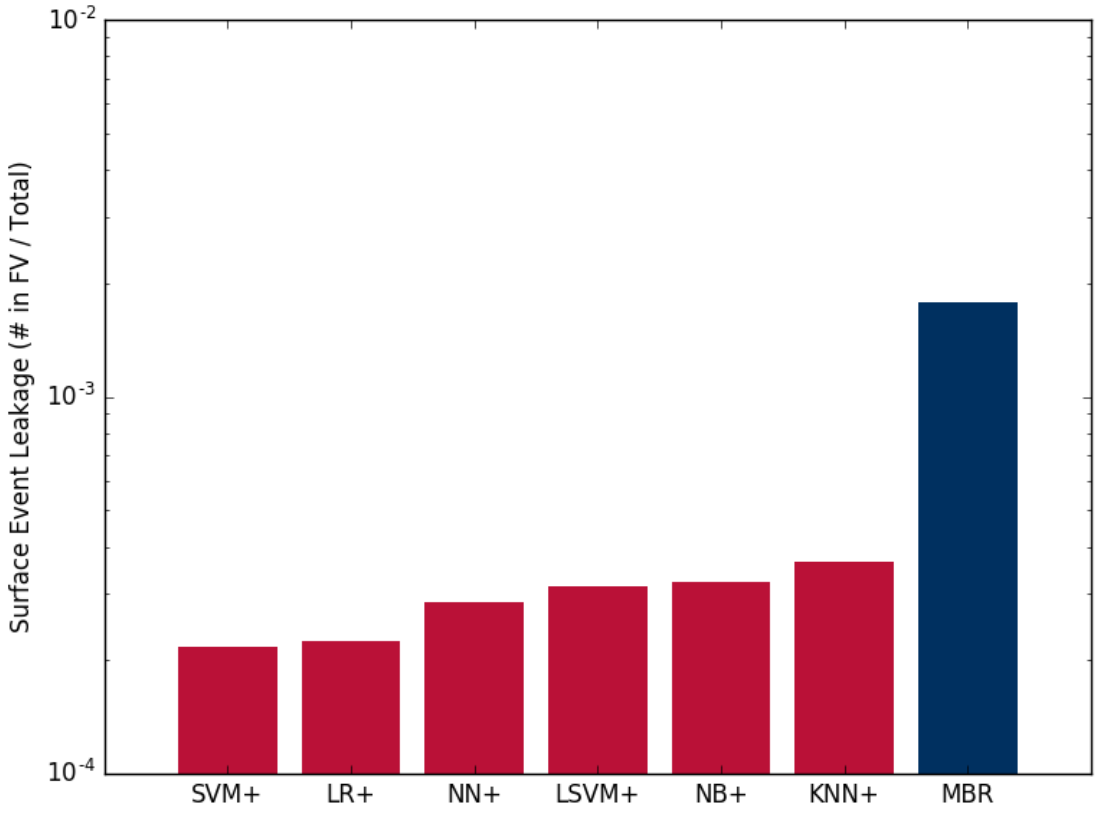

Figure 4.4: Leakage of each algorithm with MBLikelihood boosted. MBLikelihood individual performance included for comparison.

Chapter 5

Trained on Real Data

The Machine Learning algorithms have performed very well when trained and tested on Monte-Carlo simulaiton data. A natural concern for both MBLikelihood and the Machine Learning algorithms is that having trained them with simulated data, they have picked up on some aspect specific to the simulator. It would be preferable to train the algorithms with detector data. Unfortunately, detector data does not come with the Monte-Carlo simulation truth information that was used for training in previous chapter. In this Chapter it is demonstrated that it is possible to train the Machine Learning algorithms with real data thanks to the classification paradigm.

5.1 ^{39}Ar Beta Decay Events

The isotope ^{39}Ar is found in argon with an activity of roughly 1 Bq/kg in beta decays (endpoint of 565 keV) (Benetti et al. 2007). In a detector with several tonnes of liquid argon, this results in a kilohertz rate for this signal. This can be reasonably efficiently discriminated from a WIMP signal using pulse shape discrimination (Boulay and Hime 2006), but is still useful for calibrating various aspects of the analysis. For example, the light yield of the detector can be calibrated using the ^{39}Ar signal. While

the beta events produce scintillation with a different pulse shape in time than alphas or WIMP-like events, they are all isotropic sources of light and so should deposit light in the PMTs in a similar way. The uniform distribution of ^{39}Ar in the detector has been used to validate MBLikelihood and other position reconstruction algorithms. Now it may be possible to train the Machine Learning algorithms with these data.

5.2 ^{39}Ar Event Selection

^{39}Ar beta events are the dominant signal in DEAP-3600 by rate. It is possible to get a relatively clean sample just by selecting the energy ROI, however further cuts were applied to remove the next leading background sources. First, a cut in fprompt was used to eliminate events outside the range [0.2, 0.4], this eliminates many pileup ^{39}Ar beta events and almost any nuclear recoil event. Fmaxpe was required to be less than 0.4 to eliminate Cherenkov events. The calcut and dtmtrigsrc (a quality flag from the event trigger module) variables were used to remove cases of high event rate or timing mismatches in the trigger modules. The neck veto cut was used to eliminate events that occurred within the neck of the detector. A cut was used to remove any cases with multiple events in the same timing window. Any events which occurred within 20 μs of another were also removed, so that late light did not contaminate the signal. Finally, for the events used as argon volume events, it was required that their MBLikelihood z-axis value was below the fill level by 5 cm.

5.3 Training with Real Data

The exact location of any individual ^{39}Ar event is not known and so it is not possible to use this for training the position reconstruction algorithms. This is where an advantage of position classification becomes apparent. A simple position reconstruction method used by the collaboration is called the Centroid algorithm. It computes a 3D position for each event by taking the weighted average of the position of each PMT where the weight is the amount of charge registered by that PMT. While the Centroid algorithm is not nearly accurate enough to be used to train a position reconstruction algorithm such as MBLikelihood, it can reliably label very low radius events and very high radius events. There will be some contamination of the datasets, but the dominant signal in the sample will be correct labels for real detector data. This means that ^{39}Ar data can be used to train all of the Machine Learning algorithms discussed in this paper. Figure 5.1 shows the MBLikelihood radius versus the Centroid radius. It shows that they are in general agreement, with high Centroid radius events consistently also labelled as high radius by MBLikelihood.

Using ^{39}Ar data, each Machine Learning algorithm was trained using the same procedure as was done with the simulated data except for the source of the training data. With 50,000 surface and 50,000 fiducial volume events labelled by Centroid, each algorithm optimized a decision plane to divide these datasets. To select the best hyper-parameters, simulated data was used as described in Section 4.2. The simulated data is still the best reference for comparing surface event leakages and so is used for hyper-parameter selection and testing. It was possible that the Machine Learning algorithms only learned to reproduce Centroid; the coefficients for the linear algorithms (LR, LSVM) were checked and found to have learned, based on the data,

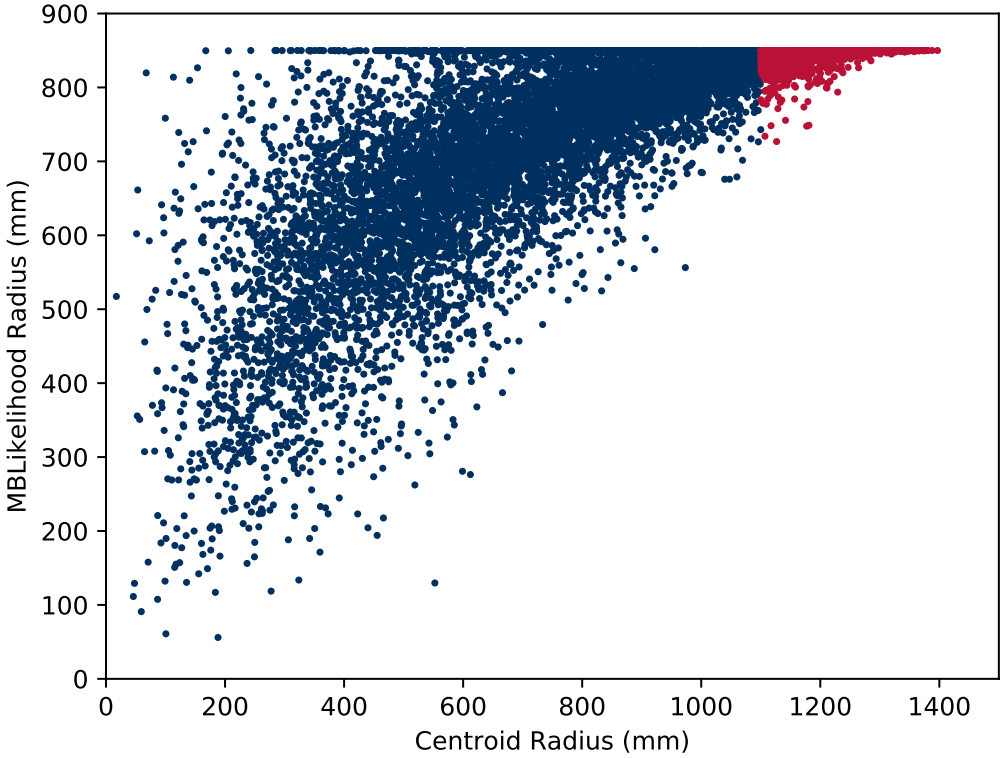

Figure 5.1: MBLikelihood radius vs Centroid radius with surface (red) and argon volume (blue).

that the detector is in a partial fill configuration. The coefficients for the PMTs above the fill level had significantly shifted values compared to those below the fill level. This is not encoded in the Centroid algorithm and so represents extra information that the Machine Learning algorithms learned.

5.4 Individual Performance

After the Machine Learning algorithms were trained using detector data and validated using simulated data, they were tested with a new sample of simulated data. The

resulting leakage for each model on its own can be seen in Figure 5.2. While the results are not identical for each algorithm, it is clear that the Machine Learning algorithms are still able to achieve similar individual performance as in Section 4.6. Naive Bayes is still the worst performing, with the others again achieving similar leakage values. As with the simulated data, MBLikelihood has the best performance at roughly half the leakage of any other algorithm. The fact that the algorithms perform similarly regardless of which dataset is used to train them is reassuring that the Monte-Carlo simulation is behaving as expected.

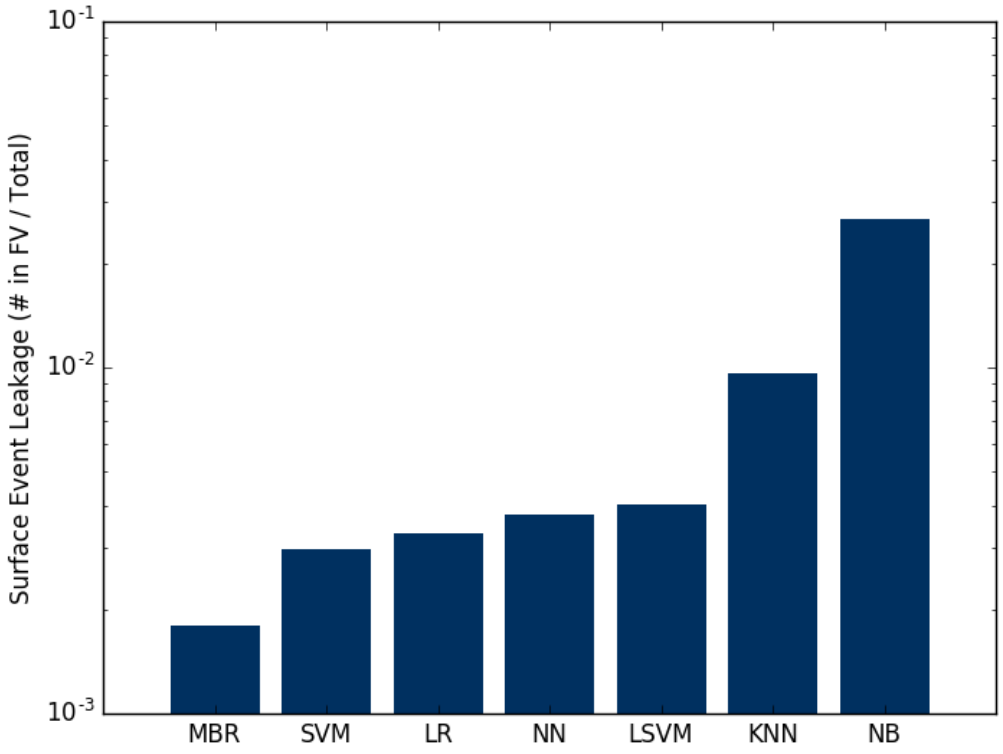

Figure 5.2: Leakage of each algorithm independently. There is a strong dependence on performance based on which algorithm is chosen.

5.5 Boosted Performance

The performance of the Machine Learning algorithms was similar when trained with detector data to the performance with simulated data. This is reassuring and so it can be expected that they will still be able to boost the MBLikelihood surface event leakage. Figure 5.3 shows the boosted surface event leakage for each algorithm. The roughly order of magnitude reduction in surface event leakage is maintained with the new algorithms. This is very reassuring as it was possible that in Chapter 4 the algorithms had simply learned to exploit a non-physical aspect of the simulator. As they have been trained using real data, the fact that the boost remains is strong evidence that they will be able to help MBLikelihood reject surface events in the real data as well.

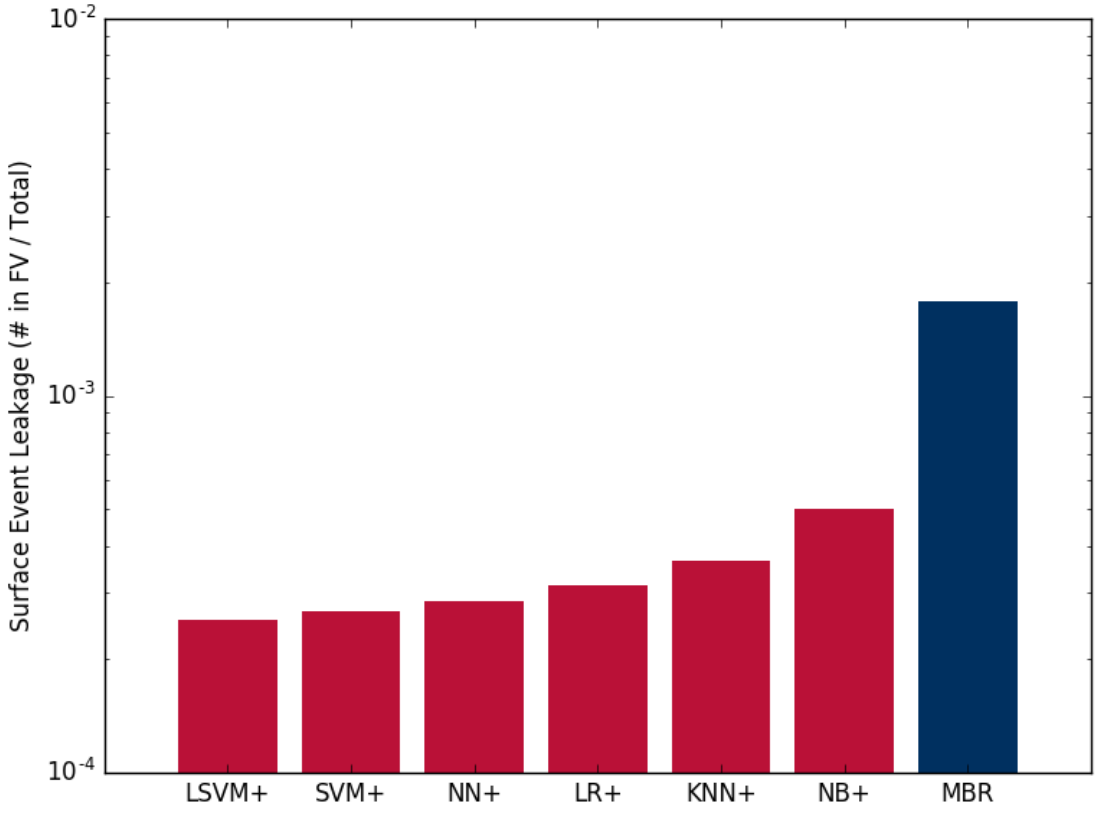

Figure 5.3: Leakage of each algorithm with MBLikelihood boosted. Shown are the Machine Learning algorithms combined with MBLikelihood and the surface event leakage of MBLikelihood on its own.

Chapter 6

Tests with Alternate Datasets

6.1 Modified Optical Model

In the simulator it is possible to modify optical properties of the detector. This can be used to test how stable the algorithms are. If small changes to the detector cause large changes in the performance of the Machine Learning algorithms, while not causing the same changes in MBLikelihood then that would be cause for concern. As the surface event leakage cannot be accurately measured with real data, it is important to be able to use simulated data to characterize the Machine Learning performance. This section demonstrates that for several reasonable changes to the optical model, the Machine Learning algorithms track the same changes in performance that MBLikelihood experiences, validating their robustness.

As changes in the optical model are intended to represent the range of uncertainty in those parameters when translating between the simulator and the real detector, it would not make sense to optimize the surface event leakage in the same way as in Chapter 4. To boost the algorithms in the modified optical model, instead of selecting the combination of thresholds that gets the minimum surface event leakage

(described in Section 4.7) one of the thresholds is fixed at the value used for the default optical model boosting. With one threshold fixed, the other is adjusted until 1 tonne of fiducial volume is obtained. Because of the ^{39}Ar signal, it will always be possible to calibrate a 1 tonne fiducial volume in real data, but optimizing surface event leakage may not be possible.

6.1.1 TPB Thickness

A layer of wavelength shifter (TPB) between the acrylic and argon volume was deposited in the final stages of the detector construction. This process is very complicated and is discussed in detail in Broerman et al. (2017). The deposition process was designed to create a uniform layer that is 3 µm thick, however there is some uncertainty on the value. Changing the thickness of this layer has a large impact on the topology of an event. Light that gets wavelength shifted is scattered isotropically, which tends to erase some of the position information. The wavelength shifted light can be scattered multiple times if it does not enter a LG, further erasing the information about the initial position of the event in the detector. In the limit as this layer gets very large, DEAP-3600 becomes an integrating sphere and no position information can be gained by observing the PMT charges. To study the effect of different TPB thicknesses, these values were chosen: 2 µm, 2.4 µm, 2.7 µm, 3.3 µm, 3.6 µm , and 4 µm. They were chosen as the uncertainty in the TPB thickness is ± 0.9 µm and so they extend to roughly $1\,\sigma$. Figure 6.1 shows the performance of each algorithm at the different TPB thicknesses. It is clear that the TPB thickness has a large impact on the performance of each algorithm, including MBLikelihood. The hierarchy of performance is always maintained; MBLikelihood on its own performs worse than the

boosted analysis.

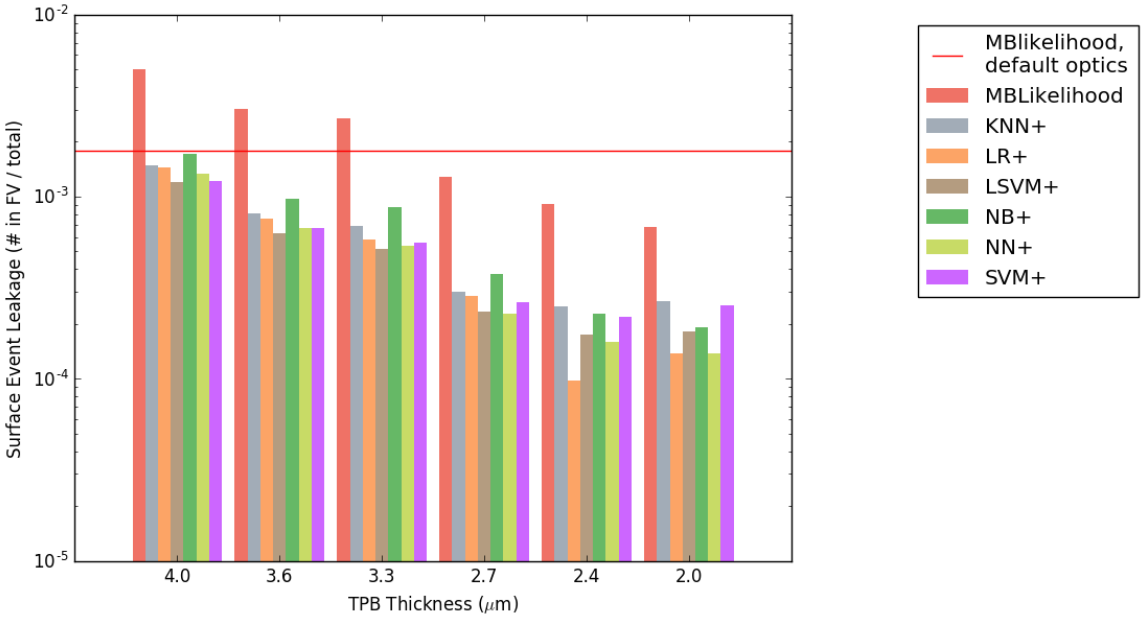

Figure 6.1: Surface event leakage for each algorithm with different TPB thicknesses. X-axis shows 6 different TPB thicknesses and a bar for each algorithm's performance.

6.1.2 Fill Level

DEAP-3600 was designed to be completely filled with liquid argon, but due to a complication in the cool down process it was only filled to about 55 cm above the equator. The fill level in the detector can be measured with an uncertainty of about 3 cm using multiple methods. Having a liquid-gas boundary changes the optics of the detector asymmetrically. It has been observed that the liquid-gas boundary is not properly reconstructed by MBLikelihood as reflection and refraction at the surface break assumptions about the spherical symmetry of the detector. The result is a bent boundary in the MBLikelihood reconstructed position instead of the expected

flat/constant z value. However, it was not known before this analysis if the liquid-gas boundary had an impact on the radial distribution. In the simulator one can try varying the fill level to observe if this has an effect on the surface event rejection efficiency. To study the effect of different fill levels, these values were chosen: 51 cm, 53 cm, 54 cm, 56 cm, 57 cm , and 59 cm. These were chosen as the uncertainty in the fill level is 3 cm and so represent roughly a $1\,\sigma$ variation. Figure 6.2 shows the performance of each algorithm at the different fill levels. It seems that these small deviations have no effect on the performance of each Machine Learning algorithm.

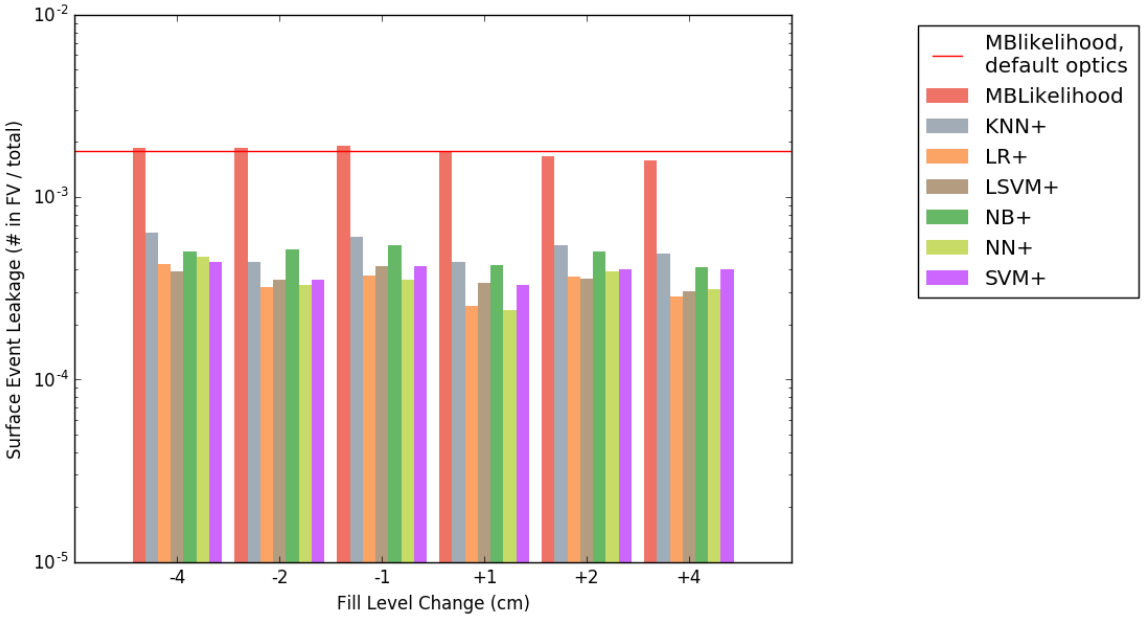

Figure 6.2: Surface event leakage for each algorithm with different fill levels in the simulator. X-axis shows 6 different fill levels and a bar for each algorithm's performance.

6.1.3 Scattering

A Detailed analysis has not yet been performed in-situ on the scattering length of argon in the DEAP-3600 detector. In RAT the scattering length is defined as a

piecewise function of wavelength (Caldwell 2014). Changing the scattering length in argon has an effect analogous to that of changing the TPB thickness. With a shorter scattering length there is more information lost while light is travelling through the argon volume. In the limit as the scattering length becomes very small, DEAP-3600 would again lose all position information for each event. However, this optical parameter is different than the TPB thickness as it affects light as it travels though liquid argon, instead of only at the liquid-acrylic interface.

In RAT, the scattering length is wavelength dependent and so it is recorded internally for multiple wavelengths and interpolated to get a full wavelength range. As an optical parameter, the scattering length is controlled by a scaling factor that is multiplied with all the recorded values. To study the effect of different scaling lengths, these values were chosen for the scale factor: 5, 3, 2, 0.5, 0.33 , and 0.2. Studies to determine the scattering length within the detector are ongoing, these values were chosen to represent a large uncertainty. Figure 6.3 shows the performance of each algorithm at the different scattering length adjustments. The large variation in performance is similar in magnitude to the variations seen from changing the TPB thickness. The important result of this test is that the hierarchy is maintained between MBLikelihood and the boosted analysis.

6.1.4 Absorption Fraction

As with the scattering length, there has not yet been an in-situ validation of the absorption length in liquid argon, for scintillation light it is assumed that there is no absorption. Argon scintillates by creating dimers and then decaying back to individual atoms. As the energy level structure of the dimers and atoms are sufficiently different,

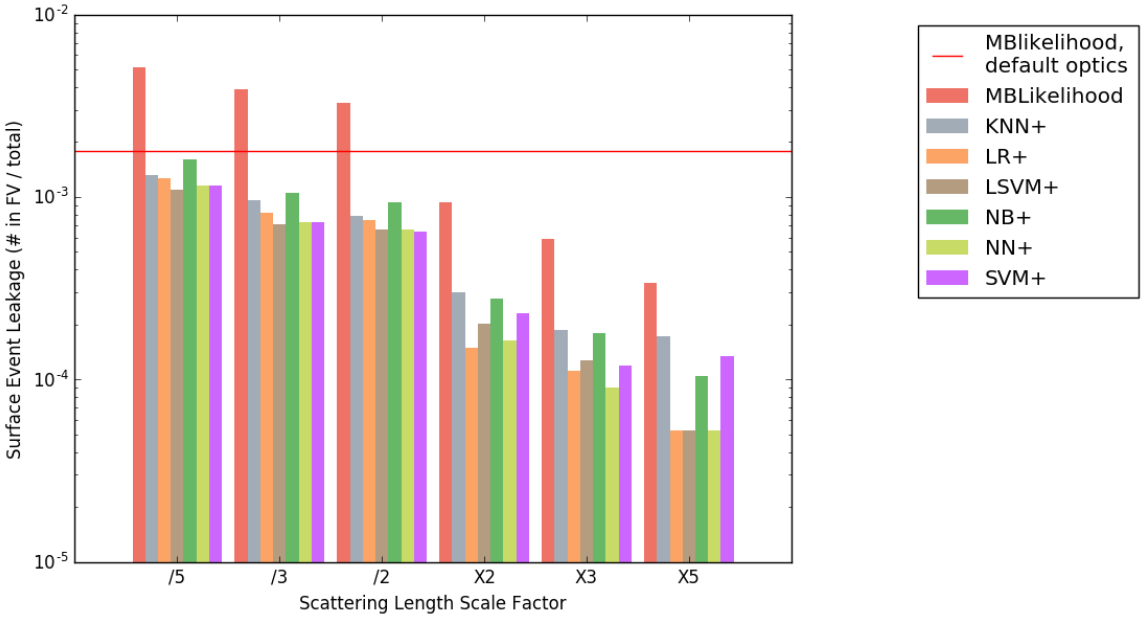

Figure 6.3: Surface event leakage for each algorithm with different scattering length scale factors in the optical model. X-axis shows 6 different scattering length factors and a bar for each algorithm's performance.

scintillation light should not be absorbed as it travels in the argon volume (DEAP-3600 Collaboration et al. 2017). However, there are possible absorption mechanisms, for example dimers may form in the liquid argon without external excitation and those could absorb scintillation light. Increasing the amount of absorption will decrease the amount of information that one has for a given event. For surface events, the absorption length will have a non-uniform effect on the light since some only has to travel a short distance to reach the TPB, while light going to the opposite side of the detector has to travel nearly the entire diameter of the vessel. One could imagine that this non-uniform effect would be a benefit for position resolution as it would contribute to the concentration of charge that is characteristic of surface events. At some point the loss of information would begin to dominate and position resolution

would get worse. It is not clear how the two effects will balance out, and so it is worthwhile to investigate.

In RAT, the absorption length is not recorded directly, but instead as a fraction of scattering events. The absorption fraction determines the number of scattering events that, instead of scattering, simply eliminate the photon from the simulation. Changes to the absorption fraction change both the scattering length and the absorption length. To study the effect of different absorption lengths, these values were chosen for the scale factor: 0.99, 0.98, 0.97, 0.96, 0.95, 0.94. These fractions represent having very little absorption, yet likely represent more than would ever be reasonable in the detector. Figure 6.4 shows the performance of each algorithm at the different absorption fraction values. It is clear that the absorption length does not have a large impact on performance for any of the algorithms in the range tested.

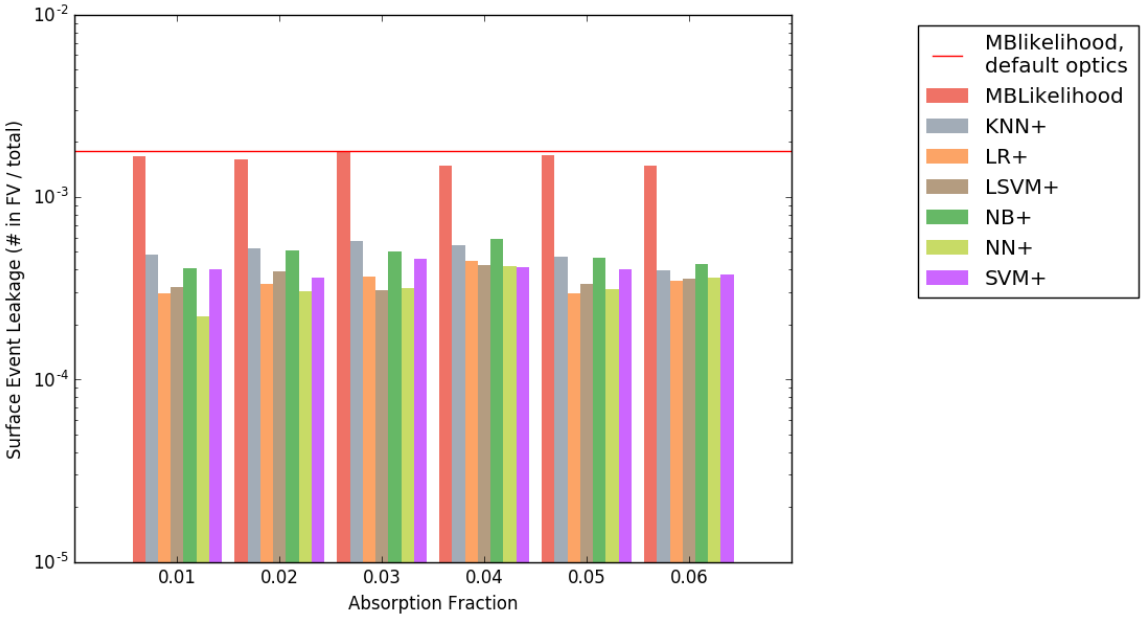

Figure 6.4: Surface event leakage for each algorithm with different absorption length scale factors in the optical model. X-axis shows 6 different absorption length factors and a bar for each algorithm's performance.

6.2 Americium Beryllium Neutron Events

To approximate the signal that a WIMP would create, a neutron source was placed next to the DEAP-3600 detector. The FBs and LGs perform well as neutron shields for the argon volume, however some neutrons still made it into the detector. The resulting events can be used to test various aspects of the analysis. The neutrons do not interact uniformly in the detector, instead their distribution is exponential in radius and the majority interact on the side of the detector that has the source. Neutrons may also scatter and interact multiple times, as long as the interactions are close to each other this should not impact position reconstruction analysis. Some may scatter a large distance in the detector before interacting again, these events are not useful for position reconstruction analysis, but are difficult to discriminate against. This makes the AmBe source a poor test for the performance of each algorithm, but can be useful as a check that they treat events similarly. For this thesis it will be useful to look at these events simply to confirm that the Machine Learning algorithms do not behave pathologically on a true WIMP-like signal.

In total 6633 events from the AmBe source passed a basic version of the ROI without a fiducial volume. MBLikelihood accepted a fraction of 0.073 of these events with a radius set for 1 tonne of fiducial volume using real ^{39}Ar events. The Machine Learning algorithms had similar acceptances of the neutron events ranging between 0.065 and 0.081 indicating that they all have the same acceptance to within 1σ. Similarly, the boosted analysis stayed within 1σ acceptance of MBLikelihood for these events, as shown in Figure 6.5. This suggests a consistent representation of the fiducial volume for MBLikelihood and the Machine Learning algorithms in the energy ROI. It is not possible to determine if the boosted analysis is improving surface event rejection

with this dataset. The primary goal of this section was simply to demonstrate, using real data, that MBLikelihood and the boosted analysis should have similar WIMP acceptances.

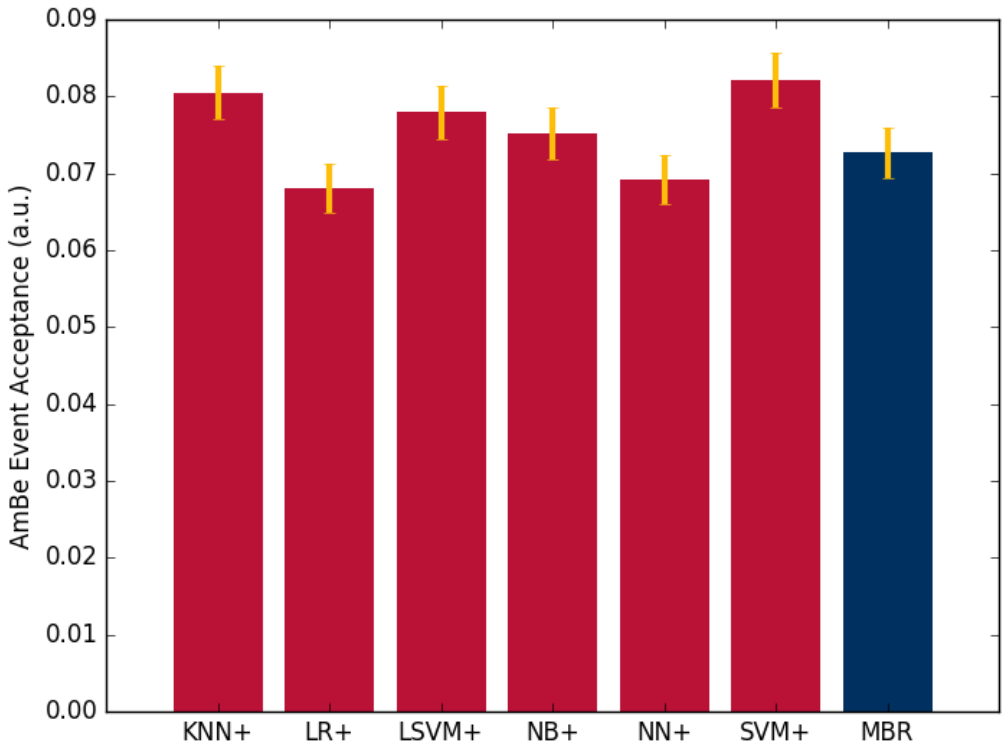

Figure 6.5: Acceptance of neutron events into a 1 tonne fiducial volume for boosted Machine Learning algorithms and MBLikelihood.

6.3 ^{210}Po Alpha Events

During construction of the detector small amounts of radon diffused into the acrylic vessel. This element has a decay chain that ends in stable ^{206}Pb. A long lived intermediate step is ^{210}Po which decays via an alpha particle. The alpha particle has

a high energy by the standards of DEAP-3600 and most of the time is far outside of the energy region of interest. However, a small number of them will travel through the TPB layer losing energy until they reach liquid argon and can mimic a WIMP signal, falling into the energy ROI of $(120, 240)$ QPE. As the alpha particles have very short mean free paths in liquid argon, all of the events that appear in the window can be assumed to have come from the surface of the detector. These are to be eliminated by choosing a fiducial volume in the argon volume.

Unfortunately it is not possible to consistently identify ^{210}Po events at energies corresponding the energy ROI, only the high energy ones can be identified easily. The majority of the ^{210}Po events can be identified within a window of high QPE $(10000, 25000)$ and fprompt $(0.6, 0.8)$. DEAP-3600 has been carefully designed and built to have a low background rate; this applies to ^{210}Po as well. The rate of ^{210}Po alphas in DEAP-3600 is approximately 1.6 mBq (Lehnert 2018) and so only a small collection (compared to ^{39}Ar betas) of them can be found in the data set, totalling roughly 6000. Because of their high energy it would be expected that this dataset of ^{210}Po events should be easily distinguished from fiducial volume events due to the high signal to noise ratio.

Analysis thus far has been concerned with events in the energy ROI, or QPE values in the range $(120, 240)$. To analyze the leakage of the ^{210}Po events initially, a dataset of ^{39}Ar events was used to define the fiducial volume with QPEs in the energy ROI. This meant that the argon volume events and surface events in this analysis had vastly different energy scales. When the MBLikelihood surface event leakage was computed for this dataset, the ^{210}Po event leakage that resulted was 5.4 %. This is much larger than expected, implying that perhaps the low energy ^{39}Ar events do not

define the fiducial volume properly for high energy events. The ^{39}Ar beta spectrum ends at roughly 4000 QPE and so a collection of events near this endpoint was used. The resulting ^{210}Po event leakage reduced to 3.8 %, which is still far larger than was expected.

Figure 6.6 shows the leakage of ^{210}Po events into a fiducial volume defined by real ^{39}Ar events from the high energy dataset for each algorithm boosted with MBLikelihood and including MBLikelihood for comparison. Interestingly, for the ^{210}Po events the Machine Learning algorithms provide very little benefit to MBLikelihood if at all. In fact, the benefit could quite easily be within error as the improvement is less than 3σ. It is reassuring that the leakage does not increase when using the machine learning algorithms on their own or in the boosted analysis. However, due to the dependence of position reconstruction at different energy scales it is not clear that this sample of ^{210}Po events makes any significant statement about the performance of the boosted analysis in the energy ROI.

These results indicate that efforts to eliminate the energy dependence of the machine learning analysis were not entirely successful for extreme differences in energy. There must be some topological difference between surface events in the range (120, 240) QPE and the range (10000, 25000) QPE. One possible explanation is that the low energy events are much more strongly affected by noise sources such as Poisson statistics, dark noise, and after-pulsing. At high energy the signal to noise ratio is very different; this would change the topology of high energy versus low energy events. It is also possible that other effects, such as clipping and PMT saturation are impacting the results (Lindner 2015). These effects are known to occur in this energy range and can change the amount of charge in a PMT, breaking assumptions that are made in

the low energy regime.

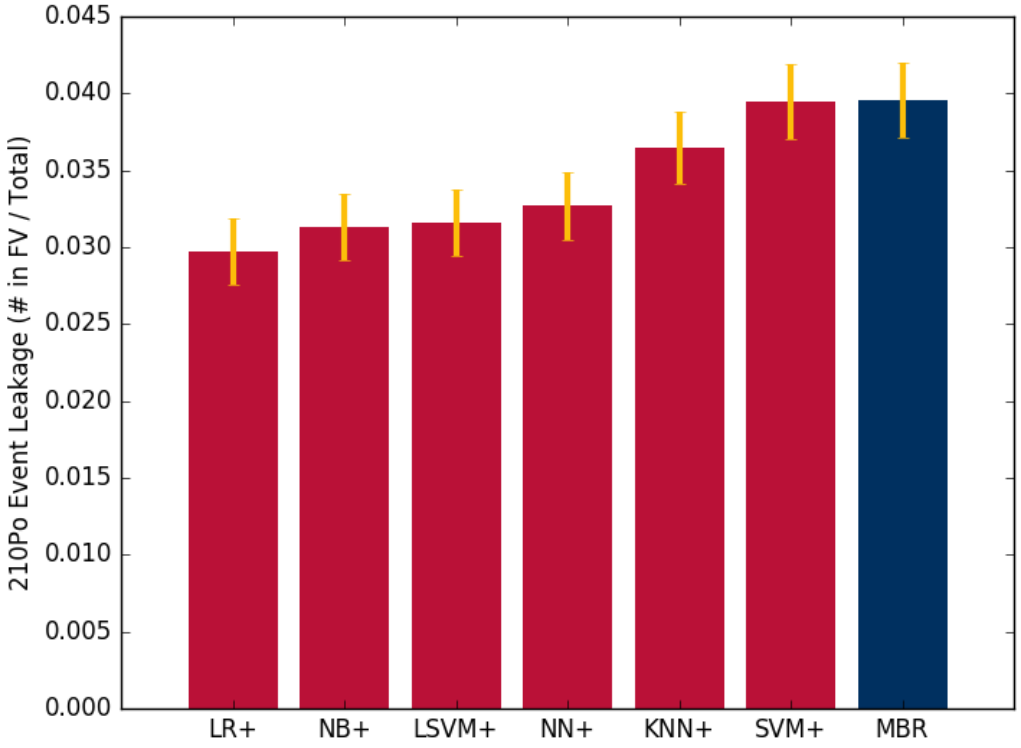

Figure 6.6: Leakage of ^{210}Po events at high energy into a fiducial volume defined to keep 1 tonne using real ^{39}Ar from the higher energy tail of the beta spectrum.

Chapter 7

Further Characterization

Fiducial volume event selection is a complicated problem, this chapter investigates a few of the questions left open by the analysis in earlier chapters. Chapter 4 and Chapter 5 focused solely on reducing the surface event leakage in the ROI, however that is not a complete story. With the ROI pushed to as low an energy as possible, Poisson statistics fluctuations can become a dominating factor in performance. It is interesting to determine how the performance of each algorithm changes as a function of energy within the ROI. Another aspect of the analysis is the size of the fiducial volume, it is possible that the relative performance of each algorithm is dependent on the size of the fiducial volume. Another brief analysis shows why it is only useful to use a single Machine Learning algorithm to boost with MBLikelihood instead of two or more. In the last section there is a discussion of the runtime for each algorithm, showing that some of the Machine Learning algorithms take far longer to perform a classification than the others.

7.1 Energy Dependence

The ROI for the DEAP-3600 dark matter search will ideally be pushed to the lowest energy possible. The number of photo-electrons generated in the PMTs from an expected WIMP signal will be of the order $\sim 100 - 250$, which is less than the number of PMTs. This means that the amount of information that an algorithm has to work with is very limited and errors are dominated by statistics. To observe this effect, surface event leakage is plotted as a function of the total number of photo-electrons measured in all PMTs. Figure 7.1 shows that for each algorithm the leakage increases consistently as the total charge decreases, including for MBLikelihood. This is as expected, and there is not much that can be done to mitigate this effect, it is purely determined by the light yield of the detector. It is worthwhile to observe these trends to make sure that the Machine Learning algorithms (and MBLikelihood) have not simply learned to discard all events with low energy, or any other energy dependent decision boundary. It was plausible before doing this analysis that the algorithms had been able to minimize leakage by rejecting all events below a certain energy, which are the most difficult to identify. It is undesirable for the position reconstruction algorithms to do this as the low energy threshold should be decoupled from the fiducial volume analysis. Based on Figure 7.1, it is clear that the energy dependence is consistent with MBLikelihood and so likely the decision boundary is not energy dependent, it is only being limited due to the statistical noise.

7.2 Differences in Simulated and Real Data

It is expected that the simulated data and the real data will be slightly different. The model of the DEAP-3600 detector is undergoing continuous improvement, especially

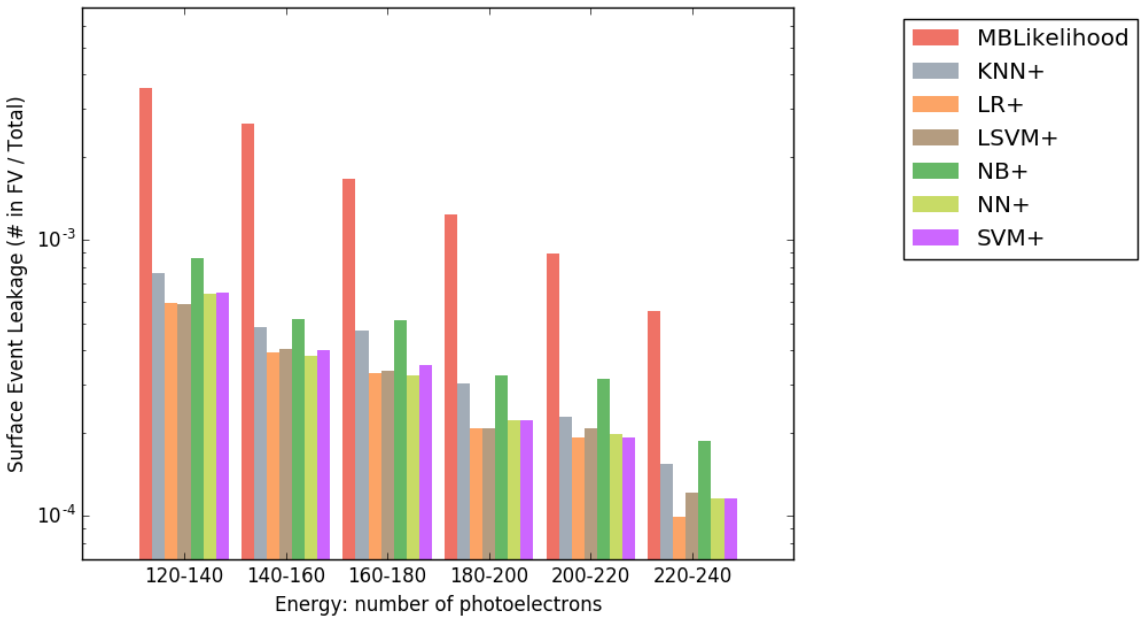

Figure 7.1: Surface event leakage as a function of the energy (QPE) of an event. Statistical noise is a major factor in surface event discrimination and so a measurable dependence on charge is seen for all algorithms in the ROI.

as the first data are being analyzed. It is therefore not surprising that the Machine Learning algorithms and MBLikelihood behave differently on the two datasets. The simplest way to investigate the difference is with data that are uniformly distributed in the detector. While almost any dataset can be achieved in Monte-Carlo simulation, the real data is more complicated, but ^{39}Ar beta decays provide a uniform signal that is easy to acquire in large numbers. Figure 7.2 shows a comparison of the output parameter for a few of the Machine Learning algorithms and MBLikelihood. There are a few main features to notice.

The first is that the difference is much more pronounced in the Machine Learning algorithms. This indicates that the Machine Learning algorithms are more sensitive than MBLikelihood to changes in the detector configuration, however Section 6.1

shows that they should be robust to small changes. By computing a fiducial volume threshold for each dataset, it is expected that discrimination power will still be preserved and relative performance in simulation will translate to relative performance on data. The differences that are seen suggest that it would be unwise to make claims about the absolute surface event leakage, but the statement that the boosted performance is superior to MBLikelihood on its own is considered robust. It is possible that the Machine Learning algorithms could be used to identify specific differences between the simulation and real data, however this is beyond the scope of this thesis.

The second striking feature is that when moving from data to simulation, the fiducial volume thresholds all move in the same direction. Specifically the fiducial volume threshold for data is higher than the threshold for simulation. This suggests, somewhat inconclusively, that each algorithm is picking up on the same difference between real and simulated data. The shift from data to simulation has events classified systematically more like fiducial volume events. This could indicate that the noise model in the simulator is in fact set too high, as this would cause events to appear systematically lower in radius. However, any strong statements about the noise model or other reasons for this systematic shift are outside the scope of this thesis.

The third feature that is apparent in Figure 7.2 is that the shape of the distributions for the Machine Learning algorithms are different. The simplest explanation for the different shapes is from the way each algorithm defines its decision boundary. For example, the Logistic Regression (LR) algorithm requires the output to be in the range $[0, 1]$ meaning that a bias in the events in either direction (surface or fiducial) would cause a pileup of events near the 0 or 1 boundary. This is only a problem with the histogram as it uses equal width bins, the output is continuous and

so there should always be a one-to-one mapping that would realign the distribution in whichever direction is desired.

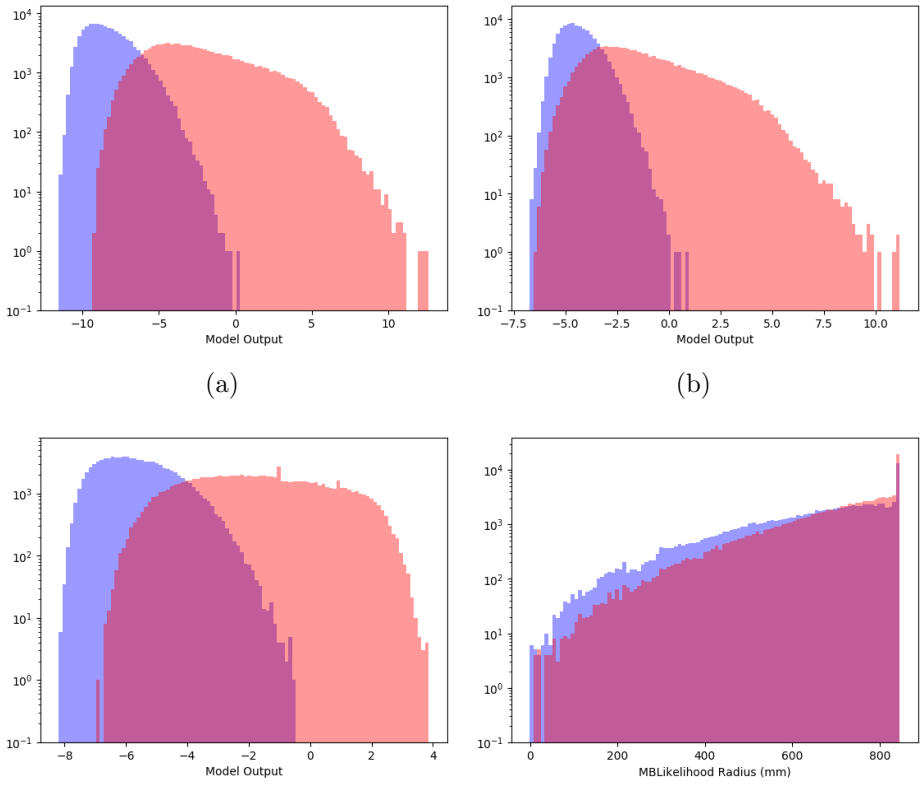

Figure 7.2: Comparison of the output parameter for Logistic Regression (a), Linear Support Vector Machine (b), Support Vector Machine (c), MBLikelihood (d) on real (red) and simulated (blue) data. Datasets are assumed to be uniformly distributed in the detector. For the sake of comparison, only the dot product output is used for LR as the $1/(1+e^{-\vec{a}\cdot\vec{b}})$ function does not allow real and simulated distributions to be visually compared.

7.3 Fiducial Volume Size

All analyses performed thus far have assumed a 1 tonne fiducial volume. With better surface event discrimination it may be possible to increase that. Figure 7.3 shows the

surface event leakage as a function of fiducial mass for simulated ^{40}Ar events using the default optical model. The figure shows that when using the boosted algorithms the surface event leakage increases exponentially with fiducial volume, while the MB-Likelihood algorithm has a slight curve in its surface event leakage as a function of fiducial volume. To match the surface event leakage of MBLikelihood with a 1 tonne fiducial volume the boosted analysis can accept 1.5 tonnes. This is an incredible increase in fiducial volume and could translate to significant increases in exposure, or reduced runtime for the experiment to achieve its physics goals. It is also interesting to note that the boosted algorithm has less of an impact as the fiducial mass gets very large. Above a fiducial volume of 2 tonnes there is little benefit at all from using the boosted analysis.

7.4 Multi-Boosting

After training the Machine Learning algorithms and using them individually to boost MBLikelihood it is natural to ask if the leakage can be further reduced by using all algorithms together. The answer is that there is no benefit from combining them. This is because they all learned roughly identical decision boundaries. A histogram matrix in Figure 7.4 shows that most of the Machine Learning algorithms have an output that is strongly related. Using a dataset of real ^{39}Ar beta events the output parameter for each algorithm was plotted against each of the other algorithms to create the off diagonal images. On the diagonal, a histogram is included which shows how the events are distributed for the corresponding algorithm. The different shapes in the scatter plot matrix are created by different definitions of distance from the decision plane. For example, the linear Support Vector Machine algorithm uses the

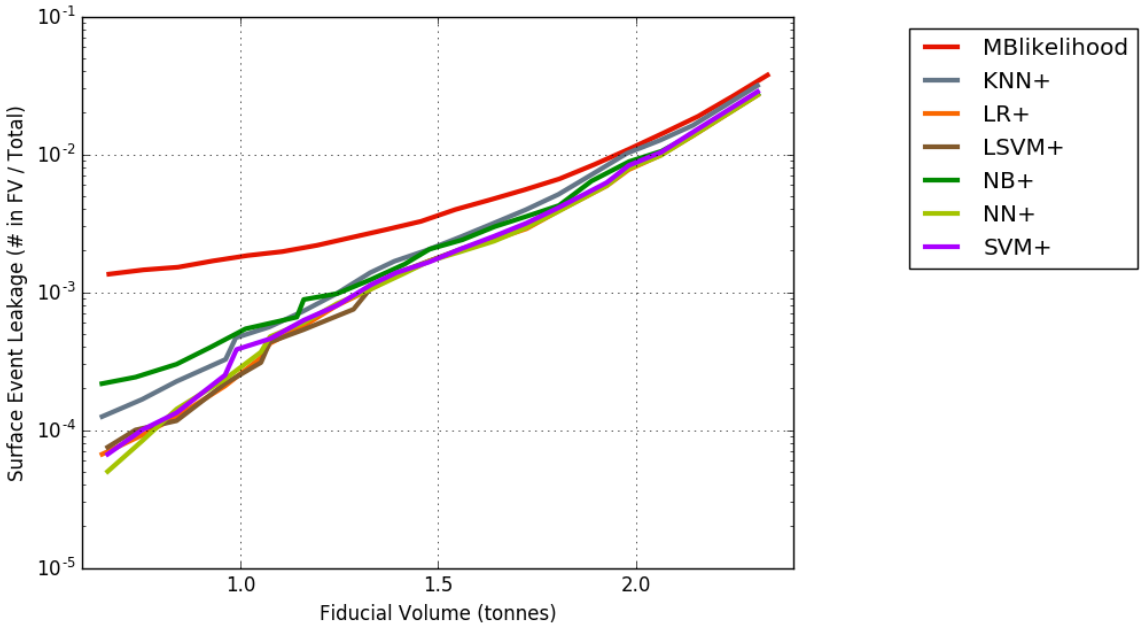

Figure 7.3: Surface event leakage as a function of chosen fiducial volume for MB-Likelihood and MBLikelihood boosted with the corresponding Machine Learning algorithm. The smaller the fiducial volume, the greater impact the boosted analysis has on surface event leakage. This can be seen in the narrowing of the gap between MBLikelihood and the boosted algorithms at higher fiducial volume.

Euclidean distance from the decision plane as the output while Logistic Regression uses the Euclidean distance mapped into the range $(0, 1)$. These are two of the best performing algorithms, and in Figure 7.4 they are extremely tightly correlated. There is, however, a significant amount of scatter between MBLikelihood and any of the Machine Learning algorithms. Thus, MBLikelihood is the only one to benefit from boosting. The Naive Bayes algorithm also shows lots of scatter with each of the other Machine Learning algorithms, however this is just due to its poor individual performance and not because of any extra information.

7.5 Runtime

While many of the algorithms seem to have similar performance, a major factor in their possible implementation is how fast they can make a classification. To be run on the live analysis that is performed as data are recorded from DEAP-3600 an algorithm must be very fast, preferably less than 0.01 seconds per event. This is due to the high frequency ^{39}Ar signal at over 3 kHz; the events are processed in parallel, but resource limitations still constrain runtimes. The value of 0.01 seconds per event is not a hard limit, but is roughly the time used by MBLikelihood to make a classification (on the hardware used for this analysis) and so it makes an appropriate comparison value. Figure 7.5 shows the average runtime of each algorithm (based on 100,000 events). For the fastest algorithms, such as the linear Support Vector Machine the runtime in RAT would likely be dominated by the time to read in the PMT charge information. The slowest algorithm, K-Nearest Neighbours, likely could not be run in the live analysis unless it was only applied to a subset of events such that it's slow runtime does not greatly impact the runtime.

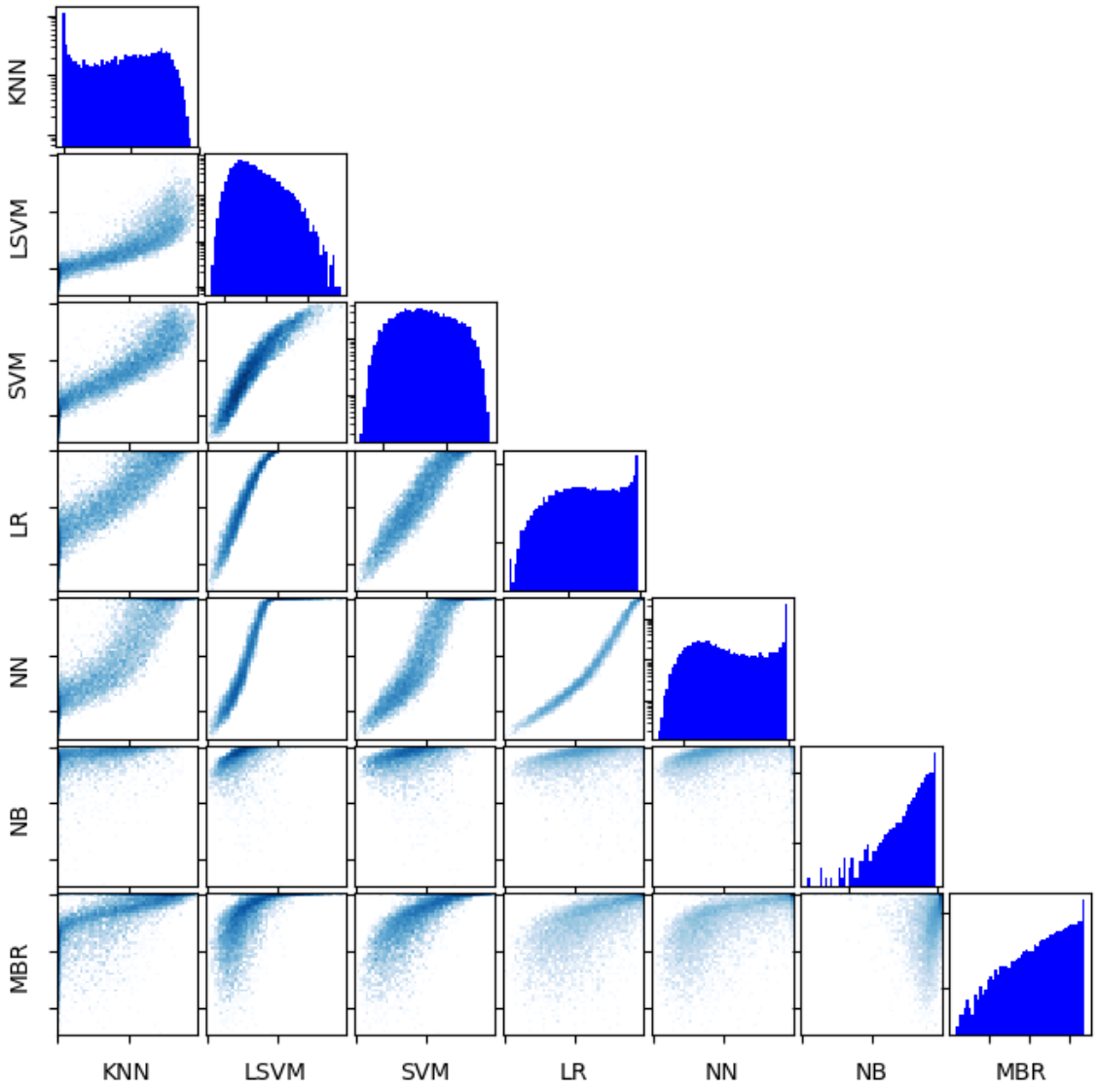

Figure 7.4: Scatter Plot Matrix showing the predictions for each algorithm on the same dataset. The diagonal has a histogram of the output for an algorithm, the off diagonal entries are a 2D histogram for the output of two algorithms. The bin counts for both the 1D and 2D histograms have been displayed in log scale.

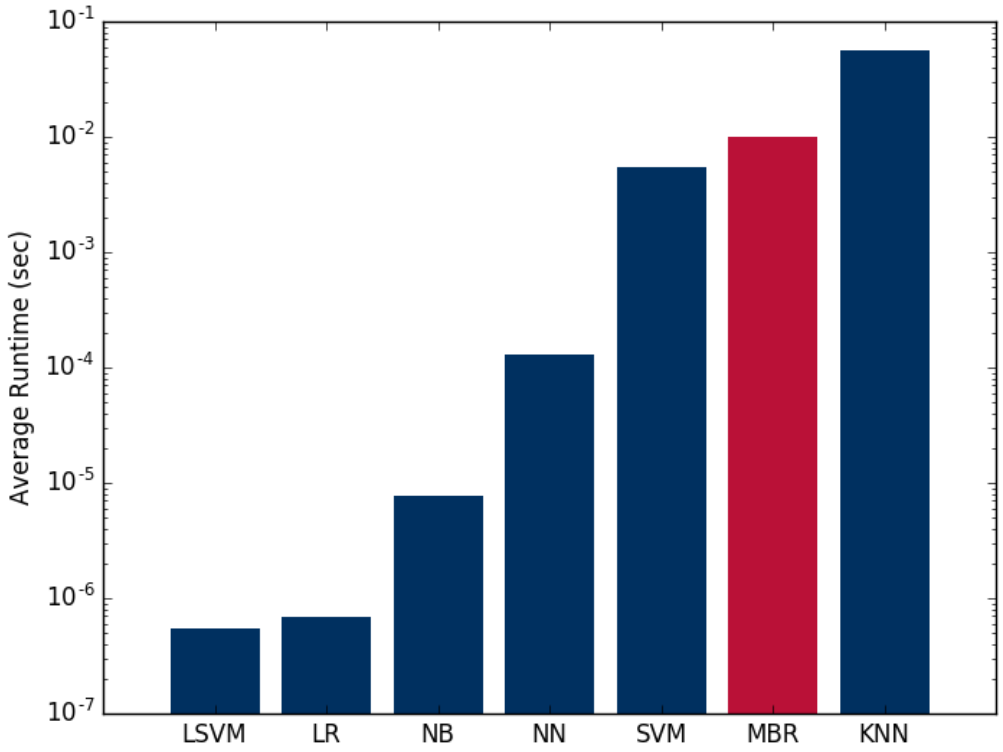

Figure 7.5: Average runtime for each Machine Learning algorithm. There are dramatic differences due to the way that they represent their decision boundaries.

Chapter 8

Summary and Conclusions

8.1 Summary

The direct detection of a dark matter signal is an intense focus of the astroparticle physics community. The DEAP-3600 dark matter direct detection experiment is made of an active volume of liquid argon contained in an acrylic vessel with 255 Photo-Multiplier Tubes. To detect a faint dark matter signal DEAP-3600 has been designed to have ultra-low background rates. There is still a small amount of contamination at the argon-acrylic boundary which is to be removed by selecting a fiducial volume. This thesis explored the use of several Machine Learning algorithms for the purpose of improving the ability of MBLikelihood to select events from a fiducial volume within the active volume. With an initial analysis using simulated data it was found that MBLikelihood outperformed the Machine Learning algorithms, but when used together it was possible to greatly improve performance. Next, the same algorithms were trained using real detector data and it was found that the same hierarchy of performance was maintained. For further tests of the algorithms several datasets were compiled to check that the algorithms were robust to all relevant situations. When

the optical model of the detector was varied, each algorithm tracked MBLikelihood in its performance. When the boosted algorithms were tested using high energy, real surface events it was found that they were able to reject them even more efficiently than MBLikelihood alone. When neutron source data was tested, it was found that the algorithms had a reasonable acceptance indicating that they would not lose a WIMP signal. Next, a brief analysis was performed on some more fine detail questions on the leakage. It was found that the surface event leakage as a function of energy was dependent in roughly the manner expected. By changing the desired fiducial volume, it was found that for MBLikelihood accepting a 1 tonne fiducial volume the boosted analysis could accept 1.5 tonnes with the same surface event leakage. This would lead to a significant increase in sensitivity for DEAP-3600. Finally, a comparison on the output values of the Machine Learning algorithms showed that using many of them to boost MBLikelihood would not be effective.

8.2 Future Work

This analysis has made it clear that MBLikelihood can be improved with respect to fiducial volume calculations. It may be possible to gain the same reduction in surface event leakage as the boosted analysis by making modifications directly to the MBLikelihood algorithm. However, it is only the Machine Learning algorithms that can be trained on real detector data. This suggests that it may be beneficial to invest time in improving the Machine Learning algorithms. There are several options for improvement, including the removal of the normalization preprocessing step. It would be worthwhile to redo the analysis in this thesis without normalization, but with a much larger training sample. Having information about the total charge of the event

(this is the information that normalization removes) might allow the algorithm to adapt to topological differences over different energies. Care would need to be taken to ensure that the algorithms do not learn pathological energy dependent leakages, but may allow the algorithms to individually achieve the same performance as the boosted analysis.

There are several more combinations of training data that would be interesting to use. For example, a mix of simulated and real data may ensure that performance is consistent between the detector and the simulator. While DEAP-3600 was under vacuum, a uniform source named the *laser ball* was deployed to calibrate the detector. Using the simulator, it would be possible to determine if the Machine Learning algorithms continue to perform normally when the detector is in vacuum. If this is the case, then data taken using the laser ball could be included in the training or testing data.

The results in this thesis are very promising and suggest that a similar analysis could be beneficial in other aspects of the detector. Another potential source of events that leak into the fiducial volume is related to the neck geometry. Events which occur within the neck can distribute light in a way that biases the reconstructed 3D position to be centred in the detector. It is possible that a Machine Learning algorithm trained on these events would be able to identify a specific signature and efficiently remove them. Of course, there are many more interesting avenues to explore in future work.

During the writing of this thesis, the simulation software (RAT) has gone through several iterations of improvement to better reflect the detector. This has been done thanks to standard improvements, as well as a better understanding of the detector now that it is online. These updates were not considered in this thesis, but have

been included in an updated version of MBLikelihood. Early indications are that the MBLikelihood performance has improved and so it will be worthwhile to see if an updated version of the analysis presented here will continue to give such large improvements.

8.3 Conclusion

fiducial volume selection is a critical component of the DEAP-3600 background rejection analysis, improving this can have a large impact on the final result. This thesis has analyzed the use of several Machine Learning techniques for the purpose of boosting the standard fiducial volume algorithm. It is found that the discrimination power can be increased significantly using a boosted analysis framework and the effect is dependent on the chosen fiducial volume and ranges from an order of magnitude improvement at 1 tonne to the performance at 1.5 tonnes which matches the MBLikelihood performance at 1 tonne.

Great care has been taken to validate the robustness of the Machine Learning algorithms. They have been shown to behave similarly to MBLikelihood under changes in the detector optics using simulated data. The benefits gained from boosting remain in the modified optics testing as was seen in the default optics. The acceptance for the boosted analysis was shown to be equivalent to MBLikelihood on real neutron events that mimic a WIMP signal. Nevertheless, the boosted analysis showed fewer ^{210}Po surface events in a fiducial volume than MBLikelihood on its own.

It is further comforting that the algorithms are trained using real data, they cannot have picked up on any peculiarities of the simulator. The sample of real data is blemished by the fact that it was classified by the Centroid algorithm, however the

relative simplicity of the algorithm is reassuring. The extensive testing on simulated data is considered a strong validation that boosting does indeed confer a reduction in leakage into the fiducial volume. This thesis concludes that the Logistic Regression or the linear Support Vector Machine should be used in tandem with the MBLikelihood algorithm for the purpose of selecting fiducial volume events. Many of the algorithms had similar performance, however those two run the fastest while maintaining that performance.

Machine Learning is a powerful and easy to use tool for data analysis. This thesis work has demonstrated that with relatively little effort it is possible to improve on custom built algorithms. The majority of the effort in this thesis has been in the process of validation, which is the ideal situation for algorithms which will hopefully be applied to the WIMP search analysis. There are many more avenues to apply Machine Learning in DEAP-3600 and other experiments.

Part II

AN EXTENSIVE STUDY OF THE SCATTER IN THE RADIAL ACCELERATION RELATION IN SPIRAL GALAXIES

Chapter 9

Outline

Part II investigates the scatter in the Radial Acceleration Relation (RAR) to determine if observational uncertainties alone can account for the measured scatter. The RAR is a spatially-resolved scaling relation that compares the observed radial acceleration (V_{rot}^2/R) with acceleration expected from visible matter alone (M_{bar}/R^2 where M_{bar} is the mass of all visible matter up to radius R). Before analyzing the data, Chapter 10 presents relevant information and techniques. Section 10.1 gives a broad overview of scaling relations in the context of galaxy physics, motivating studies such as the one presented here. Section 10.2 and Section 10.3 discuss relevant measurements used in analyzing galaxies, along with their typical uncertainties. They are crucial in the calculation of radial accelerations and must therefore be treated with care. Section 10.4 presents details of the RAR and the motivation for performing this analysis. Finally, Section 10.5 introduces the statistical methodology for studying the scatter in the RAR.

This extensive analysis of the RAR relies on collating several galaxy surveys whose available data include the required elements for the construction of the RAR, such as resolved rotation curves and light profiles. Chapter 11 presents an overview of each

survey. A short description of each is given in Sections 11.1 - 11.6. Additional information about each survey, figures with relevant parameters, and how these compare among the surveys are presented in Appendices A and B.

With relevant background and datasets in place, Chapter 12 examines the scatter in the RAR. This is the main focus of Part II: addressing the claim by Lelli et al. (2017) that the RAR relation is consistent with zero intrinsic scatter. The scatter in the observed RAR is calculated in Section 12.1. As a further check, the scatter in the RAR is plotted against several galaxy properties in Section 12.1.2 to test against any additional correlations. In Section 12.2 a Monte-Carlo simulation is used to determine what the scatter would be if only observational uncertainties contributed to the scatter. This simulated scatter for the RAR is then compared with the observed value; the former is shown to be consistently smaller than the latter, demonstrating that the intrinsic scatter is non-zero. This is in tension with the suggestion by Lelli et al. (2017), the RAR is found to be consistent with Λ Cold Dark Matter (ΛCDM) and standard galaxy formation models.

Chapter 10

Background

This chapter presents the necessary material for our RAR analysis. To motivate the study of scaling relations, Section 10.1 discusses how they are constructed and used to study galaxy formation models and cosmology. Section 10.2 and Section 10.3 explain the basic galaxy measurements and calculations that go into the RAR analysis. These are the inputs to equations presented in Section 10.4, which explains and motivates the study of the RAR. To ensure the robustness of our analysis, statistical techniques described in Section 10.5 are utilized to accurately evaluate the observational scatter in the RAR, which is the central focus of this work.

10.1 Scaling Relations

One of the most common signatures of galaxy formation and evolution is scaling relations. Structural parameters of various kinds can be measured for a given galaxy, how they correlate is of paramount interest. A scaling relation takes two or more of these measured values and connects them by a functional relationship, typically linear in log space. The relation is then described by the slope, intercept, and scatter about the best fit line (Courteau et al. 2007). These 3 parameters are then used to

tune or test galaxy formation and cosmological models (Trujillo-Gomez et al. 2011; Dutton et al. 2017).

The most powerful of these relationships are ones with low scatter and that interrelate distance-dependent and distance-independent quantities. The canonical example of this relates total luminosity and rotational velocity of a galaxy (Tully and Fisher 1977; hereafter Tully-Fisher relation); while it is relatively easy to measure the total magnitude of a galaxy and its rotational velocity, converting magnitude into luminosity requires knowledge of distance which is notoriously uncertain (Willick 1999). A relation between the star formation rate of a galaxy and its stellar mass called the the star formation rate main sequence, is an excellent example of a scaling relation that can be used to test galaxy formation models over cosmic time (Noeske et al. 2007; Santini et al. 2017).

The study of scaling relations for galaxies is central in our modern understanding of astronomy. The Tully-Fisher relation allows for straightforward distance measurements, making studies of the large scale structure of the universe possible. Galaxy simulations can also be tested against scaling relations; their ability to reproduce many relations will attest to the validity and reliability of the simulation (Dutton et al. 2017). Simulations can also use scaling relations as input to the sub-grid physics, such as the relation between gas density and star formation rate density, called the Kennicutt-Schmidt law (Schmidt 1959; Kennicutt 1998).

10.2 Galaxy Measurements

10.2.1 Surface Brightness Profiles

While the era of multi-messenger astronomy is certainly well under way (Abbott et al. 2017; Aartsen et al. 2018), the primary source of information on astrophysical objects is still almost always light. With sensitive telescopes and clever techniques, a great deal of information can be gleaned from the light emitted by a galaxy. One of the most fundamental measurements made of a galaxy is a light profile, which is a measure of the amount of light in a galaxy as a function of distance from the centre. For details about the extraction of galaxy light profiles and the structural parameters that can be gleaned from them, see Courteau (1996).

Light profiles can be measured in different wavebands across the electromagnetic spectrum which are all sensitive to different astrophysical phenomena. For instance, blue light samples young stars but suffers extinction from interstellar dust, whereas red light is more sensitive to old stars, is less sensitive to dust effects, and is a better tracer of the stellar mass. For this reason, our study will favour light profiles measured at redder bands.

Once the image of a galaxy has been taken, it must be reduced in order to provide useful information. This includes subtracting dark, sky, and bias frames as well as flat fielding (Courteau 1996; Mohr et al. 2012). It is often assumed that galaxies are axisymmetric such that the light in an image can be azimuthally averaged to produce a profile of the surface brightness as a function of radius, called a surface brightness profile (equivalently, one can construct a curve of growth). For inclined galaxies, these surface brightness profiles are obtained via fitting of elliptical isophotes (Courteau 1996; Hall et al. 2012; Gilhuly and Courteau 2018).

Light profiles of spiral galaxies are often approximated by exponential profiles $I(R) = I_0 e^{-R/R_d}$, where I_0 is the central surface brightness and R_d is the disk scale length (Freeman 1970; Lin and Pringle 1987). In practice, only $\sim 30\,\%$ of the galaxy population harbours pure exponential disks (Roediger et al. 2012). The broader range of galaxy light profiles can be described by various types (I, II, III) indicating if the disks are purely exponential (Freeman 1970 Type I), depressed in the inner parts (Freeman 1970 Type II), or (anti-)truncated in the outskirts (Erwin et al. 2005 Type III). In addition to the disk, a central, nearly spherical feature is often visible. Called a bulge, it is typically populated by older stellar populations and must be accounted for if performing a decomposition of the light profile.

10.2.2 Rotation Curves

Beyond measuring the distribution of light, one may examine spectroscopic profiles as well. Of interest here are techniques for measuring the line of sight velocity along the principle axis of a galaxy. Radial velocity information is especially useful in rotationally supported systems to trace their gravitational potential. This information is also useful in scaling relations as velocities are distance independent (modulo redshift corrections; Weinberg 1972); when paired with a distance dependent quantity they can be used as distance indicators (Tully and Fisher 1977). Another distance indicator available in the same spectroscopic information is the systematic velocity of a galaxy, parametrized as a redshift, this can be used as a distance indicator via the Hubble flow (Riess et al. 2016; Planck Collaboration et al. 2016b).

Various tracers of rotational motion include Hα, HI, and CO emission lines; globular clusters and satellites; in some cases even individual stars (Sofue and Rubin 2001).

To properly interpret rotation curve data, one must consider the possibility of non-circular motions (Sofue and Rubin 2001; Spekkens and Sellwood 2007). A significant source of non-circular motions in spiral galaxies is a bar, which drives some material in towards the centre, and other material out to the faint disk. The impact of a bar on a rotation curve depends on its orientation to the observer, but can be as high as a few tens of $\mathrm{km\,s^{-1}}$ in the inner parts of the galaxy (Spekkens and Sellwood 2007). Spiral arms can also move material along their length (Schoenmakers et al. 1997), thus contributing additional non-circular motions. These are not as strong as those from a bar, but are more ubiquitous (Trachternach et al. 2008). Hayashi and Navarro (2006) also find it is possible to have large non-circular motions near the centre of a galaxy as a result of a triaxial dark matter halo. Altogether, non-circular motions due to bars, spiral arms, and halo triaxiality are mostly dominant in the centres of galaxies.

Atmospheric turbulence (called seeing) can impact the inner parts of measured light profiles and rotation curves as well (Förster Schreiber et al. 2006). The effect of seeing is to redistribute the incoming light into neighbouring pixels on an image. Velocity and light information is therefore spread out as in a convolution. This effectively dampens the profiles wherever there would be large gradients, typically at the centre. For this thesis, the inner 5 arcseconds of each galaxy is cut from the analysis to mitigate the effects of seeing and non-circular motions which can dominate the central regions.

10.2.3 Size Measurements

One of the structural parameters that can be extracted directly from a light profile is the galaxy's size. The angular extent of a galaxy on the sky might seem simple to evaluate at first, but its definition is actually quite subjective. There are no clear edges to most galaxies as their surface brightness profiles can be measured out to the detectable limits. At some point the galaxy must end and the intergalactic medium must dominate. The boundary at which this occurs is not clearly defined and detecting it is beyond current technical capabilities. In some cosmological studies, it is informally assumed that galaxies simply extend until they meet the next galaxy (Ramella et al. 2001). In the absence of a clearly defined edge, some marker must be used to evaluate the size of a galaxy. The three most common markers are radii enclosing a specified fraction of the total light (R_e), isophotal radii ($R_{23.5}$), and model scale lengths (R_d).

The effective radius is the radius enclosing half of the total light from a galaxy. The advantage of an effective radius measurement is that it is independent of the total light in a galaxy, it only traces the distribution of the light. It is thus convenient to study the radial distribution of a property scaled by the effective radius when comparing galaxies.

Alternatively, one can use an isophotal radii which reports the radius at which a specific surface brightness is reached. Any surface brightness can be chosen and multiple isophotal radii can be used. A common choice is $R_{23.5}$ (R corresponding to $23.5\,\mathrm{mag\,arcsec^{-2}}$), which strikes a balance between probing the outskirts of a galaxy and having high signal to noise. The advantage of isophotal radii is that they are independent of the galaxy structure internal or external to that radius.

Another common radial scale comes from fitting a model to the surface brightness profile and using the scale parameters as a radius. In an exponential profile $I(R) = I_0 e^{-R/R_d}$, the parameter R_d is called a disk scale length. While this radius may be physically motivated, it only applies naturally to a subset of all galaxies (Type I). R_d can also be challenging to measure consistently and becomes less meaningful in non-Type I profiles.

All of these size measure techniques can be performed independent of distance and used to scale the radii in a profile. Thus large surveys can meaningfully compare many galaxies with measures taken at analogous locations in the profile (Binney and Tremaine 2008). It is also possible to compare the radii to each other, or with different versions of the same type. A good example of a distance independent parameter is the concentration index which takes the ratio of radii enclosing two different fractions of the total light (say 50 % and 90 %). The concentration index is a common measure for automated morphological classification (Kent 1985; Park and Choi 2005).

10.3 Galaxy Structural Parameters

10.3.1 Inclination

Spiral galaxies are flattened systems. The circular velocity of a rotating disk must therefore account for inclination effects (Byun et al. 1994). A face-on view of the disk is given an inclination of 0° while an edge-on view of the disk is given an inclination of 90°. From the photometry one can extract an axis ratio b/a (identified as q), where a and b are the major and minor axes respectively, at large radii by fitting elliptical isophotes; this gives the inclination by $i = \cos^{-1}(q)$. Since disks have a non-zero thickness, an edge on disk will still appear as a thin ellipse (axis ratio q_0) on the

sky instead of an infinitesimal line. The axis ratio of a disk when seen edge-on is labelled q_0 and can be used as a correction for other inclinations. A correction can be determined by assuming that the intrinsic thickness of the disk is also elliptical, giving Eq. 10.1 (Hubble 1926)

$$\cos(i) = \sqrt{\frac{q^2 - q_0^2}{1 - q_0^2}}. \tag{10.1}$$

The exact value of q_0 is poorly known, but is generally found to be near 0.2 and can vary by 0.1 between galaxies (Holmberg 1946; Haynes and Giovanelli 1984). A second complicating factor is that galaxies are not perfectly circular (Ryden 2004), this effect is not large and is not accounted for in this thesis.

Inclinations are used to correct rotation curves, making them a key structural parameter in this analysis. As galaxies are viewed in projection, the line of sight velocity is projected as well and so only the component along the line of sight is measured. Correcting velocities using inclinations is discussed in Section 10.4.2.

10.3.2 Colours and Mass to Light Ratios

The choice of photometric waveband determines which aspect of the galaxy is probed. As stated earlier, redder wavelength bands will be less affected by dust and can observe populations that are otherwise obscured in the visible (Draine 2003). It is also possible to use multiple filters and get separate images, as is done in the Sloan Digital Sky Survey (SDSS). Measuring a resolved galaxy spectrum over much of the disk is challenging due to the long integration times required. The Spectral Energy Distribution (SED) of a galaxy (energy as a function of frequency) can be quite complex, and so it is important to determine which components are relevant for a

given analysis. In some cases only rough spectral information is needed, for example the slope of the SED in a given region. This is achieved by taking the ratio of fluxes, or equivalently the difference in magnitude, from two different wavelength bands. Such a colour term is always written with the longer wavelength subtracted from the shorter wavelength band.

A mass-to-light ratio is a conversion factor between the amount of light observed and the amount of mass in a system. The light in a waveband comes from a population of stars which must be accurately modelled to derive a representative mass-to-light ratio. For main sequence stars, that conversion goes roughly as $L \propto M_*^{3.5}$ (Boselli 2011). This implies that high mass stars have much higher luminosity, and likely contribute little to the total mass, instead generating most of the galaxy light.

To understand a mass-to-light ratio, several ingredients must be put together: the stellar initial mass function, a library of stellar SEDs, well sampled isochrones, an initial-final mass function, and a star formation history (Conroy 2013). These are assembled in two stages: first a Simple Stellar Population (SSP) model is created by assuming that a large population of stars are made at one instant in time and then evolve together (said to be co-eval). To do this, an initial mass function is needed to determine the distribution of masses for the stars, usually over the whole range from $0.08\,\mathrm{M}_\odot$ to $100\,\mathrm{M}_\odot$ where stars burn hydrogen. A library of SEDs is needed to perform a weighted sum to compute the observed total SED of the population (e.g. Bruzual and Charlot 2003). Finally a library of isochrones are needed to track the evolution over time of the stellar population in the Hertzsprung-Russell diagram (Bertelli et al. 2008), and thus the evolution of their SEDs. Each of the components must be known as a function of metallicity as well. The sampling of

these can be done theoretically or observationally. In the case of theoretical libraries, the sampling can be more complete, but not all physical processes are modelled. Notably, cool stars are not properly represented and some lines (e.g. CH and MgH) or the TiO bands are incorrectly modelled (Martins and Coelho 2007; Allard et al. 2011). Empirical libraries have the opposite problems of theoretical ones. The physics is, by definition, modelled correctly, but the libraries are sparsely sampled and stars may be assigned incorrect parameters (Sánchez-Blázquez et al. 2006). To handle sparse sampling of the libraries, one must interpolate which is challenging given the large range of stellar properties (Vazdekis et al. 2010).

The SSP models can then be combined to form a realistic galaxy. Using a star formation history of a galaxy, one can consider SSPs of appropriate ages and metallicities to construct an SED. While any star formation history could in principle be used, usually simple prescriptions are chosen such as an exponentially decreasing star formation rate (closed box) or one which rises (hierarchical formation) (Schmidt 1959; Finlator et al. 2007; Papovich et al. 2011). For a given age, the initial-final mass distribution can be used to determine the number of stellar remnants (white dwarfs, neutron stars, black holes) contributing to the mass. This is not a negligible amount, and can represent 25 % of the remaining stellar mass that has not been recycled into nebula and dust (Conroy 2013). Finally a dust model must be included to account for absorption/emission spectra that alter the SED emitted by a galaxy. The resulting model couples the measurable SED of a galaxy to the total stellar mass, giving a stellar mass-to-light ratio; it is called a Stellar Population Synthesis (SPS) model.

Measuring the SED of a galaxy is challenging and sometimes only photometry is available. The SPS models can be projected into photometric bands by integrating

the filter response curves, losing much of the well-resolved information. In most cases a single photometric band cannot distinguish between different SPS models, due to the large luminosity evolution of stellar populations over time. However, with flux measurements in two bands the slope of the SED can be estimated with their ratio. Surprisingly, this provides enough information to determine the stellar mass-to-light ratio to within less than 0.3 dex (dex indicating the value is in $\log_{10}$ space), of which only about 0.13 dex is random error (Roediger and Courteau 2015). The colour to stellar mass-to-light ratios used in this thesis are taken from Zhang et al. (2017), which provide a detailed analysis of the uncertainties for their relations.

10.3.3 Distances

One of the great challenges in astronomy is the determination of cosmic distances (Hodge 1981; Freedman and Madore 2010). Many galaxian parameters rely on knowledge of the distance to an object, such as the luminosity or physical size. Both of these quantities are important for this thesis work; the luminosity is needed for calculations of the baryonic mass, and sizes are needed to compute the radial acceleration. There are three typical methods by which distances are calculated in astronomy: standard candles, standard rulers, and Hubble flow. A standard candle is an object of known luminosity, thus when it is placed at a distance one can use the apparent brightness to calculate distance (Kowalski et al. 2008). Similarly, a standard ruler is an object of known size, when placed at a distance one can use the angular size projected on the sky to determine distance (Djorgovski and Davis 1987). While the first two are common distance measuring techniques, the Hubble flow is unique to astronomy (Riess et al. 2016). As was discovered by Hubble (1929), objects which

are further away also have larger line of sight recessional velocities. The recessional velocity as a function of distance, the so-called Hubble-Lemaitre constant (H_0) is generally computed in two ways. From the Cosmic Microwave Background (CMB): $H_0 = 67.8 \pm 0.9\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Planck Collaboration et al. 2016b), or with distance measurements to nearby galaxies: $H_0 = 73.2 \pm 1.74\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Riess et al. 2016).

Distance measurement uncertainties play an important role in this thesis. The Hubble flow accounts for the global expansion of space, but does not include peculiar velocities of individual galaxies. The magnitude of peculiar velocities is uncertain, Kessler et al. (2009) take $\sigma_v = 300\,\mathrm{km\,s^{-1}}$ and Wood-Vasey et al. (2007) use $\sigma_v = 400\,\mathrm{km\,s^{-1}}$, it is known to be higher in cluster environments as well (Girardi et al. 1993). Thus, distances determined by this method are uncertain by the relative magnitude of the Hubble flow velocity and the peculiar velocity of a galaxy. Distances for the samples in this thesis are typically beyond 50 Mpc, with a conservative assumption of $\sigma_v = 500\,\mathrm{km\,s^{-1}}$, this gives a distance uncertainty of $\sim$15 %. For any dataset that does not report distance uncertainties, a common 15 % value is assumed in this thesis, except for the Ouellette et al. (2017) sample which is in a cluster environment and is assigned a 19 % uncertainty (see Section 11.6).

10.4 The Radial Acceleration Relation (RAR)

10.4.1 What is the RAR?

In a Modified Newtonian Dynamics (MOND) framework, the acceleration in Newton's second law is modified below a certain scale. While the exact modelling of this modification is not well defined, it must have the properties described in Section 1.2.3. To test MOND with rotation curves, it is natural to examine galaxies in acceleration

space. Another related avenue to explore in a MOND framework is the baryonic acceleration. As the model does not include dark matter, gravitation is solely a consequence of the visible matter distribution. Lelli et al. (2017) have presented an analysis with these expectations. Using a sample finely selected to cover a wide parameter space (see Lelli et al. 2016), they compute the observed radial acceleration versus the baryonic acceleration as expected in regular Newtonian mechanics. The data populate a tight relation in this space which they parametrize using Eq. 10.2.

$$g_{obs} = \frac{g_{bar}}{1 - e^{-\sqrt{g_{bar}/g_\dagger}}}, \tag{10.2}$$

where g_{bar} is the baryonic radial acceleration and g_{obs} is the observed radial acceleration, $g_\dagger$ is a characteristic scale that determines the transition from a Newtonian dynamics to a MOND regime. This formulation is chosen because it reproduces the MOND prescription of matching Newtonian dynamics for acceleration above the scale $g_\dagger$, while below $g_\dagger$ the acceleration goes as $g_{obs} = \sqrt{g_\dagger g_{bar}}$, smoothly transitioning between them (see: Section 1.2.3). In a MOND framework, the RAR will apply to all galaxies regardless of evolution history, morphology, etc. The only systems that would not obey this relation are ones not in equilibrium configurations, those with systematic observational uncertainties such as non-circular motions, or systems containing dark matter. Lelli et al. (2017) measured the scatter of the forward residuals (residuals in the y-axis) about this relation to be 0.11 dex using a Gaussian fit, or 0.13 dex as the root mean squared. This remarkably small scatter is reported to be consistent with the average observational uncertainties in the survey. Suggesting that the relation may have zero intrinsic scatter. If true, this would challenge the current ΛCDM paradigm of structure formation in our universe.

10.4.2 Calculating the Observed Acceleration (g_{obs})

The RAR is based on both the baryonic and observed radial accelerations. From a galaxy rotation curve one may calculate the observed radial acceleration using $g_{obs}(R) = V_{rot}^2/R$, where R is the physical radius at which V_{rot} and g_{obs} are evaluated. This assumes that the observed velocities are from tracers travelling in circular orbits. R must be converted from the observed angular radius θ to a physical radius (in km) via the formula $R = \theta D$ where D is the distance. This introduces a distance uncertainty into the calculation of g_{obs}. V_{rot} is measured from a rotation curve (see Section 10.2.2). Critically, the measured line of sight velocities (V_{los}) must be corrected to the rotational velocity V_{rot} using $V_{rot} = V_{los}/\sin(i)$, where i is the inclination of the galaxy (redshift corrections are negligible in the distance regime considered). As discussed in Section 10.3.1, inclination estimates have significant uncertainties that must be accounted for. For low inclination galaxies ($i < 30°$), the inclination and its uncertainty can have a large impact on the inferred rotational velocity.

10.4.3 Calculating the Baryonic Acceleration (g_{bar})

The baryonic acceleration must be inferred from the measured light profile (see Section 10.2.1). The light must first be transformed into a stellar mass measurement using a mass-to-light ratio (see Section 10.3.2). For a curve of growth, that produces a mass profile ($M_\star(R)$); for a surface brightness profile it yields a surface density profile ($\Sigma_\star(R)$). Simplifying assumptions must then be made regarding the 3D structure of the galaxy in order to compute the potential $\phi_\star$ as a function of position. Finally, radial acceleration can be extracted from the potential as: $g_\star = \frac{\partial \phi_\star}{\partial R}$. In this thesis it is assumed that the stellar acceleration $g_\star$ is identical to the baryonic acceleration

g_{bar}. For galaxies with high gas masses this is certainly not the case, however Lelli et al. (2017) show that the scatter is unaffected by using $g_\star$ instead of g_{bar}.

Under the assumption that galaxies are spherical systems, only the enclosed mass is needed to compute the potential. The acceleration can then be computed from Newton's Law of Universal Gravity: $g_{bar} = \frac{GM(R)}{R^2}$, where G is the gravitational constant. While a sphere may be a reasonable approximation for the bulge of a galaxy, the disk is far from spherical. This can cause discrepancies between a spherical and a flattened model (Freeman 1970; Binney and Tremaine 2008).

Under the assumption that galaxies are flattened systems, the calculations become more involved. First, a disk flattening prescription must be invoked. Assuming axisymmetry and an isothermal vertical distribution the 3D density distribution can be written as $\rho_{bar}(R, z) = \Sigma_{bar}(R)\mathrm{sech}^2(z/z_0)/(2z_0)$, where ρ_{bar} is the density, z is the height above the disk, and z_0 is the scale height (van der Kruit and Searle 1981). We use $z_0 = q_0 R_d$ from Dutton et al. (2005) where R_d is the disk scale length and q_0 the intrinsic thickness described in Section 10.3.1.

With a prescription for the density it should now be possible to determine the gravitational potential by solving Poisson's equation $\nabla^2\phi = 4\pi G\rho$ (Binney and Tremaine 2008). For this, we take advantage of the software *galpy*, by Bovy (2015). *Galpy* solves Poisson's equation via the self-consistent field method; this involves a multipole expansion that can be taken to arbitrary accuracy and is described by Hernquist and Ostriker (1992).

10.4.4 The RAR as an Alternative to the Baryonic Tully-Fisher Relation

The RAR as presented in Lelli et al. (2017) has spurred a rich literature (Bílek and Samurović 2017; Famaey et al. 2018; Garaldi et al. 2018; Rodrigues et al. 2018). One especially insightful paper by Wheeler et al. (2018) investigates the connection between the RAR and the Baryonic Tully-Fisher Relation (BTFR). Wheeler et al. (2018) showed that the tightness of the RAR is strictly a consequence of its construction; their argument is reproduced here for completeness. In Eq. 10.2 for the limit as the baryonic acceleration is small (below the characteristic scale $g_\dagger$), the observed acceleration will tend to $\mathrm{g_{obs}} = \sqrt{\mathrm{g_\dagger g_{bar}}}$. Most spiral galaxy rotation curves are flat out to large radii and thus can be represented by a single value $V_{\rm flat}$. The BTFR is a scaling relation connecting the baryonic mass to $V_{\rm flat}$ with the relation: $M_{bar} = aV_{\rm flat}^4$ (note that the power differs from other BTFR studies, Hall et al. 2012; Ouellette et al. 2017; and references therein). A significant amount of the total baryonic mass in a galaxy is concentrated in the centre, meaning that the approximation $\mathrm{g_{bar}} \approx \frac{\mathrm{GM_{bar}}}{\mathrm{R^2}}$ applies in the outer parts of a galaxy. Consequently, one can derive the RAR as follows:

$$\mathrm{g_{bar}} = \frac{GM_{bar}}{R^2} = \frac{GaV_{\rm flat}^4}{R^2} = Ga\left(\frac{V_{\rm flat}^2}{R}\right)^2 = Ga\left(\mathrm{g_{obs}}\right)^2$$

$$\text{Let} : g_\dagger = (Ga)^{-1}$$

$$\mathrm{g_{obs}} = \sqrt{\mathrm{g_{bar} g_\dagger}}$$

Thus, the RAR can be thought of as a consequence of the BTFR or vice-versa. Wheeler et al. (2018) also examine the scatter in this model. Representing a galaxy that deviates from the BTFR by a factor $1+\delta$, the BTFR becomes: $M_{bar} = aV_{\rm flat}^4(1+\delta)$. Carrying this through the derivation gives: $\mathrm{g_{obs}} = (1+\delta)^{-1/2}\sqrt{\mathrm{g_{bar} g_\dagger}}$. As the RAR is represented in log space, the deviation is reduced by a factor of 1/2 implying a low

scatter compared to the already tight BTFR.

This argument broadly explains why the RAR is so tight in terms of the BTFR, but still leaves open some questions. First, it only applies in the outer regions of a galaxy, as it was assumed that all the baryonic matter is enclosed and the potential can be treated as spherical, yet the relation is seen to be obeyed at all radii. Second, the question of the origin of the BTFR is still left open, as both relations imply a "conspiracy" between baryonic and dark matter (however, see Dutton et al. 2007). Third, if this fully explained the reduction in scatter, then the predicted observational scatter should be reduced as well if propagated through the same equations.

10.5 Statistical Techniques

10.5.1 Bootstrap Sampling

The central limit theorem states that the sum of many independent random variables will tend toward a normal distribution. It is common for scientific measurements to fit this description (for example, the counts in a well illuminated pixel), greatly simplifying analysis. However, this is not always the case (for example, Poisson statistics). In some situations a variable that is normally distributed will then be transformed through a function such as $1/x$ or $\log(x)$ which will distort the shape of the distribution asymmetrically. In still more situations, there is a bias in the data such as for flux limited measurements that may reduce the number of samples on one region of parameter space or increase the number of samples elsewhere (for example Malmquist bias). It is therefore desirable to use techniques that are not based on an assumption of normality.

Bootstrap sampling is perhaps the simplest technique for computing measures of

uncertainty without assuming a probability distribution. In this method, a dataset is used as its own distribution function from which more samples can be drawn (Davidson and Hinkley 1997). For a dataset of N measurements a new sample of size N is drawn by randomly choosing elements from the dataset (with replacement). This can be repeated many times, for each re-sampling one can compute any metrics that would have been computed on the original data such as a median. The result is that instead of having one value for the metric, a distribution is acquired. From this distribution one can report an interval, which for this thesis is computed as half that enclosing 68.3 % of the data such that the same number of elements are on each side of the interval. This interval will be indicated as Δ while the standard deviation ($\sqrt{\sum(x_i - \bar{x})^2/N}$ where $\bar{x}$ is the mean) will be indicated with σ. In the limit of large amounts of normally distributed data this technique will yield the same value for the uncertainty as a standard deviation. However, if the data are not normally distributed then the uncertainties will still give reasonable values that are robust against outliers (DiCiccio and Efron 1996).

The bootstrapping method does have its own drawbacks (Laberge 2011). Most notably (as used in this thesis), there is an assumption that the uncertainty for each sample is the same and that samples are all independent. This is not the case for the samples in the RAR (for example $\delta V_{\rm obs}$ varies throughout a rotation curve), however the variations in uncertainty do not span such a wide range that it is unreasonable to use bootstrapping. The samples are also not all independent, for example the distance and inclination are shared over a whole galaxy.

10.5.2 Regression Analysis

When performing a regression to fit parameters to data, the choice of fitting algorithm can significantly impact the results (e.g. Hogg et al. 2010). The standard choice in many fitting scenarios is a forward least squares fit, where the model is chosen to minimize $\sum_i [y_i - f(x_i)]^2$, where f is the model and x_i, y_i are measurement pairs where x_i is assumed to be error free. This technique is strongly biased when there is uncertainty on both variables. The inverse least squares fit reverses the roles of x_i and y_i by assuming y_i has no error. Thus, it has the opposite bias and a straight forward method of rectifying this is to take the bisector between the forward and inverse fits. The bisector method is thus easy to implement, as one only needs to perform two least squares fits (Isobe et al. 1990). Another option is to minimize the orthogonal distance between the model and each data point (Boggs et al. 1992). Figure 10.1 demonstrates how these four methods perform on a sample of 200 points drawn from a 2D Gaussian distribution. The distribution is constructed in arbitrary units with a covariance matrix of $\begin{bmatrix} 0.85 & 1.0 \\ 1.0 & 0.85 \end{bmatrix}$ thus laying on the 1:1 line. The first three panels show the resulting fits under three conditions: all data is available, the data is truncated in the x-axis (indicated by vertical lines), and data truncated in the y-axis (indicated by horizontal lines). The bottom right panel shows the forward scatter σ against the slope for each fit with lines indicating the true values. It is clear from Figure 10.1 that the orthogonal and bisector methods can be far less biased in cases where both axes have uncertainty and/or truncations. For this thesis, fits are performed using the orthogonal distance regression as implemented for python in scipy (Jones et al. 2001).

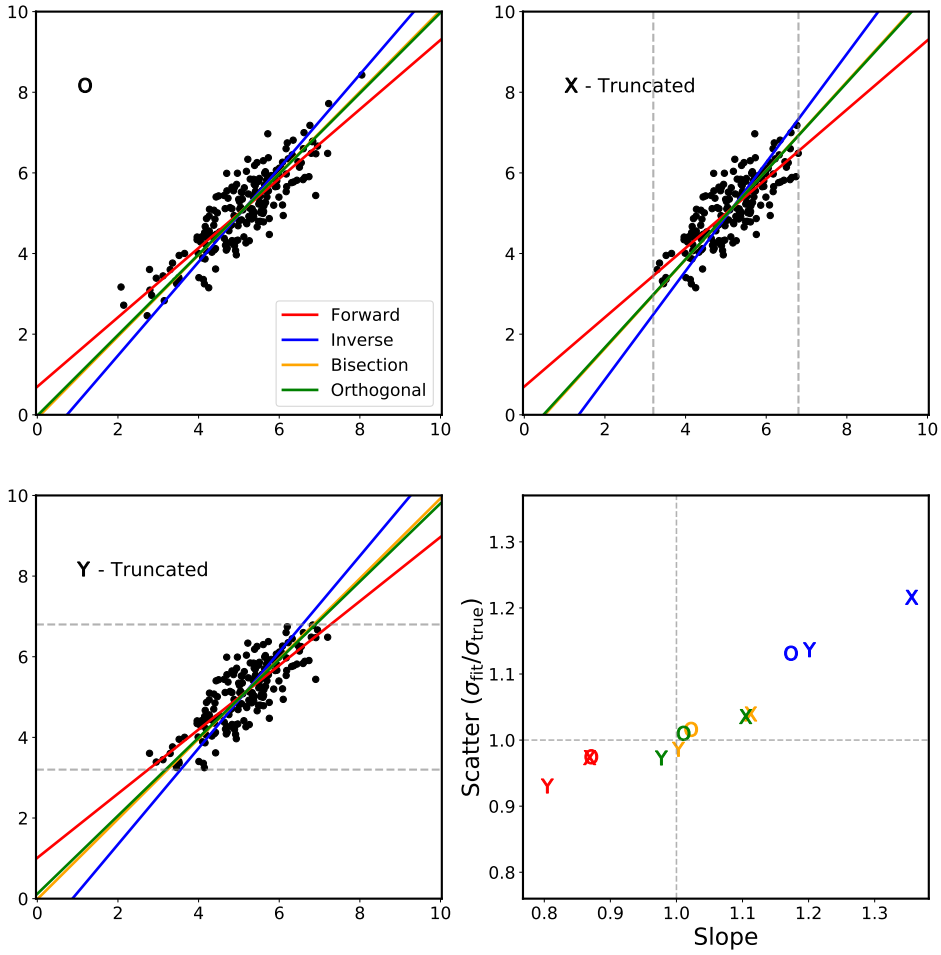

Figure 10.1: Example of fitting with several different regression algorithms. The bottom right panel shows the range of fit parameters depending on the method chosen.

10.5.3 Monte-Carlo Sampling

Propagating uncertainties through a function can be achieved to first order by summing uncertainties in quadrature multiplied by the partial derivative of the function evaluated at that point, as in Eq. 10.3 if only the first term is used.

$$\sigma_f^2 = \sum_i \left[\left(\frac{\partial f}{\partial x_i} \right)^2 \sigma_{x_i}^2 + \frac{1}{2} \sum_j \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)^2 \sigma_{x_i}^2 \sigma_{x_j}^2 \right] + \mathcal{O}(3), \tag{10.3}$$

where $f(\vec{x})$ is a function of several variables and σ_i is the uncertainty in each variable x_i, the result of Eq. 10.3 is the uncertainty σ_f in the value $f(\vec{x})$. This technique can result in several problems. If the second order contributions are significant, then the uncertainty value will be unrepresentative. Second order contributions can be large for highly non-linear functions, such as $\frac{1}{x}$ in the region near zero. They can also be large for strongly correlated variables such as in a light profile, the surface brightness and colour are correlated when they share a waveband. Finally, if f involves operations through which the derivatives cannot be evaluated then Eq. 10.3 is unsuitable. An example of this would be a function which includes a cutoff threshold, or a fitting routine. Each of these cases are seen in this analysis, necessitating a more robust uncertainty propagation technique.

With the power of modern computers a full treatment of uncertainties can readily be achieved. The variables x_i can be sampled from a distribution then propagated through the function f normally. The simplest case for the sampling distribution is a Gaussian with standard deviation σ_i. As the sampling is discrete, it is possible to impose extra constraints or priors to make the results more realistic as well. For example some variables are only physical for positive values, a condition that can be

easily accommodated in this scheme. Done many times, this procedure will produce a distribution for f that represents the combined uncertainty from all σ_i. If it is Gaussian, that distribution for f can then be parametrized with a mean and standard deviation. Often when using non-linear functions the result will not be Gaussian, in which case other metrics to represent the uncertainty can be used such as the interval described in Section 10.5.1.

Monte-Carlo sampling can treat a whole analysis routine as f by presenting it with re-sampled "observations" and allowing the code to run normally (Kroese et al. 2014). Even conditional statements can be properly handled, which may be important whenever a sample is near a region that it is triggered. Also, in the case where f is linear and all variables x_i are independently distributed Gaussians, Monte-Carlo sampling will reproduce the standard uncertainty as in Eq. 10.3. The sampling operation can then be understood as a generalization to infinite order in the limit that a large sample is drawn. There are, however, some drawbacks to this technique. One is that it is far more computationally expensive; while modern computers are powerful, they are often applied to complex problems, constraining the number of times that analyses can be re-run. Another drawback is the level of complexity that can be applied with Monte-Carlo sampling, if not explained fully it becomes challenging for others to reproduce the results.

Chapter 11

Datasets

To study the RAR as extensively as possible, several distinct datasets with the necessary measurements have been collated. Sections 11.1 - 11.6 introduce the six studies used in this analysis in chronological order. The original motivation for each is discussed as well as any unique properties of the dataset that are relevant to this thesis. Section 11.7 then explains, in detail, the data quality cuts that are applied to all samples to prepare them for the RAR analysis. Wherever possible, a uniform approach is applied to all datasets.

11.1 Mathewson et al. (1992; M92)

The survey by Mathewson et al. (1992; hereafter M92) represents galaxies visible from the southern hemisphere, complementing other surveys which had primarily focused on the northern hemisphere. This survey, along with a second released in 1996, comprise the largest samples used in this thesis. A total of 705 galaxies have the measured quantities needed for this study. While they do not have resolved colour information, most galaxies have a globally measured B-I colour. Where colours were not reported in the survey, colour information was retrieved from the NASA/IPAC

Extragalactic Database (NED) accounting for 69 of the galaxies. The uncertainty on the surface brightness values is not reported in this dataset and so a uniform uncertainty of 0.05 mag is assumed, corresponding to roughly 5 % uncertainty in flux space as suggested in M92. This is a small uncertianty compared to others in the analysis, but it is kept for completeness and to asses potential scatter introduced by the analysis proceedure which must make slight extrapolations of the surface brightness profile.

11.2 Mathewson and Ford (1996; M96)

Expanding upon the M92 dataset, Mathewson and Ford (1996; hereafter M96) sampled over 1000 galaxies making it one of the largest samples of resolved rotation curves ever assembled. In this analysis, only galaxies which were not in M92 are used from the M96 sample. Using the same methodology as M92, Hα rotation curves and I-band photometry were measured allowing the two datasets to be easily used in tandem. These data were originally collected for studies of peculiar velocities using the Tully-Fisher relation. In this thesis, we now use the data for our extended analysis of the RAR.

For the surface brightness measurements we assume a uniform uncertainty of 0.05 mag (M96). Only 19 galaxies have measured B-I colours. Since colour information is needed in calculating mass-to-light ratios, colour values (B,V,R,I) were retrieved from NED, accounting for 415 galaxies. Where no colour information was available, a median value of 1.0 $M_\odot/L_\odot$ was used for 886 of the galaxies. The 68.3 % interval about this median mass-to-light ratio is 0.2 and thus the variation is smaller

than the 0.13 dex uncertainty assumed for mass-to-light ratios in this thesis. 38 galaxies had to be rejected as they had no morphological type information. In total 1280 galaxies had available information and could be used for our RAR analysis, some of which may have been removed by further cuts described in Section 11.7.

11.3 Courteau (1997; C97)

Courteau (1997; hereafter C97) compiled a sample of 300 Sb and Sc type galaxies with Hα rotation curves and r-band photometry for the purpose of cosmic-flow studies. Using multiple measurements, specific focus was given to understanding the uncertainties on observational values. Care was taken with data modelling techniques (Courteau 1996; C97). Some galaxies have as many as 4 rotation curve measurements, and more than half of the surface brightness profiles have 2 or more measurements. These multiple measurements were used to asses the accuracy of uncertainty measures on the profiles. As there are multiple measurements for most galaxies (photometry and/or rotation curves), a scheme must be devised for selecting which measurement to use. In this thesis, the deepest (greatest radial extent) profile is chosen for our analysis as is common for multiple measurements (C97).

11.4 ShellFlow (SF)

ShellFlow is a sample of over 150 galaxies originally designed to measure the cosmological bulk flow of galaxies in a shell at redshift between 4500 and 7000 $\mathrm{km\,s^{-1}}$ (Courteau et al. 2000; hereafter SF). With both V and I-band photometry, SF represents one of two datasets in this thesis for which resolved colour profiles are available. Hα rotation curves are also available. The SF data were optimized for bulk flow measurements

as can be seen in Appendix A. As SF was designed to fill a shell, the galaxies are sampled from both the northern and southern hemisphere Overlapping regions of the sky were used as extra calibrations for the telescopes and so this represents a very high quality dataset.

11.5 Spitzer Photometry and Accurate Rotation Curves (SPARC; SP)

SPARC is an amalgamation of over 50 small studies totalling 154 spiral galaxies with high quality HI and Hα rotation curves as well as Spitzer 3.6 μm photometry spanning a wide parameter space in luminosity, morphology, and surface brightness (Lelli et al. 2016; hereafter SP). SP photometry is in the Spitzer 3.6 μm band, meaning that it traces low mass stars which contribute most to the stellar mass of a galaxy (see Section 10.3.2). Resolved rotation curves for this dataset are available mostly in HI; for the inner regions, $\sim 1/3$ of SP galaxies have hybrid HI-Hα data to achieve higher resolution. HI rotation curves are very useful since they can extend beyond the Hα rotation curves. Hα rotation curves have higher spatial resolution, and the two agree for the most part in the overlapping regions (SP). Included with this dataset is a quality factor; all low quality (level 3) galaxies are removed from the analysis. Distance uncertainties are reported for this dataset (median uncertainty of 14 %), and so the provided values are used in the analysis. A mass-to-light ratio of $0.5\,M_{\odot}/L_{\odot}^{3.6\mu m}$ is adopted for all galaxy disks as in SP, the whole galaxy is considered a disk in this analysis.

11.6 The Spectroscopy And H-Band Imaging Of Virgo Cluster Galaxies Survey (SHIVir; SV)

SHIVir is an ongoing project to accurately measure photometry and spectroscopy for Virgo Cluster galaxies (Ouellette et al. 2017; hereafter SV). The resulting dataset can be used for analysis of scaling relations in a cluster environment by taking advantage of the closest galaxy cluster in the sky. In their paper SV showed that SHIVir data are of high quality and cover a wide range of galaxy morphological types. This dataset consists of 46 galaxies with resolved Hα rotation curves and SDSS + H-band photometry, allowing for the determination of resolved colours. While it is the smallest survey used in this thesis, its high quality data due to consistent methods and resolved colour information are valuable for testing the RAR in a cluster environment. Distances and their uncertainties are provided for most of the galaxies in this survey (median uncertainty of 19 %); when missing, a distance of 16.5 Mpc is used with a 19 % uncertainty (SV).

11.7 Data Quality Selections

A strength of this analysis is that it combines several surveys simultaneously in a uniform fashion. In that spirit, many of the data quality selection cuts are identical for all surveys. This section explains our data cuts and why they were needed. Some selections were also made on a per-survey basis and are described in the corresponding section for each survey.

The first level of selection is that each galaxy must have certain measurements for the analysis to proceed. Each galaxy must have: a rotation curve, surface brightness profile, distance, inclination, and a morphological type. Ideally, uncertainties are also

be available for these parameters. If an uncertainty is not provided, an average value is used instead (see Sections 11.1 - 11.6).

For inclinations, uncertainties are universally missing and a model was developed to express the uncertainty which is described in Section 12.2.1. A distance uncertainty of 15 % is used for the surveys: M92, M96, SF, and C97. Almost all galaxies in these samples are beyond 50 Mpc, thus the line-of-sight peculiar velocities would have to be above 500 $\mathrm{km\,s^{-1}}$ to cause a greater than 15 % error (see Section 10.3.3). Discrepancies in H_0 values from Planck Collaboration et al. (2016b) and Riess et al. (2016) do not enter into this analysis as they represent a constant scaling that will not change the scatter of the RAR in log space.

The rotation curves are further examined for suitable use in RAR calculations. Local velocity uncertainties must be less than 10 % of the velocity measurement; this is the same cut used in Lelli et al. (2017) and ensures the scatter is being probed by high quality data instead of outliers. Rotational velocities are also removed from the analysis if they are less than 10 $\mathrm{km\,s^{-1}}$ to prevent large logarithmic scatter in the $\mathrm{g_{obs}}$ calculation which goes as V^2. Radii less than 5 arcsec from the galaxian centre are also removed from the analysis as this region can be influenced by seeing effects.

The surface brightness profiles are also subject to selection cuts. First, any regions fainter than 25 $\mathrm{mag\ arcsec^{-2}}$ are removed from the analysis. While some surveys can sample the light at fainter levels, uncertainties grow rapidly. Likewise, any surface brightness uncertainties greater than 0.1 $\mathrm{mag\,arcsec^{-2}}$ are also removed from the calculations. For all surveys but SP, the curve of growth and surface brightness profiles are sampled at the same radii and so data quality cuts are used to remove points from both profiles.

The morphological type is restricted to the range Sa-Sm. Inclinations must exceed 30° to avoid serious deprojection errors (Courteau 1997). Whenever data must be interpolated, a cubic function is fit to nearby points and then sampled, this arises when g_{bar} and g_{obs} must be related but the surface brightness profile and rotation curve are sampled at different radii. When data must be extrapolated, a linear function is fit to the nearby points and then sampled, this arises when doing a multi-pole expansion of the baryonic gravitational potential.

Chapter 12

The Scatter in the RAR

We now examine the RAR and its scatter in detail. Section 12.1 describes fitting the RAR to the data. The results of the fits are presented, including the scatter in the relation. In Section 12.1.2, we examine any correlations of the scatter with various galaxian parameters. Then, in Section 12.2 the Monte-Carlo simulation with zero intrinsic scatter is presented. The details of its construction are discussed and the resulting fits with scatter are examined. It is found that the Monte-Carlo simulation scatter is below the observed scatter by a significant amount, with an intrinsic scatter on the order of 0.13 dex.

12.1 The Scatter in the Observed RAR

12.1.1 Fitting the RAR

With several datasets collected, it is possible to compute the RAR relation. The datasets are not combined into a single sample as systematic differences between them could add observed scatter. Any systematic difference which is a constant factor for a whole dataset (such as an incorrect H_0) will only impact the normalization of the RAR

Survey	$g_\dagger$ (10^{-10} m s^{-1})	$\sigma_{g_\dagger}$ (dex)	$\Delta_{g_\dagger}$ (dex)	N Galaxies	N Samples
SV	2.99 ± 0.11	0.285 ± 0.009	0.182 ± 0.012	33	1562
SP	2.38 ± 0.04	0.138 ± 0.002	0.131 ± 0.005	115	2297
SF	2.79 ± 0.02	0.160 ± 0.001	0.146 ± 0.002	165	17363
C97	2.29 ± 0.02	0.187 ± 0.001	0.177 ± 0.003	297	18088
M92	2.72 ± 0.03	0.198 ± 0.002	0.172 ± 0.003	699	12666
M96	2.29 ± 0.02	0.191 ± 0.002	0.163 ± 0.003	1240	17287

Table 12.1: Parametrization for the RAR fit to each dataset. Column $g_\dagger$ is the parametrization for Eq. 10.2 and its bootstrap uncertainty, $\sigma_{g_\dagger}$ and $\Delta_{g_\dagger}$ are the standard deviation scatter and the 68 % interval scatter respectively reported with their bootstrap uncertainty. The final two columns represent the number of galaxies and number of samples in the RAR.

(in log space) and not the scatter which is of interest for this study. Figure 12.1 shows a two dimensional histogram of measured points in the RAR for each dataset. Each dataset has been fit individually with Eq. 10.2 using orthogonal distance regression in $\log_{10}(g_{bar})$, $\log_{10}(g_{obs})$ space. The results of each fit can be seen in Table 12.1, where each row represents one dataset, the scale parameter $g_\dagger$ and scatter measurements $\sigma_{g_\dagger}$, $\Delta_{g_\dagger}$ about the relation are given in columns 2, 3, and 4 respectively. Bootstrap uncertainties are presented alongside their corresponding quantity. Columns 5 and 6 give the number of galaxies and points in the RAR for each dataset. Of interest for this analysis is the random scatter about the RAR and not its exact parametrization ($g_\dagger$). Two measures of scatter are used, with $\sigma_{g_\dagger}$ representing a standard choice for scatter measurements and $\Delta_{g_\dagger}$ used as it is robust to outliers. Figure 12.2 presents a histogram of the residuals for each dataset along with a Gaussian fit for comparison; this figure demonstrates the non-Gaussian nature of these calculations.

All fits return values for $g_\dagger$ that are comparable, however they differ by far more than the bootstrap random uncertainty. These are signatures of systematic differences

between the surveys, as expected, but are not the focus of this work. The $\sigma_{g_\dagger}$ scatter about each relation has a median value of 0.19 dex, while the median $\Delta_{g_\dagger}$ interval scatter is 0.17 dex. The larger $\sigma_{g_\dagger}$ scatter for SV is due to small sample size, as can be seen in Figure 12.1 only a few galaxies that deviate significantly from the fitted RAR are increasing the scatter. The $\Delta_{g_\dagger}$ scatter is much more consistent with the others as it is robust to outliers. SP has the lowest scatter of the surveys, which may be an unintended result of the careful selection of galaxies for the study. SF has the second lowest scatter and it too had a careful selection criteria in the choice of its sample.

As stated earlier non-circular motions are a potential source of observational scatter (see Section 10.2.2). They should only be substantial in the inner regions of a galaxy, within one effective radius. To test the effect of non-circular motions, the fitting routine was applied to data only exterior to one effective radius in each galaxy. Table 12.2 presents the results of this analysis, with an identical format as Table 12.1. It is clear that the effective radius cut is quite strong, removing $\sim 40\,\%$ of the data. The $\Delta_{g_\dagger}$ observational scatter does indeed go down after this cut to a median value of 0.14 dex. The relative uncertainty ($\delta V/V$) is higher in the inner regions due to lower V values, thus it is not surprising that the scatter reduces by some amount. Section 12.2.2 will show that this reduced scatter is still not small enough to explain with observational uncertainties alone.

Survey	$g_\dagger$ (10^{-10} m s^{-1})	$\sigma_{g_\dagger}$ (dex)	$\Delta_{g_\dagger}$ (dex)	N Galaxies	N Samples
SV	2.72 ± 0.13	0.150 ± 0.010	0.120 ± 0.013	19	318
SP	2.35 ± 0.04	0.122 ± 0.002	0.121 ± 0.006	113	1799
SF	2.17 ± 0.02	0.129 ± 0.001	0.125 ± 0.002	164	10080
C97	1.71 ± 0.02	0.159 ± 0.001	0.151 ± 0.003	293	8645
M92	2.45 ± 0.03	0.173 ± 0.002	0.156 ± 0.004	685	8543
M96	2.03 ± 0.02	0.174 ± 0.002	0.151 ± 0.003	1207	11968

Table 12.2: Parametrization for the RAR fit to each dataset with all radii below one R_e removed. For a description of the columns see Table 12.1.

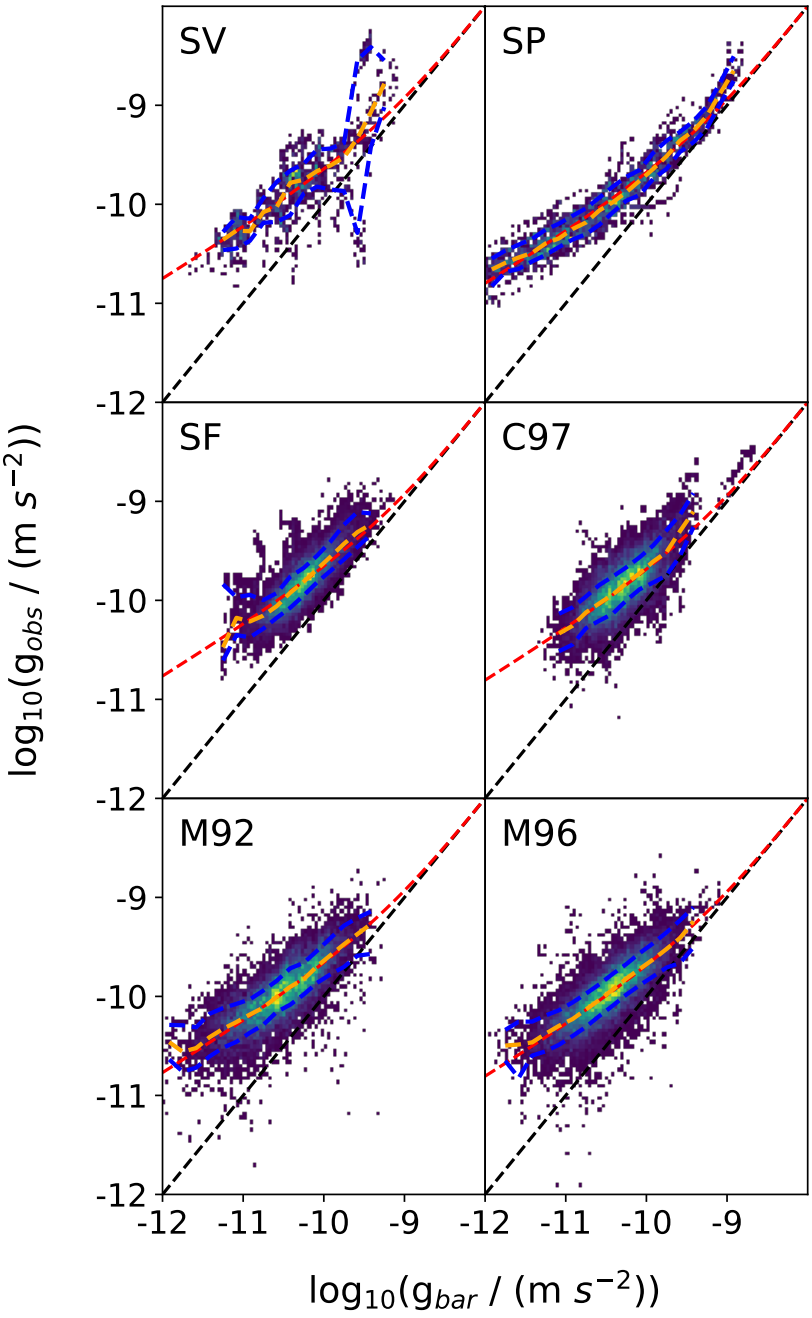

Figure 12.1: The RAR for each dataset with an independent fit of Eq. 10.2 in red. Orange shows the median, and blue represents a 68.3 % interval range for data binned in the x-axis.

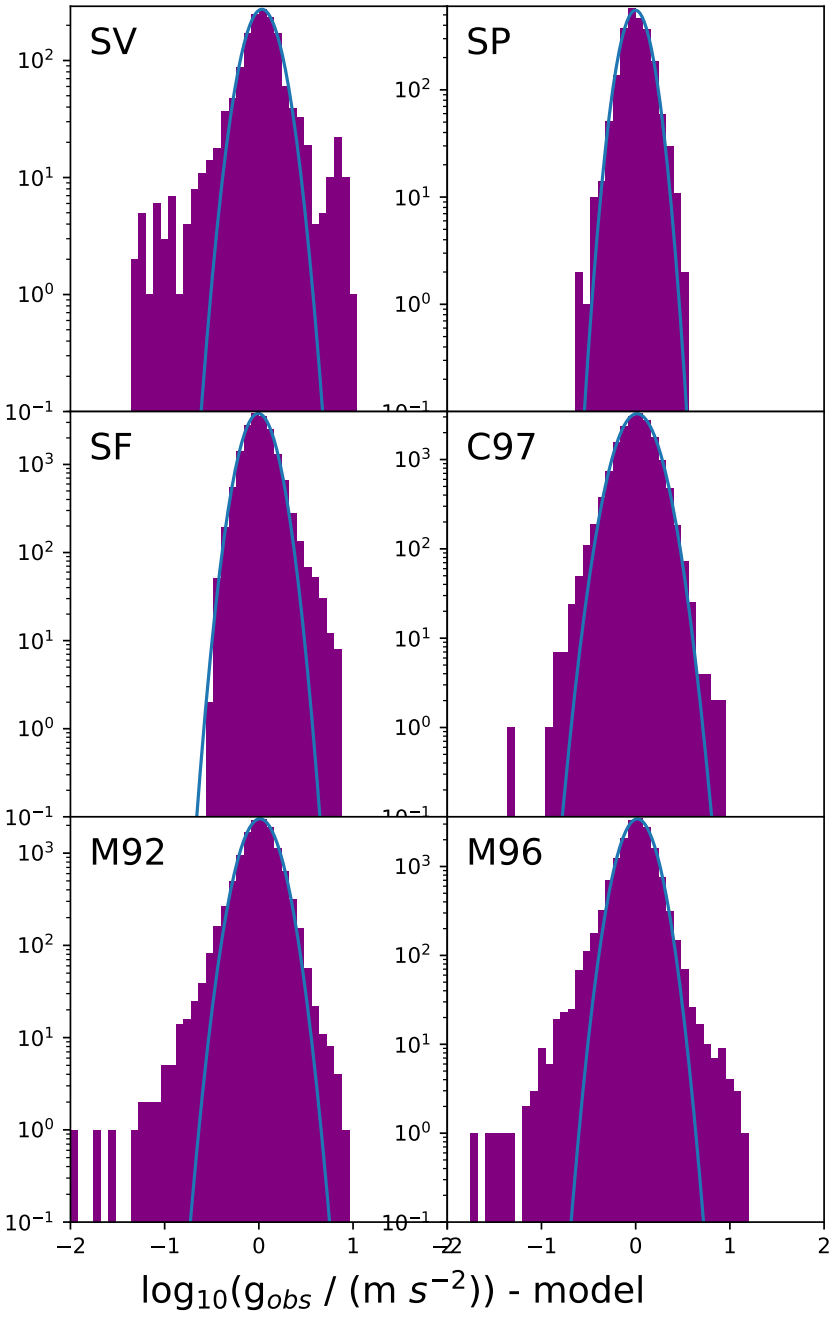

Figure 12.2: Histogram of residuals for each dataset. A Gaussian fit is overlaid in blue for comparison; the tails show non-Gaussianities of the RAR which can significantly impact scatter measures.

Parameter	SV	SP	SF	C97	M92	M96
Dynamical Mass	0.43	-0.03	0.10	0.21	0.25	0.33
Distance	0.00	-0.05	0.06	-0.06	0.05	0.05
Central Surface Brightness	-0.26	0.12	-0.01	-0.21	-0.12	-0.11
Morphological Type	-0.19	0.08	0.01	-0.08	-0.09	-0.03
Isophotal $R_{23.5}$	0.35	-0.02	0.05	-0.01	0.16	0.13
Luminosity	0.43	-0.17	0.10	0.05	0.09	0.12
Rotational Velocity	0.33	0.08	0.06	0.33	0.32	0.39
Inclination	0.11	-0.10	-0.03	0.08	0.17	0.12
Effective Radius	0.28	-0.03	0.11	0.06	0.09	0.08

Table 12.3: Pearson r correlation coefficients between the RAR scatter and several galaxy parameters. Each column represents one of the datasets while each row represents a galaxy parameter for which the correlation is computed. A significant and consistent correlation is not found for any parameter in the table.

12.1.2 Testing for Dependencies of the Residuals

One may now ask if the RAR fit scatters are correlated with any other variables. Considering how tight the RAR already is, any strong correlations would be remarkable. Plotted in Appendix C are the scatters as a function of several variables. To help visualize any trends, lines are overlaid indicating median and 68.3% intervals.

As a more quantitative measure, the Pearson r correlation coefficients are listed in Table 12.3, Appendix C presents plots of the residuals as well. Most of the parameters show no consistent correlation across all surveys, the outliers being dynamical mass and rotational velocity. These two parameters are very closely related as the dynamical mass is calculated from the velocity profile. The correlation with V is strongest, with a median correlation of 0.33 between the different surveys; this is not a strong correlation.

12.2 A Simulated RAR with Scatter

12.2.1 A Simulated Galaxy on the RAR

The central focus of this work is to determine if the scatter about the RAR can be explained by observational uncertainties. Evaluating those uncertainties can, however, be challenging. There are several sources of observational uncertainty, some of which are not best described using normal distributions. For example the distribution of corrected inclinations has a hard limit at 90°. Further, the observed quantities are transformed through several calculations before arriving at the quantities of interest which are the radial accelerations. These transformations can distort what was a normal distribution into a biased form. This is handled to first order in standard uncertainty analysis by multiplying the derivative in the expression when adding uncertainties. As the scatter is the quantity of interest for this analysis, a more advanced technique was desired.

Monte-Carlo sampling described in Section 10.5.3 allows one to fully account for all non-linearities in each transformation. Observational data are simply re-sampled and then put through the same calculations (and cuts) as the observed data to obtain new "observed" quantities. To generate these Monte-Carlo simulations a model is needed for the uncertainties at each point. Each study reports an uncertainty along with every velocity measurement and these are re-sampled in the simulation using a Gaussian distribution. Thus, if velocity data points on the outskirts of a rotation curve tend to have larger uncertainties, the Monte-Carlo simulation will properly reflect this. Although generally much smaller, the measurements of surface brightness for a light profile have associated uncertainties that are treated with a Gaussian as well.

The studies also report a distance for each galaxy, which has considerable uncertainty that can impact the RAR. SV and SP report uncertainties for some (SV) or all (SP) of the distance measurements, for the others there is no uncertainty included and 15 % distance uncertainty is used (see Section 10.3.3). To properly represent the distance uncertainty, a Gaussian is used in log space to scale the distance by the available uncertainty. The resulting distribution for distances is thus non-Gaussian in linear space; this also naturally avoids cases where the large distance uncertainties in cluster environments (SV, and some of SP) could scatter to negative values.

Inclination is given an uncertainty in two stages. First, the axis ratio is given an uncertainty using a Gaussian with $\sigma = 0.02$ corresponding roughly to a standard 5 % when combined with the q_0 uncertainty. The distribution is truncated at 0 and 1 as these are the limits of the parameter. Second, an uncertainty on the intrinsic thickness q_0 is used with a truncated Gaussian; centred on 0.2 with $\sigma = 0.03$ and truncated to be between 0.1 and 0.3 as this represents a conservatively large range in which a galaxy may fall (see Section 10.3.1). These are combined through Eq. 10.1 in the Monte-Carlo analysis. The mass-to-light ratio for each mass measurement is given a random uncertainty of 0.13 dex (Roediger and Courteau 2015). Systematic uncertainties in M_*/L can be as large as 0.3 dex which explains why the studies in this analysis are not combined. Each of these uncertainties are included in the Monte-Carlo analysis when re-sampling a point in the RAR.

Once a galaxy has been re-sampled, it can go through the same calculations as were performed on the observed data. This includes any data quality cuts, fits, and other non-linear components of the analysis. For example, a galaxy that initially passed the inclination cut ($i > 30°$ required) may randomly end up below the cutoff threshold in

the simulation and will be discarded. Similarly, re-sampled velocity points may move into or out of the allowed region ($\delta V/V < 0.1$ and $V > 10\,\mathrm{km}\,s^{-1}$). Some quantities are re-sampled for the whole galaxy, such as the distance and inclination, and thus represent a correlated uncertainty for all samples from that galaxy. Any scatter along the RAR may also be missed in a simplified analysis. Figure 12.3 shows the effect of re-sampling two galaxies, showing the simulated rotation curve and resulting position on the RAR. It can be seen that the data are sometimes scattered very far from the original position, as all points were initially on the RAR (see Section 12.2.2) before uncertainties were introduced. However, the data quality cuts generally eliminate points that scatter very far, reducing the resulting scatter. In this model it is clearly incorrect to state that each point on the RAR is independent. This also shows that it is important to include data quality cuts in the Monte-Carlo simulation.

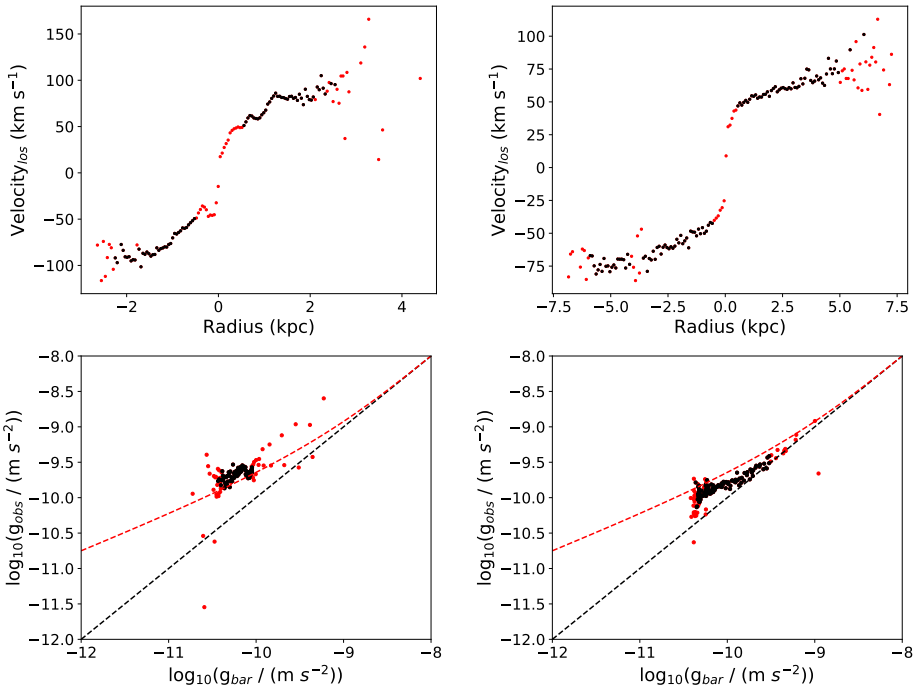

Figure 12.3: Two galaxies with simulated rotation curves and corresponding data points on the RAR. Points were forced to lay on the RAR before being re-sampled (see Section 12.2.2). The rotation curves align with the RAR plot below them, with SV VCC0801 on the left and VCC2058 on the right. All uncertainties are included, the cuts are modelled as well, with red points indicating ones that are removed from the analysis while black points passed all cuts.

12.2.2 Simulated RAR Under Zero Scatter Assumption

With a procedure for re-sampling the observed data, one may impose the zero scatter assumption and compare to observations. Putting each observation on the RAR requires choosing where on the relation it should go. One may simply keep one of g_{bar} or g_{obs} constant and move the point along the other axis, however the uncertainty in each axis is comparable and so neither could justifiably be chosen to remain constant. Instead, the nearest point on the RAR is used; this has the added benefit of moving

each point the minimum possible distance from its original observation. This point is taken as the "true" value from which observational uncertainties will be added. Once the new g_{bar} and g_{obs} are chosen, the original measurements must be altered to match. This is done by choosing a new velocity and mass-to-light ratio to re-adjust g_{obs} and g_{bar} respectively. Eq. 12.1 presents the calculation used to infer the velocity and mass-to-light ratio.

$$\begin{aligned} V_{obs} &= \sin(i)\sqrt{R g_{obs}} \\ M_*/L &= \frac{g_{bar} R^2}{GL}, \end{aligned} \tag{12.1}$$

which are simply the equations from Section 10.4 rearranged to find the measured values given acceleration. The choice of RAR parametrization used is the fitted value from Table 12.1 individually for each dataset. Note that the simulated baryonic M_*/L is chosen under the spherical potential assumption as inverting the multi-pole expansion used in the flat disk potential analysis would have added unnecessary complexity. Thus, uncertainties from the multi-pole expansion are not accounted for in this model; these are expected to be small as the expansion is taken to near the point of convergence ($< 2\,\%$ offset in tests with analytic potentials).

A new dataset is created from each study by re-sampling every point once. After the new dataset is generated, it can be treated just like the original observations with all of the same transformations and fits (except the flat disk potential). Figure 12.4 presents the simulated RAR which qualitatively matches with the observed data. For several of the studies one can see where individual galaxies have scattered coherently away from the relation, this is due to shared uncertainties such as distance and inclination. For each survey Figure 12.5 shows a histogram of the residuals; the simulated

RAR shows clear non-Gaussianity with extra observations in the tails. This effect becomes more substantial if the data quality cut $\delta V_{obs}/V < 0.1$ is relaxed. These non-Gaussianities reflect higher order effects in the g_{bar} and g_{obs} calculations, Figure 12.4 suggests that the non-linearities in g_{obs} can result in the greatest deviations from the RAR. A handful of points scatter more than an order of magnitude away from the RAR fit, this is seen in both the observed and simulated relation. Large scattering below the fit comes from velocity uncertainties which are squared in the g_{obs} calculation; this effect is mostly mitigated by the $V_{obs} > 10\,\mathrm{km\,s^{-1}}$ requirement and is more substantial if the cut is relaxed. Large scattering above the fit is mostly from distance uncertainties due to the $1/d$ in the g_{obs} calculation scattering near zero.

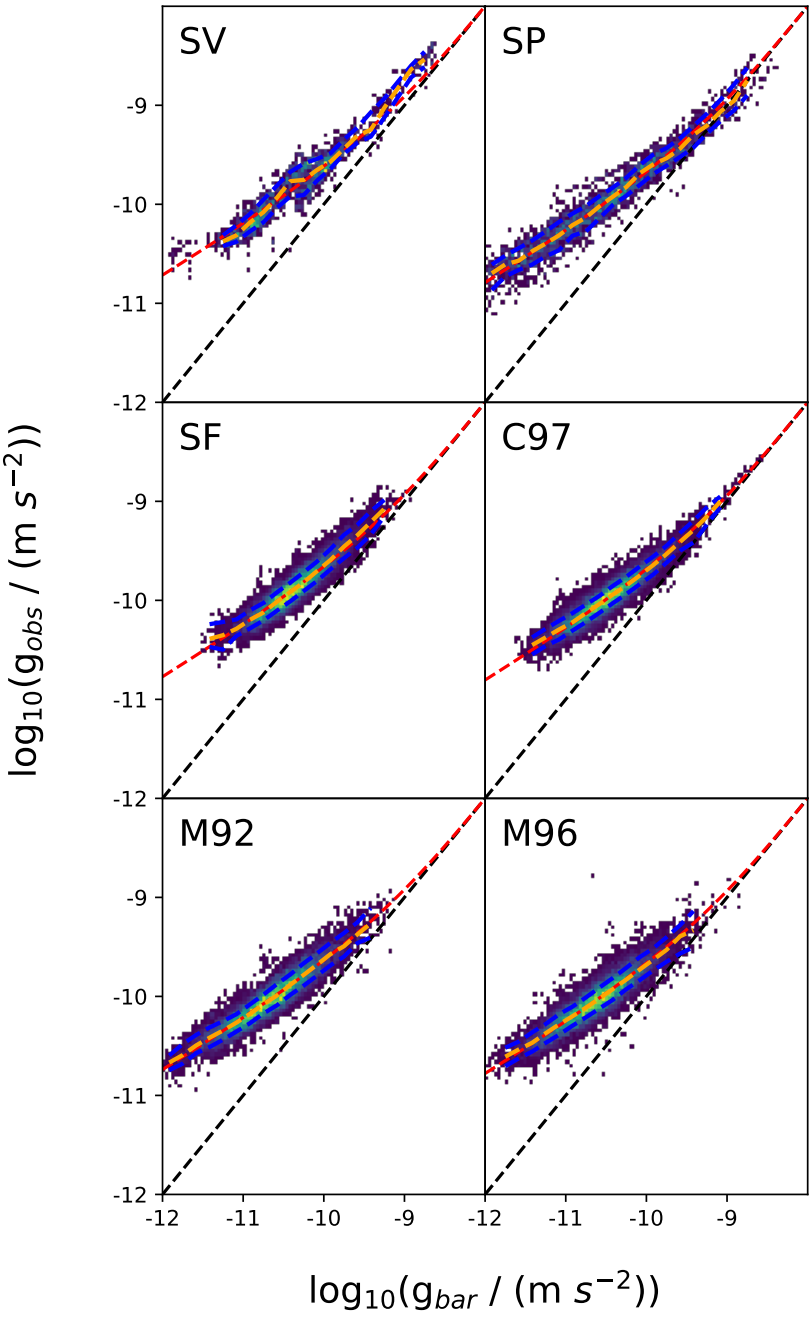

Figure 12.4: Simulated RAR sampled using values and uncertainties from each dataset.

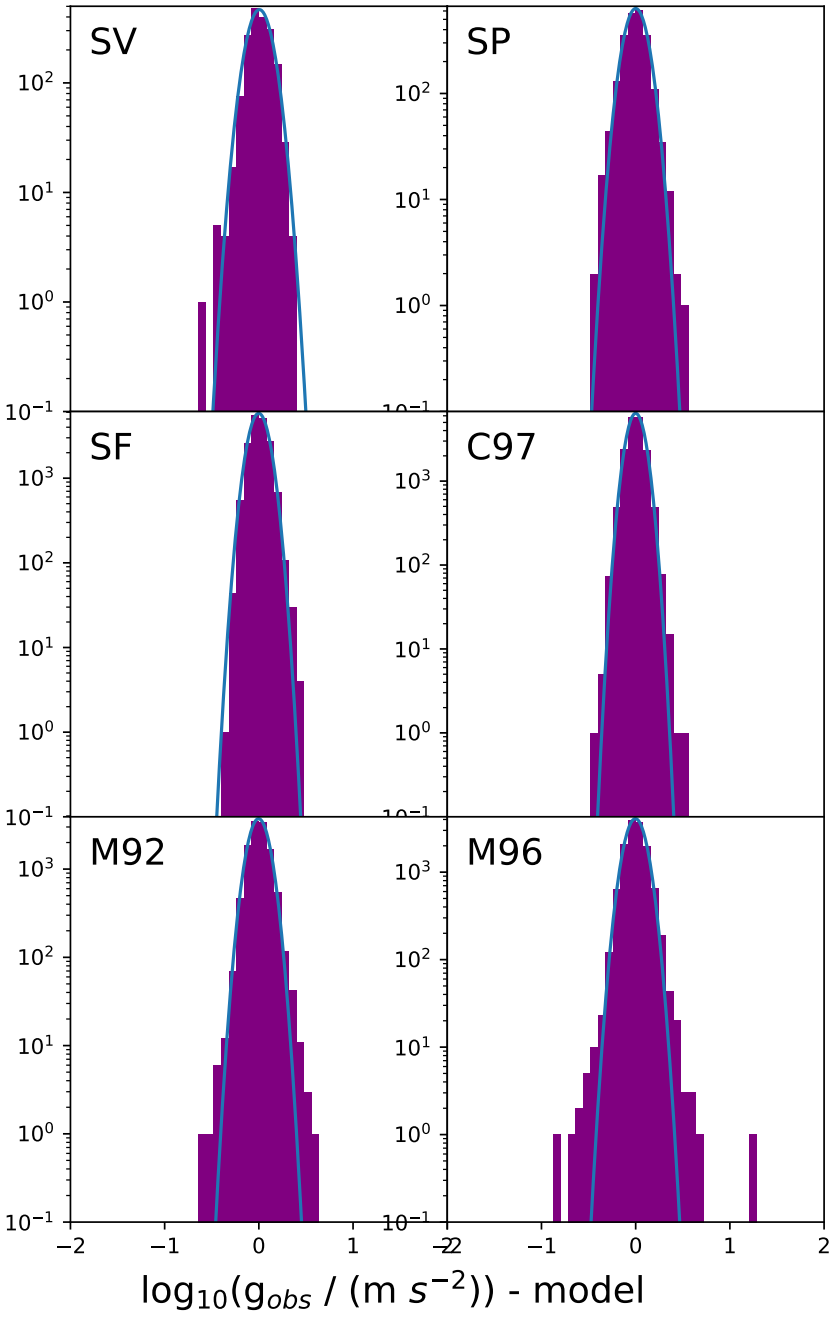

Figure 12.5: Histogram of residuals for each dataset. A Gaussian fit is overlaid in blue for comparison.

More quantitatively, Table 12.4 presents the results of fitting the simulated data in the same way as the observed data. The rows and columns are organized as in Table 12.1. The number of galaxies and number of samples in Table 12.4 is not the same as in Table 12.1 because the cuts are re-applied to the data after simulation. For every dataset, the scatter is below the scatter in the observed relation from Table 12.1, in all cases by more than 5σ, with a median $\sigma_{g_\dagger}$ scatter of 0.11 dex and median $\Delta_{g_\dagger}$ scatter of 0.10 dex. The tabulated uncertainties in the scatter must over-predict the confidence in this result as they only represent random uncertainties under the assumption that each data point is independent. The low uncertainties simply reflect the large amount of data in each study and indicate that if an identical procedure was performed many times, the same answer would result. The uncertainty in $\sigma_{g_\dagger}$ and $\Delta_{g_\dagger}$ can also be evaluated by running the Monte-Carlo simulation multiple times, doing this gives similar uncertainties as the bootstrap method, typically ~ 0.003 dex between the datasets with SV having larger values and M96 having smaller values. An avenue for future work would be to examine the effect of model choices on the scatter (i.e. disk flattening prescription) as this could cause changes larger than the bootstrap uncertainty. It is notable that the fitted values for $g_\dagger$ in the simulated data are different from the observed values for $g_\dagger$, despite the observed value being assumed during the simulation. In all cases, the simulated $g_\dagger$ value is above the observed value, this indicates a systematic shift induced by the RAR calculations. This may be an interesting avenue for future investigation as it is not the focus of this study.

The simulated and observed scatter can be compared to first order by subtracting in quadrature the $\sigma_{g_\dagger}$ (or $\Delta_{g_\dagger}$) values of the simulated RAR from the observed. Doing

Survey	$g_\dagger$ (10^{-10} m s^{-1})	$\sigma_{g_\dagger}$ (dex)	$\Delta_{g_\dagger}$ (dex)	N Galaxies	N Samples
SV	3.53 ± 0.06	0.117 ± 0.002	0.120 ± 0.005	36	1744
SP	2.48 ± 0.04	0.118 ± 0.002	0.113 ± 0.005	116	2259
SF	2.70 ± 0.01	0.093 ± 0.001	0.093 ± 0.002	165	17417
C97	2.30 ± 0.01	0.087 ± 0.001	0.084 ± 0.001	298	17511
M92	3.15 ± 0.02	0.102 ± 0.001	0.098 ± 0.002	698	11729
M96	2.61 ± 0.01	0.109 ± 0.001	0.102 ± 0.002	1225	13312

Table 12.4: Parametrization for the RAR fit to each simulated dataset. For a description of the columns see Table 12.1.

this yields a rough measure of the intrinsic scatter, as it is the component not explained by the observational uncertainties. The median value for the approximated intrinsic scatter is 0.16 dex using $\sigma_{g_\dagger}$, and it is 0.13 dex using $\Delta_{g_\dagger}$. The $\Delta_{g_\dagger}$ value should be considered more reliable. Compared to the simulated scatter, the intrinsic scatter accounts for $\sim 60\,\%$ of the observed scatter (using: $1 = (\frac{\Delta_{\rm intrinsic}}{\Delta_{\rm observed}})^2 + (\frac{\Delta_{\rm simulated}}{\Delta_{\rm observed}})^2$). Each of the surveys individually has an intrinsic scatter near 60 % of the total observed scatter except SP for which it is 26 % (inferred intrinsic scatter for each model: 56 %, 26 %, 59 %, 77 %, 67 %, and 61 % for SV, SP, SF, C97, M92, M96 respectively). This is consistent with approximately half of the observed RAR scatter being due to intrinsic galaxy variations and half due to observational uncertainties.

In Table 12.2 it was seen that the scatter reduced when data below one effective radius is removed; comparing to the simulated scatter in Table 12.4 it is clear that non-circular motions are not the cause of the discrepancy in scatter. In all cases, except SV, the simulated scatter is below the observed scatter outside one effective radius. SV has a $\Delta_{g_\dagger}$ scatter that is the same in both tables, however it has so little data (19 galaxies) when the effective radius cut is applied that it is not significant compared to all the other studies.

Parameter	SV	SP	SF	C97	M92	M96
Dynamical Mass	0.19	-0.08	-0.04	-0.04	0.03	-0.02
Distance	-0.08	-0.01	-0.15	-0.01	-0.05	-0.02
Central Surface Brightness	-0.16	0.02	-0.12	0.03	-0.04	-0.01
Morphological Type	-0.22	0.01	-0.08	-0.06	-0.06	0.02
Isophotal $R_{23.5}$	0.27	-0.05	-0.05	-0.02	0.06	0.01
Luminosity	0.23	-0.03	-0.01	-0.08	0.00	-0.03
Rotational Velocity	0.19	-0.11	0.03	-0.05	0.03	-0.02
Inclination	-0.03	0.04	0.05	0.21	0.13	0.11
Effective Radius	0.18	-0.10	-0.05	0.03	0.05	0.01

Table 12.5: Pearson r correlation coefficients between the RAR scatter and several galaxy parameters. Each column represents one of the datasets while each row represents a galaxy parameter for which the correlation is computed. A significant correlation is not found for any parameter in the table.

As with the observed data, it is possible to compare the simulated data to the simulated galaxy parameters to see if there are any correlations. This may indicate biases in the calculations that skew the relation away from the RAR systematically. Table 12.5 shows the Pearson r correlation values formatted identically to Table 12.3. As in the observed scatter, SV shows large correlations indicating that the small sample size is likely causing small fluctuations to appear as correlations. The rotational velocity is not seen to correlate as strongly in the simulated data as in the observed data, this may be a further avenue to explore in the future. For the most part, the simulated correlations are small, with fluctuations similar to those seen in the observed data.

Chapter 13

Summary and Conclusions

13.1 Summary

Scaling relations represent a critical tool for analyzing the structure and formation of galaxies as well as for cosmological studies. The Radial Acceleration Relation is a local scaling relation comparing the observed dynamics in a galaxy with that which would be expected from the visible matter alone. Lelli et al. (2017) suggest that the scatter in the relation is tight enough that it is consistent with zero intrinsic scatter for the relation. If true, this would be in strong tension with the ΛCDM paradigm. This thesis computed the RAR for an unprecedented large sample of galaxies, amounting to over 2500 galaxy observations and approximately 70,000 points on the RAR. This is more than an order of magnitude improvement over the analysis of Lelli et al. (2017). Robust fitting techniques are used to parametrize the RAR and bootstrap uncertainties determine the random error. To test the zero scatter hypothesis, a Monte-Carlo simulation is used. It is built on the zero intrinsic scatter assumption and produces the RAR as it would appear with only observational uncertainties. The

resulting scatter falls below the observed scatter for all surveys. Considering bootstrap determined random uncertainties, the differences are too large to be explained by random uncertainty. Some intrinsic scatter must be invoked to explain the distribution of galaxies in the RAR, this must be of order 0.13 dex to explain the observed scatter. This result is consistent with the ΛCDM paradigm, but unexpected in a MOND framework. To test for spurious correlations, the scatter in the RAR was tested against several galaxy parameters. Using the Pearson r correlation coefficient, we confirmed that no significant correlations exist in the scatter.

13.2 Future Work

We have clearly demonstrated that the intrinsic scatter of the RAR is non-zero. However, additional avenues may still be explored. For instance, Lelli et al. (2017) show that the stellar acceleration can be used instead of the baryonic acceleration in the RAR. This is surprising as in some galaxies the baryonic mass is less than 50 % stars. For this thesis, the stellar acceleration was indeed used as the surveys did not have resolved HI intensity maps. We could include the HI mass profiles in future work to verify that the scatter is unaffected.

Our study of galaxy light profiles may also benefit from bulge/disk decompositions to test if the scatter can be reduced by using separate mass-to-light ratios for each component. This should not strongly impact our results. SV and SF already have resolved colour profiles (i.e. variable M_*/L) and they also confirm our result of the RAR having intrinsic scatter. A more accurate model of the baryonic potential could still be generated, one with a spherical bulge and flattened disk. The current analysis assumes that the whole galaxy is flattened.

Further investigations into sources of uncertainty could be an enlightening avenue for future work. With studies of 2D velocity fields it has been found that sometimes the slit alignment for rotation curve measurement is incorrect by a few degrees. This systematically biases the rotation curve to be low, and may alter the position of a galaxy on the RAR (Aguerri et al. 2015). In this analysis only the reported rotation curve uncertainties were used to reflect these sorts of effects. However, it may be useful to include in the uncertainty model, a fraction of galaxies that are systematically biased to low velocities. Other coherent effects on the light profiles and rotation curves could be included in the uncertainty model as well, such as atmospheric smearing and non-circular motions. In this thesis, seeing effects are handled by simply removing points within 5 arcsec of the galaxy centre; yet these points typically also have high signal to noise and our analysis could be strengthened by their inclusion if modelled properly. In the outer regions of light profiles, properties such as the disk scale length used for modelling the gravitational potential also depend on the assumed disk fitting function and disk profile shape (Pohlen and Trujillo 2006; Bovy 2015). It may be instructive to look at the systematic effects of an assumed algorithm.

Perhaps the most interesting avenue for future investigation is the application of this technique to other scaling relations. Of great interest for simulations is the amount of scatter to include in sub-grid physics relations to match what occurs in nature. As the base observations are re-simulated, potentially any scaling relation which uses the measurements presented in this thesis can now be investigated with the same level of sophistication.

13.3 Conclusions

In this thesis, several datasets were collated to test the analysis of Lelli et al. (2017) who report an RAR scatter consistent with being intrinsically zero. Our analysis has over an order of magnitude more data than Lelli et al. (2017). Our method used a Monte-Carlo simulation that reflected the observational uncertainties in order to compare with the scatter in the measured relation. Several observational uncertainties were accounted for: velocity, inclination, mass-to-light ratio, distance, and surface brightness. An advantage of the Monte-Carlo simulation is that it can naturally handle complicated model uncertainties with correlations between parameters and highly non-linear calculations. The uncertainty on the scatter was evaluated using a bootstrap technique which better approximates the random uncertainty. These tools allowed for careful comparison of the observed and simulated scatter so as to produce a confident result. It is found that the intrinsic scatter of the RAR in spiral galaxies is in fact non-zero with a median intrinsic scatter of 0.13 dex, accounting for approximately half of the observed scatter. The median observed scatter was 0.17 dex while the median simulated scatter was 0.10 dex with bootstrap uncertainties of order 0.002 dex. Encouragingly, these techniques are not challenging to implement, and simply take advantage of the increased computing power available today. All things, and especially uncertainties, considered, the standard model of galaxy formation in ΛCDM is not challenged by the extensive study of the RAR presented in this thesis.

Bibliography

M. Aartsen, M. Ackermann, J. Adams, et al. Multimessenger observations of a flaring blazar coincident with high-energy neutrino icecube-170922a. *Science*, 361(6398), 2018. ISSN 0036-8075. doi: 10.1126/science.aat1378. URL `http://science.sciencemag.org/content/361/6398/eaat1378`.

B. P. Abbott, R. Abbott, T. D. Abbott, et al. Gw170817: Observation of gravitational waves from a binary neutron star inspiral. *Phys. Rev. Lett.*, 119:161101, Oct 2017. doi: 10.1103/PhysRevLett.119.161101. URL `https://link.aps.org/doi/10.1103/PhysRevLett.119.161101`.

B. Abolfathi, D. S. Aguado, G. Aguilar, et al. The Fourteenth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the Extended Baryon Oscillation Spectroscopic Survey and from the Second Phase of the Apache Point Observatory Galactic Evolution Experiment. *ApJS*, 235:42, April 2018. doi: 10.3847/1538-4365/aa9e8a.

S. Agostinelli, J. Allison, K. Amako, et al. Geant4a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250 – 303, 2003. ISSN

0168-9002. doi: https://doi.org/10.1016/S0168-9002(03)01368-8. URL `http://www.sciencedirect.com/science/article/pii/S0168900203013688`.

J. A. L. Aguerri, J. Méndez-Abreu, J. Falcón-Barroso, et al. Bar pattern speeds in CALIFA galaxies. I. Fast bars across the Hubble sequence. *A&A*, 576:A102, April 2015. doi: 10.1051/0004-6361/201423383.

D. S. Akerib, S. Alsum, H. M. Araújo, et al. Calibration, event reconstruction, data analysis, and limit calculation for the LUX dark matter experiment. *Phys. Rev. D*, 97:102008, May 2018. doi: 10.1103/PhysRevD.97.102008.

F. Allard, D. Homeier, and B. Freytag. Model Atmospheres From Very Low Mass Stars to Brown Dwarfs. In *16th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun*, volume 448, page 91, December 2011.

E. Aprile, J. Aalbers, F. Agostini, et al. Dark Matter Search Results from a One Tonne×Year Exposure of XENON1T. *ArXiv e-prints*, May 2018.

M. Arnaud. X-ray observations of clusters of galaxies. In F. Melchiorri and Y. Rephaeli, editors, *Background Microwave Radiation and Intracluster Cosmology*, volume 2005937974, page 77, 2005.

S. J. Asztalos, G. Carosi, C. Hagmann, et al. SQUID-Based Microwave Cavity Search for Dark-Matter Axions. *Physical Review Letters*, 104(4):041301, January 2010. doi: 10.1103/PhysRevLett.104.041301.

H. Baer, V. Barger, and D. Mickelson. How conventional measures overestimate electroweak fine-tuning in supersymmetric theory. *Phys. Rev. D*, 88(9):095013, November 2013. doi: 10.1103/PhysRevD.88.095013.

N. M. Ball and R. J. Brunner. Data Mining and Machine Learning in Astronomy. *International Journal of Modern Physics D*, 19:1049–1106, 2010. doi: 10.1142/S0218271810017160.

P. Benetti, F. Calaprice, E. Calligarich, et al. Measurement of the specific activity of 39ar in natural argon. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 574(1):83 – 88, 2007. ISSN 0168-9002. doi: https://doi.org/10.1016/j.nima.2007.01.106. URL http://www.sciencedirect.com/science/article/pii/S0168900207001672.

J. Bergstra and Y. Bengio. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb):281–305, 2012.

G. Bertelli, L. Girardi, P. Marigo, and E. Nasi. Scaled solar tracks and isochrones in a large region of the Z-Y plane. I. From the ZAMS to the TP-AGB end for 0.15-2.5 {M} stars. *A&A*, 484:815–830, June 2008. doi: 10.1051/0004-6361:20079165.

M. Bílek and S. Samurović. Investigating the radial acceleration relation in early-type galaxies using the Jeans analysis. *ArXiv e-prints*, November 2017.

J. Binney and S. Tremaine. *Galactic Dynamics: Second Edition*. Princeton University Press, 2008.

J. Bland-Hawthorn and O. Gerhard. The Galaxy in Context: Structural, Kinematic, and Integrated Properties. *ARA&A*, 54:529–596, September 2016. doi: 10.1146/annurev-astro-081915-023441.

P. T. Boggs, J. E. Rogers, R. B. Schnabel, and R. H. Byrd. Users reference guide for odrpack version 2.013, 1992. NISTIR 4834.

J. R. Bond and G. Efstathiou. Cosmic background radiation anisotropies in universes dominated by nonbaryonic dark matter. *ApJ*, 285:L45–L48, October 1984. doi: 10.1086/184362.

A. Boselli. *A Panchromatic View of Galaxies*. Wiley-VCH, December 2011.

A. Bosma and P. C. van der Kruit. The local mass-to-light ratio in spiral galaxies. *A&A*, 79:281–286, November 1979.

M. G. Boulay and A. Hime. Technique for direct detection of weakly interacting massive particles using scintillation time discrimination in liquid argon. *Astroparticle Physics*, 25:179–182, April 2006. doi: 10.1016/j.astropartphys.2005.12.009.

J. Bovy. galpy: A python Library for Galactic Dynamics. *ApJS*, 216:29, February 2015. doi: 10.1088/0067-0049/216/2/29.

J. Bovy and S. Tremaine. On the Local Dark Matter Density. *ApJ*, 756:89, September 2012. doi: 10.1088/0004-637X/756/1/89.

B. Broerman, M. Boulay, B. Cai, et al. Application of the tpb wavelength shifter to the deap-3600 spherical acrylic vessel inner surface. *Journal of Instrumentation*, 12(04): P04017, 2017. URL `http://stacks.iop.org/1748-0221/12/i=04/a=P04017`.

G. Bruzual and S. Charlot. Stellar population synthesis at the resolution of 2003. *MNRAS*, 344:1000–1028, October 2003. doi: 10.1046/j.1365-8711.2003.06897.x.

Y. I. Byun, K. C. Freeman, and N. D. Kylafis. Diagnostics of Dust Content in Spiral Galaxies: Numerical Simulations of Radiative Transfer. *ApJ*, 432:114, September 1994. doi: 10.1086/174553.

B. Cai, M. Boulay, B. Cleveland, and T. Pollmann. Surface backgrounds in the DEAP-3600 dark matter experiment. In R. Ford, editor, *American Institute of Physics Conference Series*, volume 1338 of *American Institute of Physics Conference Series*, pages 137–146, April 2011. doi: 10.1063/1.3579572.

T. Caldwell. Simulation of noble liquid detectors using rat, March 2014. URL `https://zzz.physics.umn.edu/lowrad/_media/meeting7/rat_aarm_2014.pdf`. talk at AARM meeting.

B. Carr and J. Silk. Primordial black holes as generators of cosmic structures. *MNRAS*, 478:3756–3775, August 2018. doi: 10.1093/mnras/sty1204.

G. F. Chapline and P. H. Frampton. A new direction for dark matter research: intermediate-mass compact halo objects. *Journal of Cosmology and Astroparticle Physics*, 2016(11):042, 2016. URL `http://stacks.iop.org/1475-7516/2016/i=11/a=042`.

D. Clowe, A. Gonzalez, and M. Markevitch. Weak-Lensing Mass Reconstruction of the Interacting Cluster 1E 0657-558: Direct Evidence for the Existence of Dark Matter. *ApJ*, 604:596–603, April 2004. doi: 10.1086/381970.

D. Clowe, M. Bradač, A. H. Gonzalez, et al. A Direct Empirical Proof of the Existence of Dark Matter. *ApJ*, 648:L109–L113, September 2006. doi: 10.1086/508162.

CMS Collaboration. Search for direct production of supersymmetric partners of the top quark in the all-jets final state in proton-proton collisions at sqrt(s) = 13 TeV. *ArXiv e-prints*, July 2017.

C. Conroy. Modeling the Panchromatic Spectral Energy Distributions of Galaxies. *ARA&A*, 51:393–455, August 2013. doi: 10.1146/annurev-astro-082812-141017.

F. Cooper, A. Khare, and U. Sukhatme. Supersymmetry and quantum mechanics. *Physics Reports*, 251(5):267 – 385, 1995. ISSN 0370-1573. doi: https://doi.org/10.1016/0370-1573(94)00080-M. URL `http://www.sciencedirect.com/science/article/pii/037015739400080M`.

C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, Sep 1995. ISSN 1573-0565. doi: 10.1007/BF00994018. URL `https://doi.org/10.1007/BF00994018`.

S. Courteau. Optical Rotation Curves and Linewidths for Tully-Fisher Applications. *AJ*, 114:2402, December 1997. doi: 10.1086/118656.

S. Courteau, A. A. Dutton, F. C. van den Bosch, et al. Scaling Relations of Spiral Galaxies. *ApJ*, 671:203–225, December 2007. doi: 10.1086/522193.

S. Courteau. Deep r-Band Photometry for Northern Spiral Galaxies. *The Astrophysical Journal Supplement Series*, 103:363, April 1996. doi: 10.1086/192281.

S. Courteau, J. A. Willick, M. A. Strauss, D. Schlegel, and M. Postman. Shellflow. I. The Convergence of the Velocity Field at 6000 Kilometers Per Second. *ApJ*, 544:636–640, December 2000. doi: 10.1086/317234.

P. Cushman, C. Galbiati, D. N. McKinsey, et al. Snowmass CF1 Summary: WIMP Dark Matter Direct Detection. *ArXiv e-prints*, October 2013.

A. C. Davidson and D. V. Hinkley. *Bootstrap Methods and their Application*. Cambridge Series in Statistical and Probabilistic Mathematics, October 1997. ISBN 9780521574716.

DEAP-3600 Collaboration, :, P.-A. Amaudruz, et al. First results from the DEAP-3600 dark matter search with argon at SNOLAB. *ArXiv e-prints*, July 2017.

T. J. DiCiccio and B. Efron. Bootstrap confidence intervals. *Statistical Science*, 11(3): 189–212, 1996. ISSN 08834237. URL `http://www.jstor.org/stable/2246110`.

S. Djorgovski and M. Davis. Fundamental properties of elliptical galaxies. *ApJ*, 313: 59–68, February 1987. doi: 10.1086/164948.

S. Dodelson and G. Efstathiou. Modern Cosmology. *Physics Today*, 57:60–61, July 2004. doi: 10.1063/1.1784308.

P. Domingos and M. Pazzani. On the optimality of the simple bayesian classifier under zero-one loss. *Machine learning*, 29(2):103–130, 1997.

B. Draine. Interstellar dust grains. *Annual Review of Astronomy and Astrophysics*, 41(1):241–289, 2003. doi: 10.1146/annurev.astro.41.011802.094840. URL `https://doi.org/10.1146/annurev.astro.41.011802.094840`.

F. Duncan, A. J. Noble, and D. Sinclair. The Construction and Anticipated Science of SNOLAB. *Annual Review of Nuclear and Particle Science*, 60:163–180, November 2010. doi: 10.1146/annurev.nucl.012809.104513.

A. A. Dutton, F. C. van den Bosch, A. Dekel, and S. Courteau. A Revised Model for the Formation of Disk Galaxies: Low Spin and Dark Halo Expansion. *ApJ*, 654: 27–52, January 2007. doi: 10.1086/509314.

A. A. Dutton, S. Courteau, R. de Jong, and C. Carignan. Mass Modeling of Disk Galaxies: Degeneracies, Constraints, and Adiabatic Contraction. *ApJ*, 619:218–242, January 2005. doi: 10.1086/426375.

A. A. Dutton, A. Obreja, L. Wang, et al. NIHAO XII: galactic uniformity in a ΛCDM universe. *MNRAS*, 467:4937–4950, June 2017. doi: 10.1093/mnras/stx458.

P. Erwin, J. E. Beckman, and M. Pohlen. Antitruncation of Disks in Early-Type Barred Galaxies. *ApJ*, 626:L81–L84, June 2005. doi: 10.1086/431739.

S. M. Faber and J. S. Gallagher. Masses and mass-to-light ratios of galaxies. *ARA&A*, 17:135–187, 1979. doi: 10.1146/annurev.aa.17.090179.001031.

B. Famaey, J. Khoury, and R. Penco. Emergence of the mass discrepancy-acceleration relation from dark matter-baryon interactions. *Journal of Cosmology and Astro-Particle Physics*, 2018:038, March 2018. doi: 10.1088/1475-7516/2018/03/038.

J. L. Feng. Dark Matter Candidates from Particle Physics and Methods of Detection. *ARA&A*, 48:495–545, September 2010. doi: 10.1146/annurev-astro-082708-101659.

K. Finlator, R. Davé, and B. D. Oppenheimer. Constraints on physical properties of z∼6 galaxies using cosmological hydrodynamic simulations. *MNRAS*, 376:1861–1878, April 2007. doi: 10.1111/j.1365-2966.2007.11578.x.

N. M. Förster Schreiber, R. Genzel, M. D. Lehnert, et al. SINFONI Integral Field Spectroscopy of z∼2 UV-selected Galaxies: Rotation Curves and Dynamical Evolution. *ApJ*, 645:1062–1075, July 2006. doi: 10.1086/504403.

W. L. Freedman and B. F. Madore. The Hubble Constant. *Annual Review*

of Astronomy and Astrophysics, 48:673–710, September 2010. doi: 10.1146/annurev-astro-082708-101829.

K. C. Freeman. On the Disks of Spiral and S0 Galaxies. *ApJ*, 160:811, June 1970. doi: 10.1086/150474.

K. C. Freeman. Galaxies, Globular Clusters, and Dark Matter. *Annual Review of Astronomy and Astrophysics*, 55:1–16, August 2017. doi: 10.1146/annurev-astro-091916-055249.

J. A. Frieman, M. S. Turner, and D. Huterer. Dark energy and the accelerating universe. *Annual Review of Astronomy and Astrophysics*, 46(1):385–432, 2008. doi: 10.1146/annurev.astro.46.060407.145243. URL `https://doi.org/10.1146/annurev.astro.46.060407.145243`.

Gaia Collaboration, A. G. A. Brown, A. Vallenari, et al. Gaia Data Release 2. Summary of the contents and survey properties. *ArXiv e-prints*, April 2018.

E. Garaldi, E. Romano-Díaz, C. Porciani, and M. S. Pawlowski. Radial Acceleration Relation of Λ CDM Satellite Galaxies. *Phys. Rev. Lett.*, 120:261301, June 2018. doi: 10.1103/PhysRevLett.120.261301.

C. Gilhuly and S. Courteau. An extensive photometric catalogue of CALIFA galaxies. *MNRAS*, 477:845–862, June 2018. doi: 10.1093/mnras/sty756.

M. Girardi, A. Biviano, G. Giuricin, F. Mardirossian, and M. Mezzetti. Velocity Dispersions in Galaxy Clusters. *ApJ*, 404:38, February 1993. doi: 10.1086/172256.

X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In Y. W. Teh and M. Titterington, editors, *Proceedings of the*

Thirteenth International Conference on Artificial Intelligence and Statistics, volume 9 of *Proceedings of Machine Learning Research*, pages 249–256, Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010. PMLR. URL `http://proceedings.mlr.press/v9/glorot10a.html`.

I. Goodfellow, J. Pouget-Abadie, M. Mirza, et al. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014. URL `http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf`.

A. Graves, A.-r. Mohamed, and G. Hinton. Speech recognition with deep recurrent neural networks. In *Acoustics, speech and signal processing (icassp), 2013 ieee international conference on*, pages 6645–6649. IEEE, 2013.

M. Hall, S. Courteau, A. A. Dutton, M. McDonald, and Y. Zhu. An investigation of Sloan Digital Sky Survey imaging data and multiband scaling relations of spiral galaxies. *MNRAS*, 425:2741–2765, October 2012. doi: 10.1111/j.1365-2966.2012.21290.x.

E. Hayashi and J. F. Navarro. Hiding cusps in cores: kinematics of disc galaxies in triaxial dark matter haloes. *MNRAS*, 373:1117–1124, December 2006. doi: 10.1111/j.1365-2966.2006.10927.x.

M. P. Haynes and R. Giovanelli. Neutral hydrogen in isolated galaxies. IV. Results for the Arecibo sample. *AJ*, 89:758–800, June 1984. doi: 10.1086/113573.

J. Heisler, S. Tremaine, and J. N. Bahcall. Estimating the masses of galaxy groups -

Alternatives to the virial theorem. *ApJ*, 298:8–17, November 1985. doi: 10.1086/163584.

L. Hernquist and J. P. Ostriker. A self-consistent field method for galactic dynamics. *ApJ*, 386:375–397, February 1992. doi: 10.1086/171025.

A. Hitachi, T. Takahashi, N. Funayama, et al. Effect of ionization density on the time dependence of luminescence from liquid argon and xenon. *Phys. Rev. B*, 27: 5279–5285, May 1983. doi: 10.1103/PhysRevB.27.5279. URL `https://link.aps.org/doi/10.1103/PhysRevB.27.5279`.

P. W. Hodge. The extragalactic distance scale. *Annual Review of Astronomy and Astrophysics*, 19(1):357–372, 1981. doi: 10.1146/annurev.aa.19.090181.002041. URL `https://doi.org/10.1146/annurev.aa.19.090181.002041`.

D. W. Hogg, J. Bovy, and D. Lang. Data analysis recipes: Fitting a model to data. *ArXiv e-prints*, August 2010.

E. Holmberg. On the Apparent Diameters and the Orientation in Space of Extragalactic Nebulae. *Meddelanden fran Lunds Astronomiska Observatorium Serie II*, 117:3–82, January 1946.

E. P. Hubble. Extragalactic nebulae. *ApJ*, 64, December 1926. doi: 10.1086/143018.

E. Hubble. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15:168–173, March 1929. doi: 10.1073/pnas.15.3.168.

T. Isobe, E. D. Feigelson, M. G. Akritas, and G. J. Babu. Linear regression in astronomy. *ApJ*, 364:104–113, November 1990. doi: 10.1086/169390.

Ž. Ivezić, S. M. Kahn, J. A. Tyson, et al. LSST: from Science Drivers to Reference Design and Anticipated Data Products. *ArXiv e-prints*, May 2008.

Ž. Ivezić, T. C. Beers, and M. Jurić. Galactic Stellar Populations in the Era of the Sloan Digital Sky Survey and Other Large Surveys. *ARA&A*, 50:251–304, September 2012. doi: 10.1146/annurev-astro-081811-125504.

C. Jillings and DEAP Collaboration. Control of contamination of radon-daughters in the DEAP-3600 acrylic vessel. In L. Miramonti and L. Pandola, editors, *American Institute of Physics Conference Series*, volume 1549 of *American Institute of Physics Conference Series*, pages 86–89, August 2013. doi: 10.1063/1.4818082.

C. J. Jillings and DEAP Collaboration. Dark matter experiment using Argon pulse-shape discrimination: DEAP-3600. Design and assembly. *Hyperfine Interactions*, 237:97, December 2016. doi: 10.1007/s10751-016-1305-6.

E. Jones, T. Oliphant, P. Peterson, et al. SciPy: Open source scientific tools for Python, 2001. URL `http://www.scipy.org/`. [Online; accessed July 2018].

R. C. Kennicutt, Jr. The Global Schmidt Law in Star-forming Galaxies. *ApJ*, 498: 541–552, May 1998. doi: 10.1086/305588.

S. M. Kent. CCD surface photometry of field galaxies. II. Bulge/disk decompositions. *The Astrophysical Journal Supplement Series*, 59:115–159, October 1985. doi: 10.1086/191066.

R. Kessler, A. C. Becker, D. Cinabro, et al. First-Year Sloan Digital Sky Survey-II Supernova Results: Hubble Diagram and Cosmological Parameters. *The Astrophysical*

Journal Supplement Series, 185:32–84, November 2009. doi: 10.1088/0067-0049/185/1/32.

V. V. Kiselev and S. A. Timofeev. Cosmological extrapolation of modified Newtonian dynamics. *Classical and Quantum Gravity*, 29(6):065015, March 2012. doi: 10.1088/0264-9381/29/6/065015.

M. Kowalski, D. Rubin, G. Aldering, et al. Improved Cosmological Constraints from New, Old, and Combined Supernova Data Sets. *ApJ*, 686:749–778, October 2008. doi: 10.1086/589937.

A. V. Kravtsov and S. Borgani. Formation of Galaxy Clusters. *ARA&A*, 50:353–409, September 2012. doi: 10.1146/annurev-astro-081811-125502.

A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

D. P. Kroese, T. Brereton, T. Taimre, and Z. I. Botev. Why the monte-carlo method is so important today. *Wiley Interdisciplinary Reviews: Computational Statistics*, 6 (6):386–392, 2014. doi: 10.1002/wics.1314. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/wics.1314`.

Y. Laberge. Advising on research methods: A consultant's companion. *Journal of Applied Statistics*, 38(12):2991–2991, 2011. doi: 10.1080/02664763.2011.559375. URL `https://doi.org/10.1080/02664763.2011.559375`.

B. Lehnert. Backgrounds in the DEAP-3600 Dark Matter Experiment. *ArXiv e-prints*, art. arXiv:1805.06073, May 2018.

F. Lelli, S. S. McGaugh, and J. M. Schombert. SPARC: Mass Models for 175 Disk Galaxies with Spitzer Photometry and Accurate Rotation Curves. *AJ*, 152:157, December 2016. doi: 10.3847/0004-6256/152/6/157.

F. Lelli, S. S. McGaugh, J. M. Schombert, and M. S. Pawlowski. One Law to Rule Them All: The Radial Acceleration Relation of Galaxies. *ApJ*, 836:152, February 2017. doi: 10.3847/1538-4357/836/2/152.

E. Levin and M. Fleisher. Accelerated learning in layered neural networks. *Complex systems*, 2:625–640, 1988.

D. N. C. Lin and J. E. Pringle. The Formation of the Exponential Disk in Spiral Galaxies. *ApJ*, 320:L87, September 1987. doi: 10.1086/184981.

T. Lindner. DEAP-3600 Data Acquisition System. In *Journal of Physics Conference Series*, volume 664, page 082026, December 2015. doi: 10.1088/1742-6596/664/8/082026.

L. P. Martins and P. Coelho. Testing the accuracy of synthetic stellar libraries. *MNRAS*, 381:1329–1346, November 2007. doi: 10.1111/j.1365-2966.2007.11954.x.

W. G. Mathews. The enormous mass of the elliptical galaxy M87 - A model for the extended X-ray source. *ApJ*, 219:413–423, January 1978. doi: 10.1086/155794.

D. S. Mathewson and V. L. Ford. Parameters of 2447 Southern Spiral Galaxies for Use in the Tully-Fisher Relation. *ApJS*, 107:97, November 1996. doi: 10.1086/192356.

D. S. Mathewson, V. L. Ford, and M. Buchhorn. A southern sky survey of the peculiar velocities of 1355 spiral galaxies. *ApJS*, 81:413–659, August 1992. doi: 10.1086/191700.

S. S. McGaugh. MOND Prediction for the Velocity Dispersion of the "Feeble Giant" Crater II. *ApJ*, 832:L8, November 2016. doi: 10.3847/2041-8205/832/1/L8.

M. Milgrom. A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis. *ApJ*, 270:365–370, July 1983. doi: 10.1086/161130.

M. Milgrom. Mond theory. *Canadian Journal of Physics*, 93:107–118, feb 2015. doi: 10.1139/cjp-2014-0211.

J. J. Mohr, R. Armstrong, E. Bertin, et al. The Dark Energy Survey data processing and calibration system. In *Software and Cyberinfrastructure for Astronomy II*, volume 8451, page 84510D, September 2012. doi: 10.1117/12.926785.

A. Y. Ng and M. I. Jordan. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In *Advances in neural information processing systems*, pages 841–848, 2002.

K. G. Noeske, B. J. Weiner, S. M. Faber, et al. Star Formation in AEGIS Field Galaxies since z=1.1: The Dominance of Gradually Declining Star Formation, and the Main Sequence of Star-forming Galaxies. *ApJ*, 660:L43–L46, May 2007. doi: 10.1086/517926.

K. S. Olsen. Analysis of alpha backgrounds in the DEAP-1 detector. *ArXiv e-prints*, art. arXiv:0906.0348, June 2009.

N. N.-Q. Ouellette, S. Courteau, J. A. Holtzman, et al. The Spectroscopy and H-band Imaging of Virgo Cluster Galaxies (SHIVir) Survey: Scaling Relations and the Stellar-to-total Mass Relation. *ApJ*, 843:74, July 2017. doi: 10.3847/1538-4357/aa74b1.

B. Paczynski. Gravitational microlensing by the galactic halo. *ApJ*, 304:1–5, May 1986. doi: 10.1086/164140.

H. Pagels and J. R. Primack. Supersymmetry, cosmology, and new physics at teraelectronvolt energies. *Physical Review Letters*, 48:223–226, January 1982. doi: 10.1103/PhysRevLett.48.223.

C. Papovich, S. L. Finkelstein, H. C. Ferguson, J. M. Lotz, and M. Giavalisco. The rising star formation histories of distant galaxies and implications for gas accretion with time. *MNRAS*, 412:1123–1136, April 2011. doi: 10.1111/j.1365-2966.2010.17965.x.

C. Park and Y.-Y. Choi. Morphology Segregation of Galaxies in Color-Color Gradient Space. *ApJ*, 635:L29–L32, December 2005. doi: 10.1086/499243.

C. Patrignani. Review of particle physics. *Chinese Physics C*, 40(10):100001, 2016. URL `http://stacks.iop.org/1674-1137/40/i=10/a=100001`. Particle Data Group.

J. A. Peacock, S. Cole, P. Norberg, et al. A measurement of the cosmological mass density from clustering in the 2df galaxy redshift survey. *Nature*, 410(6825):169, 2001.

R. D. Peccei. The strong cp problem and axions. In *Axions*, pages 3–17. Springer, 2008.

F. Pedregosa, G. Varoquaux, A. Gramfort, et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

S. Perlmutter, G. Aldering, G. Goldhaber, et al. Measurements of Ω and Λ from 42 High-Redshift Supernovae. *ApJ*, 517:565–586, June 1999. doi: 10.1086/307221.

Planck Collaboration, R. Adam, P. A. R. Ade, et al. Planck 2015 results. I. Overview of products and scientific results. *A&A*, 594:A1, September 2016a. doi: 10.1051/0004-6361/201527101.

Planck Collaboration, P. A. R. Ade, N. Aghanim, et al. Planck 2015 results. XIII. Cosmological parameters. *A&A*, 594:A13, September 2016b. doi: 10.1051/0004-6361/201525830.

M. Pohlen and I. Trujillo. The structure of galactic disks. Studying late-type spiral galaxies using SDSS. *A&A*, 454:759–772, August 2006. doi: 10.1051/0004-6361:20064883.

T. Pollmann. *Alpha backgrounds in the DEAP Dark Matter search experiment*. PhD thesis, Queen's University (Canada), 2012.

T. Pollmann, M. Boulay, and M. Kuźniak. Scintillation of thin tetraphenyl butadiene films under alpha particle excitation. *Nuclear Instruments and Methods in Physics Research A*, 635:127–130, April 2011. doi: 10.1016/j.nima.2011.01.045.

G. D. Racca, R. Laureijs, L. Stagnaro, et al. The Euclid mission design. In *Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave*, volume 9904 of *Proc. SPIE*, page 99040O, July 2016. doi: 10.1117/12.2230762.

M. Ramella, W. Boschin, D. Fadda, and M. Nonino. Finding galaxy clusters using voronoi tessellations. *Astronomy & Astrophysics*, 368(3):776–786, 2001.

R. M. Reddick, R. H. Wechsler, J. L. Tinker, and P. S. Behroozi. The Connection between Galaxies and Dark Matter Structures in the Local Universe. *ApJ*, 771:30, July 2013. doi: 10.1088/0004-637X/771/1/30.

C. Rethmeier. Monte-carlo position reconstruction analysis for deap-3600, May 2017. URL `https://indico.cern.ch/event/593812/contributions/2499758/`. CAP Congress.

A. G. Riess, A. V. Filippenko, P. Challis, et al. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *AJ*, 116: 1009–1038, September 1998. doi: 10.1086/300499.

A. G. Riess, L. M. Macri, S. L. Hoffmann, et al. A 2.4% determination of the local value of the hubble constant. *The Astrophysical Journal*, 826(1):56, 2016. URL `http://stacks.iop.org/0004-637X/826/i=1/a=56`.

D. C. Rodrigues, V. Marra, A. del Popolo, and Z. Davari. Absence of a fundamental acceleration scale in galaxies. *Nature Astronomy*, page 95, June 2018. doi: 10.1038/s41550-018-0498-9.

J. C. Roediger and S. Courteau. On the uncertainties of stellar mass estimates via colour measurements. *MNRAS*, 452:3209–3225, September 2015. doi: 10.1093/mnras/stv1499.

J. C. Roediger, S. Courteau, P. Sánchez- Blázquez, and M. McDonald. Stellar Populations and Radial Migrations in Virgo Disk Galaxies. *ApJ*, 758:41, October 2012. doi: 10.1088/0004-637X/758/1/41.

V. C. Rubin and W. K. Ford, Jr. Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *ApJ*, 159:379, February 1970. doi: 10.1086/150317.

B. S. Ryden. The Ellipticity of the Disks of Spiral Galaxies. *ApJ*, 601:214–220, January 2004. doi: 10.1086/380437.

P. Sánchez-Blázquez, R. F. Peletier, J. Jiménez- Vicente, et al. Medium-resolution Isaac Newton Telescope library of empirical spectra. *MNRAS*, 371:703–718, September 2006. doi: 10.1111/j.1365-2966.2006.10699.x.

R. H. Sanders and S. S. McGaugh. Modified Newtonian Dynamics as an Alternative to Dark Matter. *ARA&A*, 40:263–317, 2002. doi: 10.1146/annurev.astro.40.060401.093923.

P. Santini, A. Fontana, M. Castellano, et al. The Star Formation Main Sequence in the Hubble Space Telescope Frontier Fields. *ApJ*, 847:76, September 2017. doi: 10.3847/1538-4357/aa8874.

M. Schmidt. The Rate of Star Formation. *ApJ*, 129:243, March 1959. doi: 10.1086/146614.

P. Schneider. *Extragalactic Astronomy and Cosmology*. 2006.

R. H. M. Schoenmakers, M. Franx, and P. T. de Zeeuw. Measuring non-axisymmetry in spiral galaxies. *MNRAS*, 292:349–364, December 1997. doi: 10.1093/mnras/292.2.349.

S. B.-D. Shai Shalev-Shwartz. *Understanding Machine Learning: From Theory to Algorithms*. Cambrige University Press, July 2014. ISBN 9781107057135.

P. Sikivie. Dark Matter Axions. *International Journal of Modern Physics A*, 25: 554–563, 2010. doi: 10.1142/S0217751X10048846.

P. Sikivie and Q. Yang. Bose-einstein condensation of dark matter axions. *Phys. Rev. Lett.*, 103:111301, Sep 2009. doi: 10.1103/PhysRevLett.103.111301. URL `https://link.aps.org/doi/10.1103/PhysRevLett.103.111301`.

Y. Sofue and V. Rubin. Rotation Curves of Spiral Galaxies. *Annual Review of Astronomy and Astrophysics*, 39:137–174, January 2001. doi: 10.1146/annurev.astro.39.1.137.

K. Spekkens and J. A. Sellwood. Modeling Noncircular Motions in Disk Galaxies: Application to NGC 2976. *ApJ*, 664:204–214, July 2007. doi: 10.1086/518471.

V. Springel, C. S. Frenk, and S. D. M. White. The large-scale structure of the Universe. *Nature*, 440:1137–1144, April 2006. doi: 10.1038/nature04805.

J. A. Suykens and J. Vandewalle. Least squares support vector machine classifiers. *Neural processing letters*, 9(3):293–300, 1999.

C. Szegedy, W. Liu, Y. Jia, et al. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

K. P. Thompson. The Nature of Length, Area, and Volume in Taxicab Geometry. *ArXiv e-prints*, January 2011.

S. Tong and D. Koller. Support vector machine active learning with applications to text classification. *Journal of machine learning research*, 2(Nov):45–66, 2001.

A. Toomre. On the Distribution of Matter Within Highly Flattened Galaxies. *ApJ*, 138:385, August 1963. doi: 10.1086/147653.

C. Trachternach, W. J. G. de Blok, F. Walter, E. Brinks, and J. Kennicutt, R. C. Dynamical Centers and Noncircular Motions in THINGS Galaxies: Implications for Dark Matter Halos. *AJ*, 136:2720–2760, December 2008. doi: 10.1088/0004-6256/136/6/2720.

S. Trujillo-Gomez, A. Klypin, J. Primack, and A. J. Romanowsky. Galaxies in ΛCDM with Halo Abundance Matching: Luminosity-Velocity Relation, Baryonic Mass-Velocity Relation, Velocity Function, and Clustering. *ApJ*, 742:16, November 2011. doi: 10.1088/0004-637X/742/1/16.

R. B. Tully and J. R. Fisher. A new method of determining distances to galaxies. *A&A*, 54:661–673, February 1977.

S. van den Bergh. The Early History of Dark Matter. *Publications of the Astronomical Society of the Pacific*, 111:657–660, June 1999. doi: 10.1086/316369.

P. C. van der Kruit and L. Searle. Surface photometry of edge-on spiral galaxies. I - A model for the three-dimensional distribution of light in galactic disks. *A&A*, 95: 105–115, February 1981.

A. Vazdekis, P. Sánchez-Blázquez, J. Falcón-Barroso, et al. Evolutionary stellar population synthesis with MILES - I. The base models and a new line index system. *MNRAS*, 404:1639–1671, June 2010. doi: 10.1111/j.1365-2966.2010.16407.x.

S. Wang, Y.-F. Wang, Q.-G. Huang, and T. G. F. Li. Constraints on the Primordial Black Hole Abundance from the First Advanced LIGO Observation Run Using

the Stochastic Gravitational-Wave Background. *Phys. Rev. Lett.*, 120:191102, May 2018. doi: 10.1103/PhysRevLett.120.191102.

S. Weinberg. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. 1972.

S. Weinberg. A new light boson? *Phys. Rev. Lett.*, 40:223–226, Jan 1978. doi: 10.1103/PhysRevLett.40.223. URL `https://link.aps.org/doi/10.1103/PhysRevLett.40.223`.

K. Q. Weinberger and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, 10(Feb):207–244, 2009.

C. Wheeler, P. F. Hopkins, and O. Doré. The radial acceleration relation is a natural consequence of the baryonic Tully-Fisher relation. *ArXiv e-prints*, art. arXiv:1803.01849, March 2018.

S. D. M. White, C. S. Frenk, and M. Davis. Clustering in a neutrino-dominated universe. *ApJ*, 274:L1–L5, November 1983. doi: 10.1086/184139.

J. A. Willick. Measurement of galaxy distances. In *Formation of Structure in the Universe*, page 213, April 1999.

D. H. Wolpert and W. G. Macready. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82, Apr 1997. ISSN 1089-778X. doi: 10.1109/4235.585893.

W. M. Wood-Vasey, G. Miknaitis, C. W. Stubbs, et al. Observational Constraints

on the Nature of Dark Energy: First Cosmological Results from the ESSENCE Supernova Survey. *ApJ*, 666:694–715, September 2007. doi: 10.1086/518642.

J. Xu, C. Stanford, S. Westerdale, et al. First measurement of surface nuclear recoil background for argon dark matter searches. *Phys. Rev. D*, 96:061101, Sep 2017. doi: 10.1103/PhysRevD.96.061101. URL `https://link.aps.org/doi/10.1103/PhysRevD.96.061101`.

Y. B. Zel'dovich. The cosmological constant and the theory of elementary particles. *Soviet Physics Uspekhi*, 11(3):381, 1968. URL `http://stacks.iop.org/0038-5670/11/i=3/a=A13`.

H.-X. Zhang, T. H. Puzia, and D. R. Weisz. The Impact of Star Formation Histories on Stellar Mass Estimation: Implications from the Local Group Dwarf Galaxies. *ApJS*, 233:13, November 2017. doi: 10.3847/1538-4365/aa937b.

F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6:110–127, 1933.

Appendix A

Comparison of Dataset Measured Quantities

Presented here are several measured and inferred quantities for the galaxies in each dataset. The plots all have six panels, one for each survey. Within each panel is a histogram with all six studies sharing the same x-axis, the y-axis (representing counts) is allowed to vary for each panel. One must keep in mind the large range in the number of samples for each survey with SV having 46 galaxies while M96 has over 900. These plots allow for visual comparison between each survey to demonstrate the extensive parameter space covered by the combined analysis. Figure A.9 shows the location of all galaxies on the sky, with colours indicating the dataset for each point.

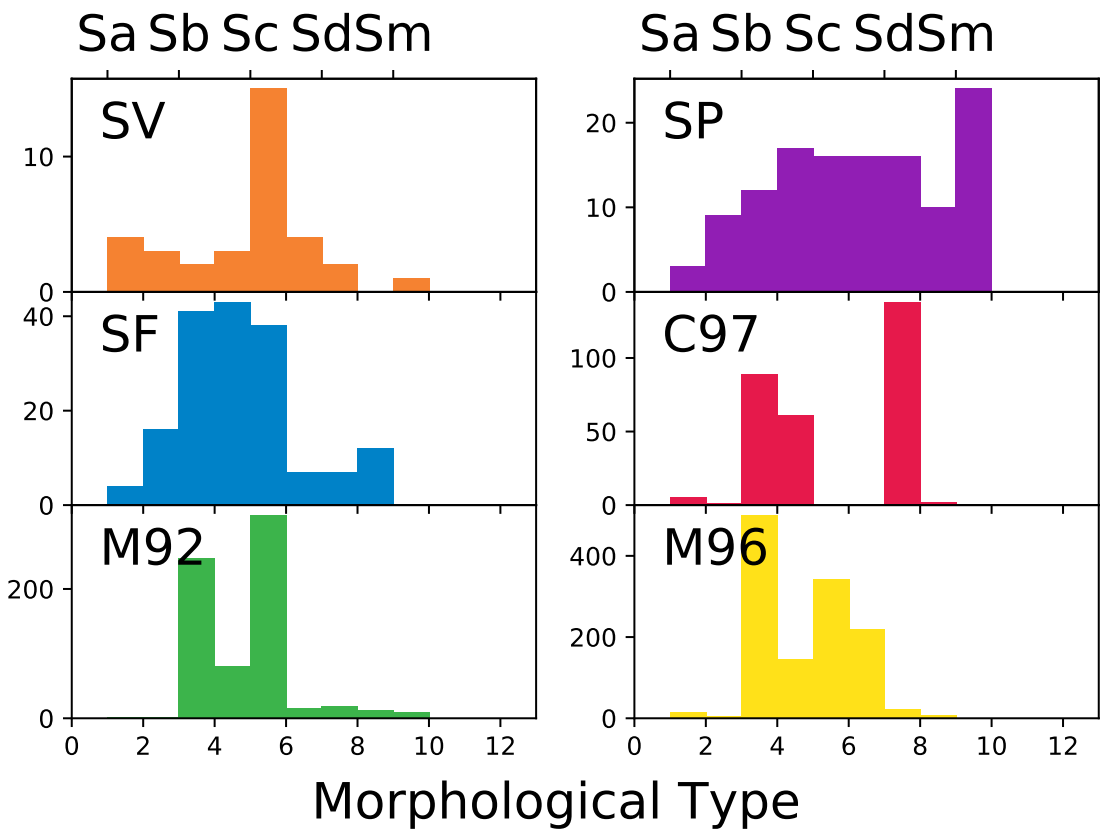

Figure A.1: Morphological type distribution for each dataset. SP clearly samples a broad range of morphological types. SV also has broad representation, while the others focus mostly on Sb, Sc, and some Sd.

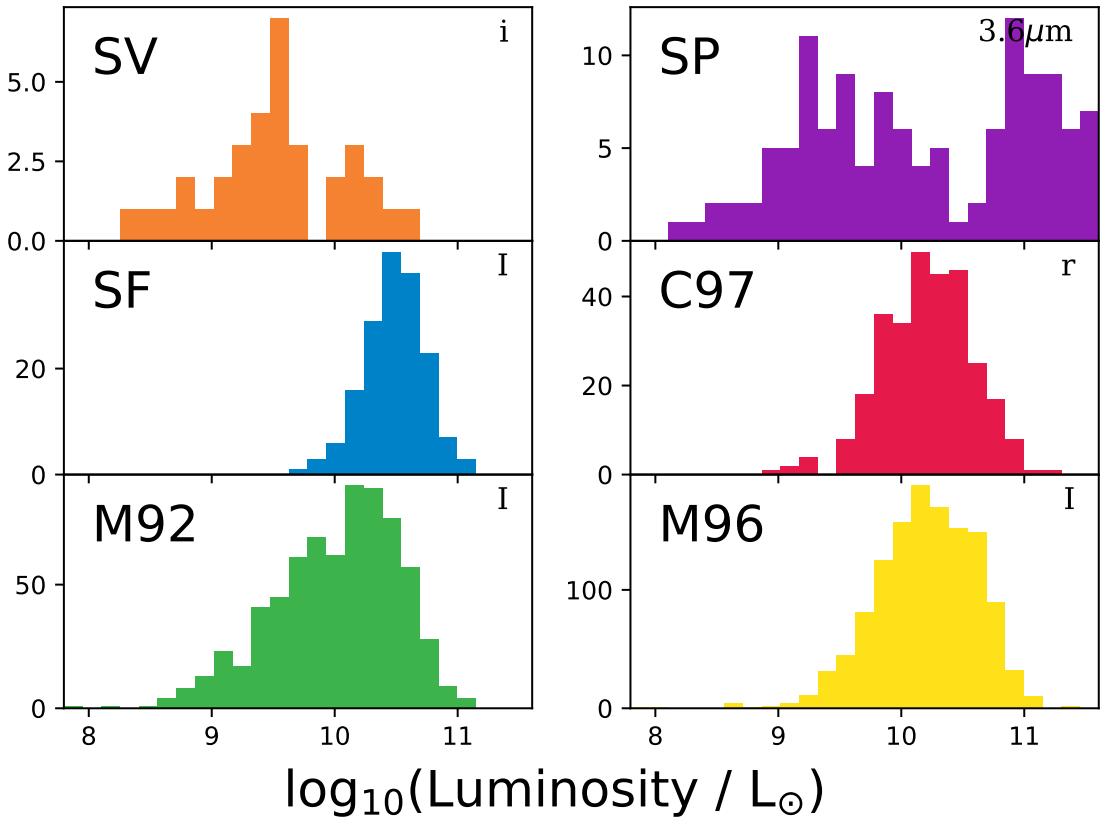

Figure A.2: The photometric band for each survey is identified in the upper right corner of each panel. As these match only for three surveys (in I-band), a direct comparison of luminosities between all the datasets cannot be made. However, one notes that SV and SP cover a broad range of luminosities. SF has a narrow distribution by design from their strict selection criteria.

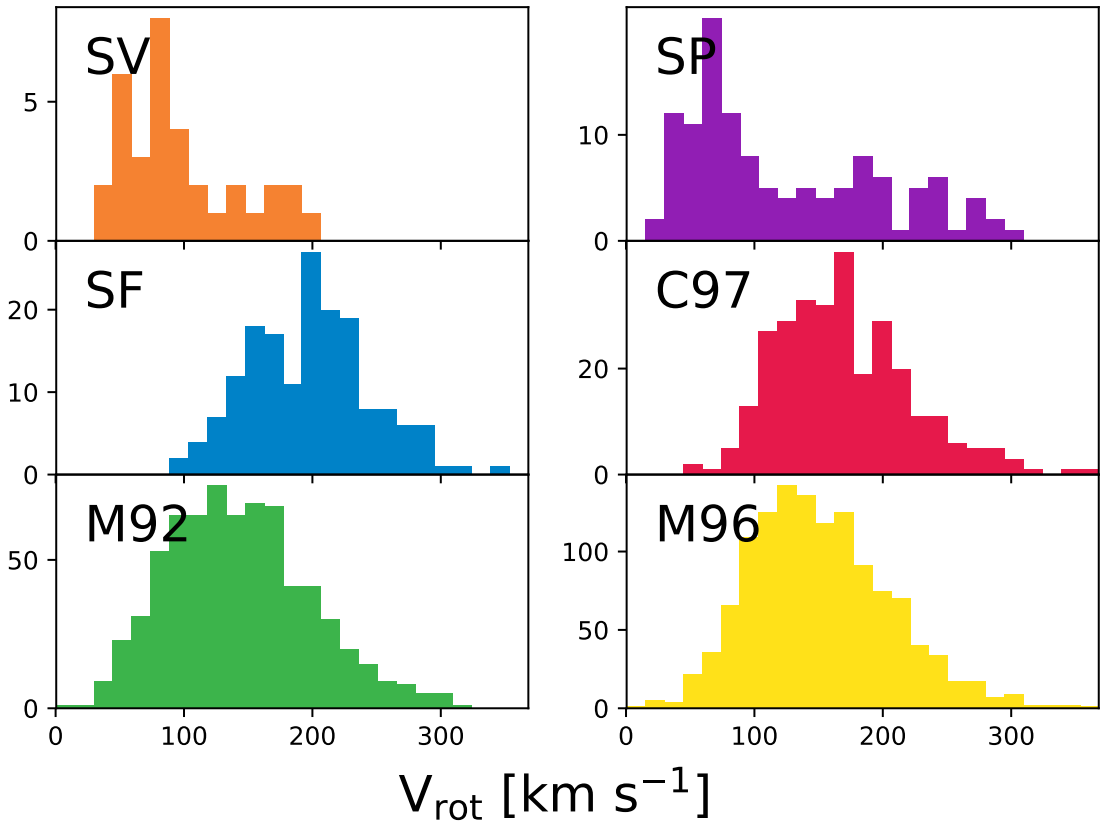

Figure A.3: Rotational velocity at the effective radius. SV and SP show a bias for slowly rotating objects while SF and C97 sample slightly faster rotators. M92 and M96 have broad distributions that peak around 120 km s^{-1}

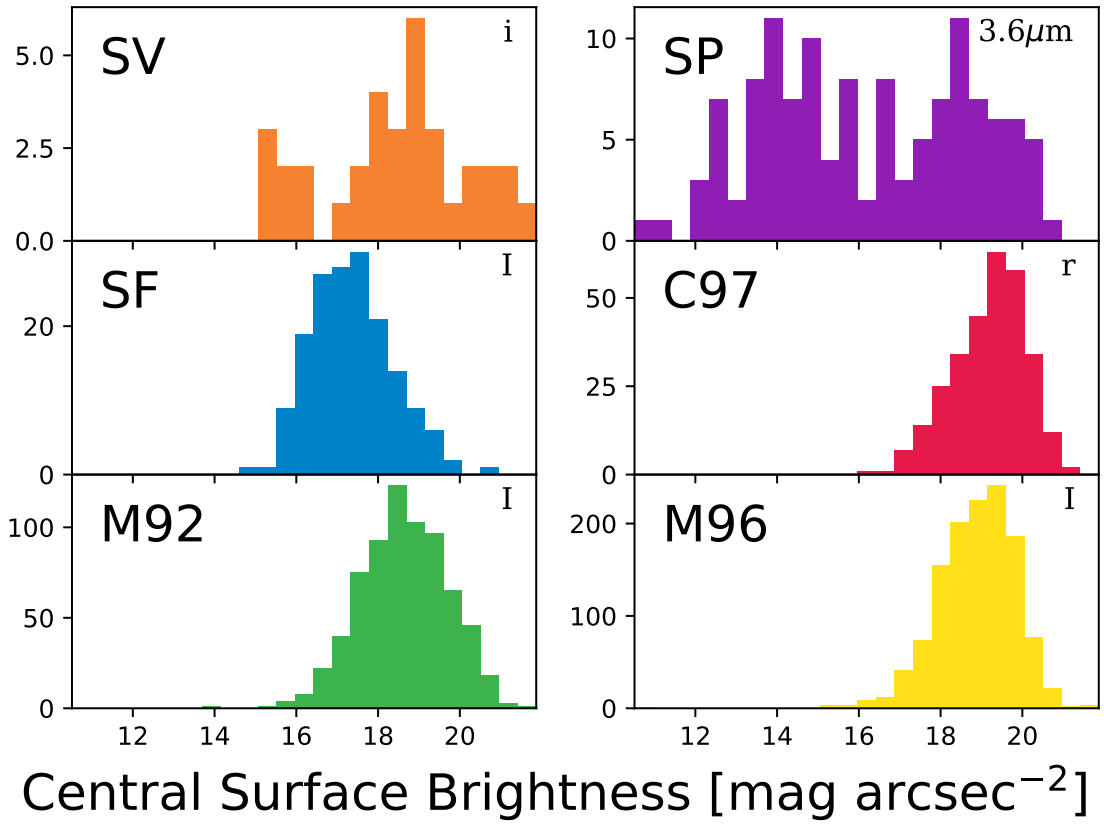

Figure A.4: The photometric band for each survey is identified in the upper right corner of each panel. As these match only for three surveys (in I-band), a direct comparison of surface brightnesss between all the studies cannot be made. However, one notes that SV and SP have broad distributions with SP showing two peaks, likely due to selection effects and not a true distribution. Compared to M92 and M96 (all I-band), SF has a slightly lower average central surface brightness.

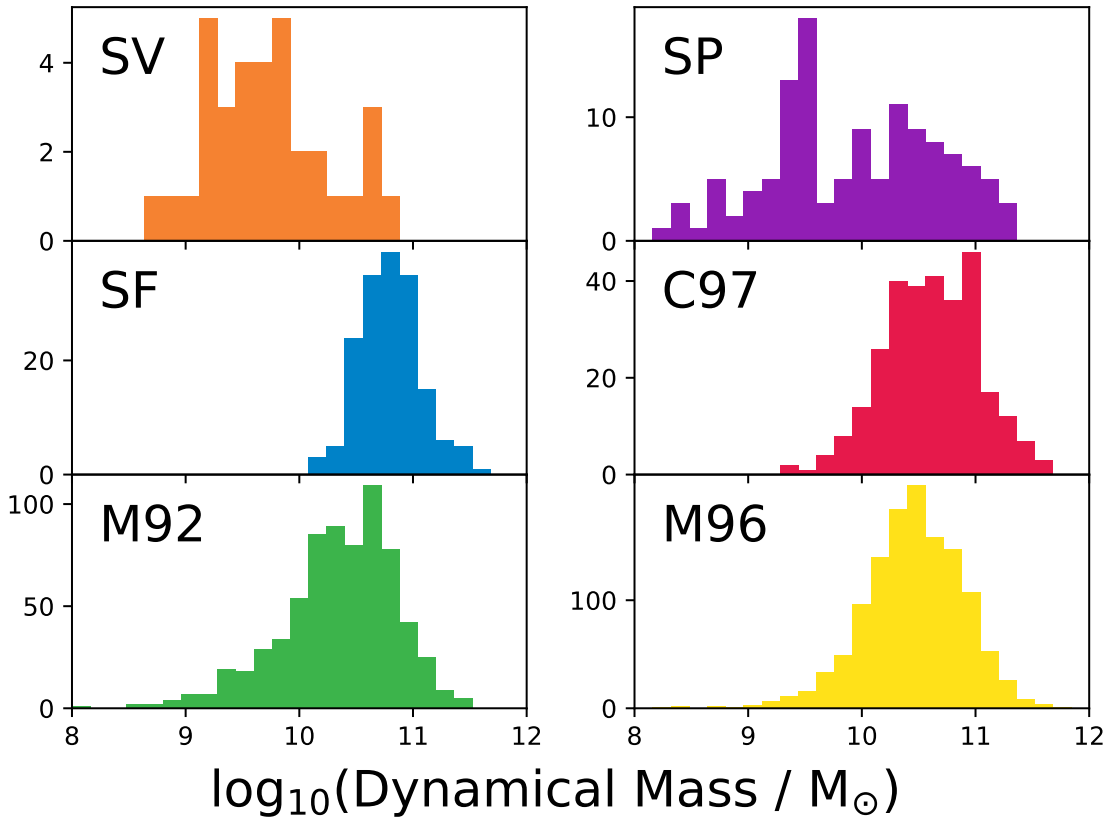

Figure A.5: Dynamical mass evaluated at the effective radius. As with luminosity, SF masses fall in a tight range compared to the others. SP has the widest range while C97, M92, and M96 are very nearly the same with a peak around $10^{10.5}\,M_\odot$.

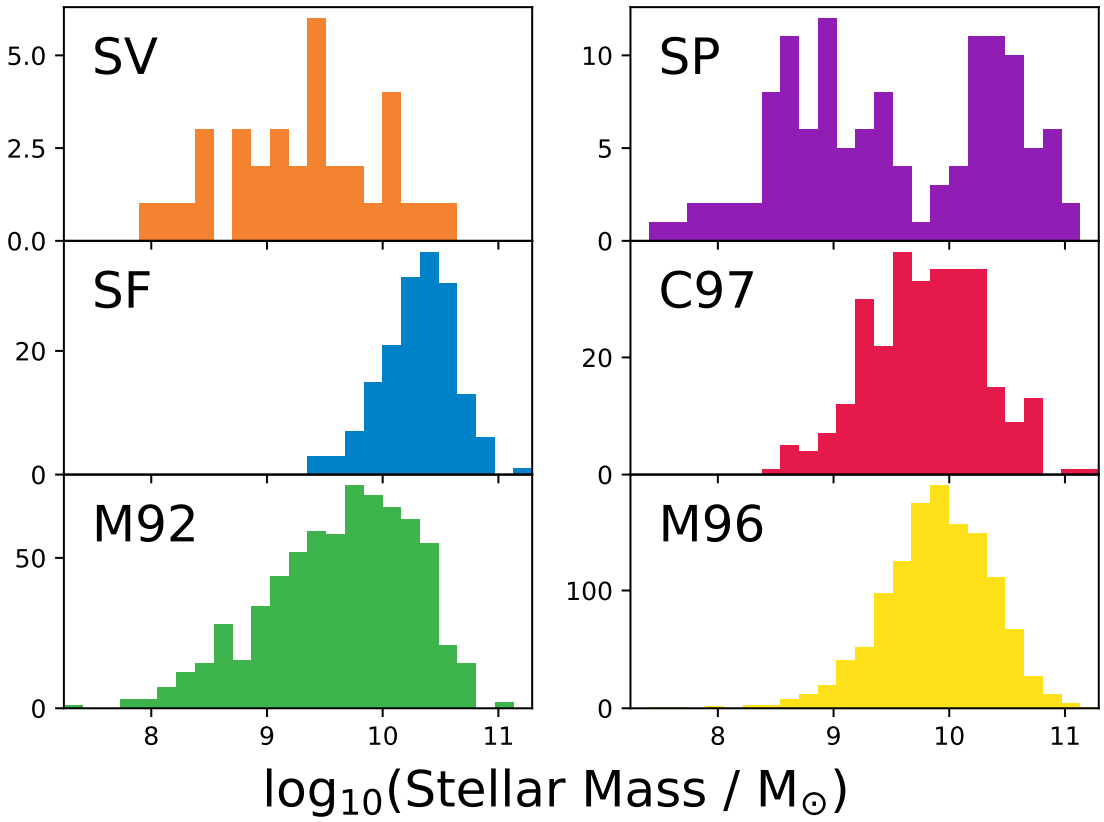

Figure A.6: Stellar mass evaluated at the effective radius. As with luminosity, SF masses fall in a tight range compared to the others. SP has the widest range and appears to have two peaks. C97, M92, and M96 are very nearly the same with a peak slightly above $10^{10}\,M_{\odot}$. SV probes some of the lowest stellar mass systems in the sample.

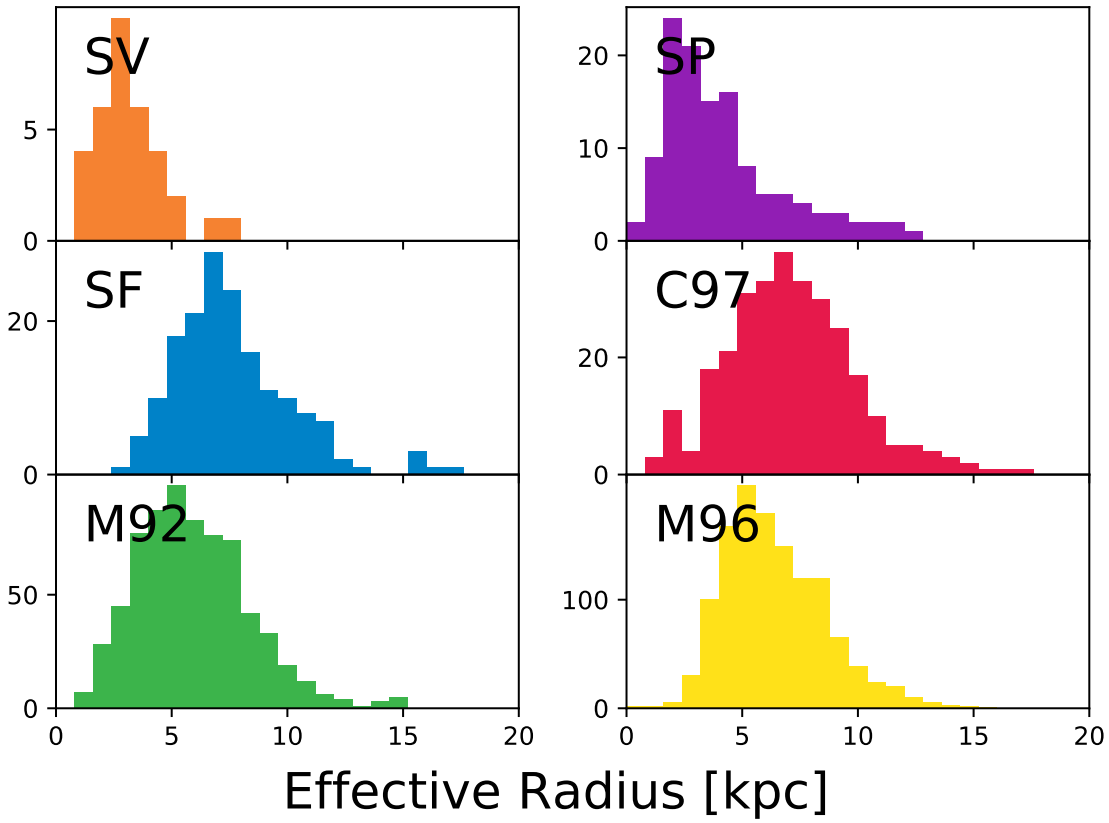

Figure A.7: SV and SP have low effective radius values compared to the other datasets. The other four are quite similar with SF and C97 being slightly higher than M92 and M96.

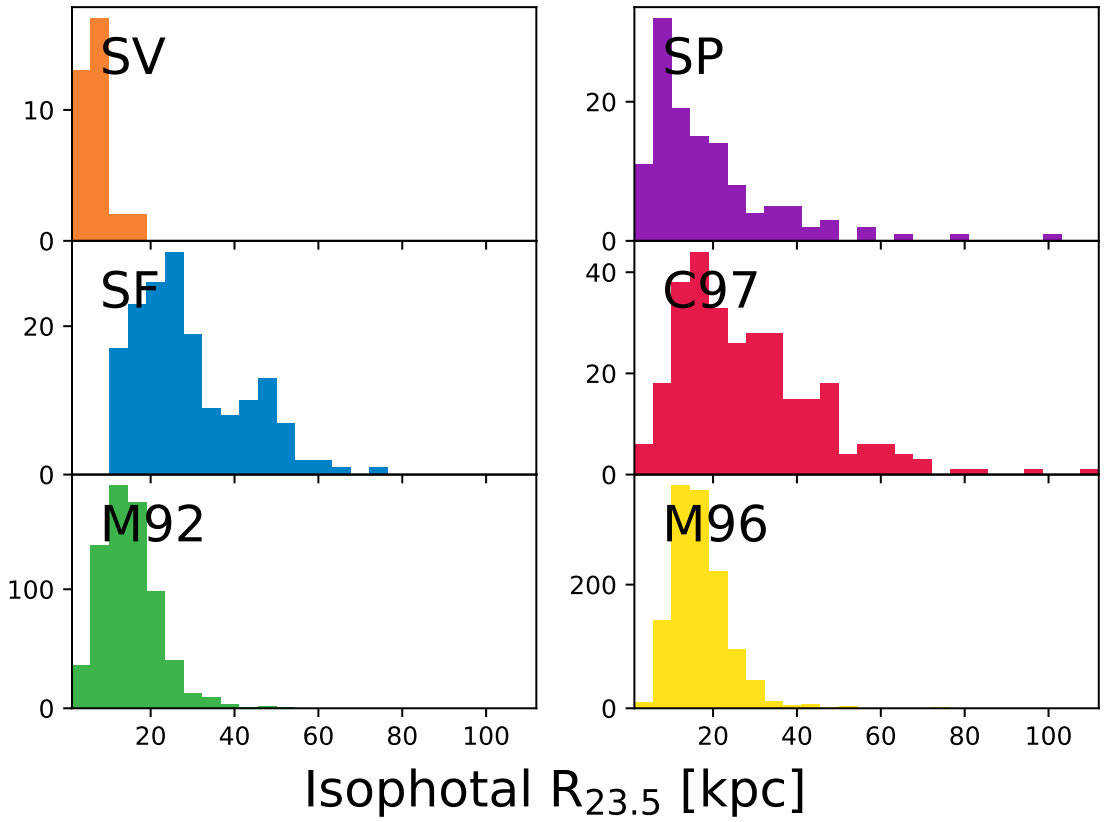

Figure A.8: SV and SP have the smallest galaxies in the sample while SF and C97 have galaxies which extend to very large radii. M92 and M96 have a narrow distribution in an intermediate range.

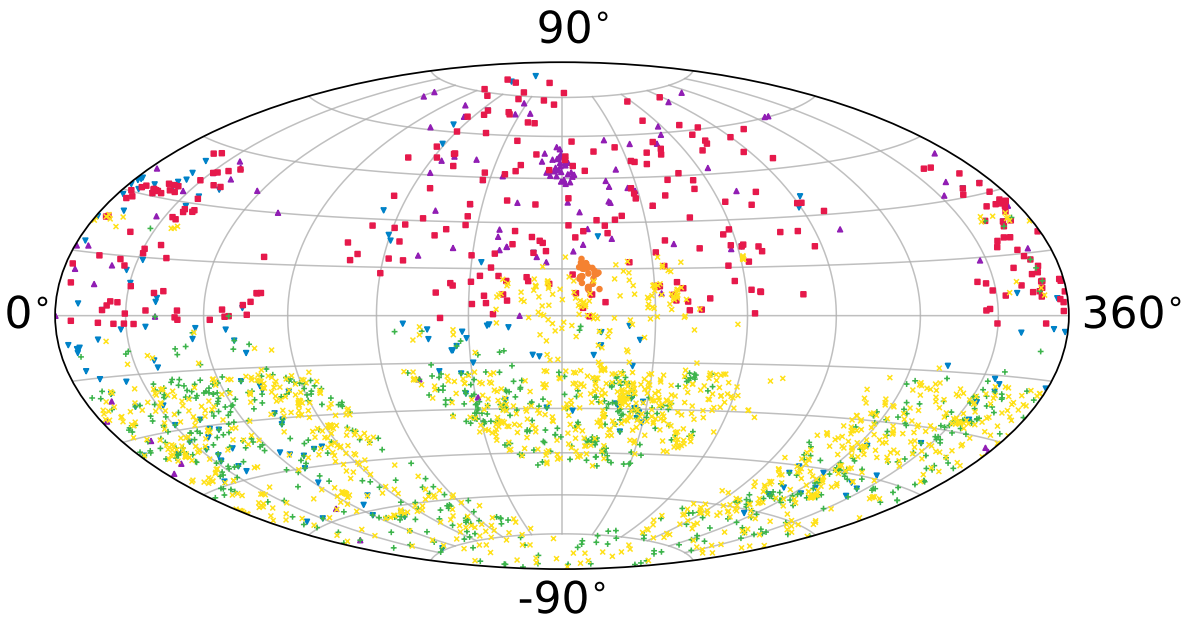

Figure A.9: Position of all galaxies for each study in a Hammer projection with Equatorial coordinates. Colours are the same as in previous plots, a different symbol is used for each dataset to aid visibility.

Appendix B

Scaling Relations for Each Dataset

This Appendix examines the correlations between available galaxy structural parameters. Each figure shows, for a single survey, a matrix of parameter relations. Along each axis is the: absolute magnitude, morphological type, $R_{23.5}$, effective radius, rotation velocity, central surface brightness, and the inclination. Luminosity is not shown as it is a direct function of absolute magnitude. The quantities: $R_{23.5}$, effective radius, and rotation velocity are plotted in log as this is the most common format for scaling relations. The lower triangle has 2D histograms showing the comparison for the data used in this thesis. The diagonal has a 1D histogram for the corresponding parameter, as in Appendix A. The upper triangle indicates the Pearson r correlation coefficient with a colour also scaled by the correlation (red for positive, blue for negative).

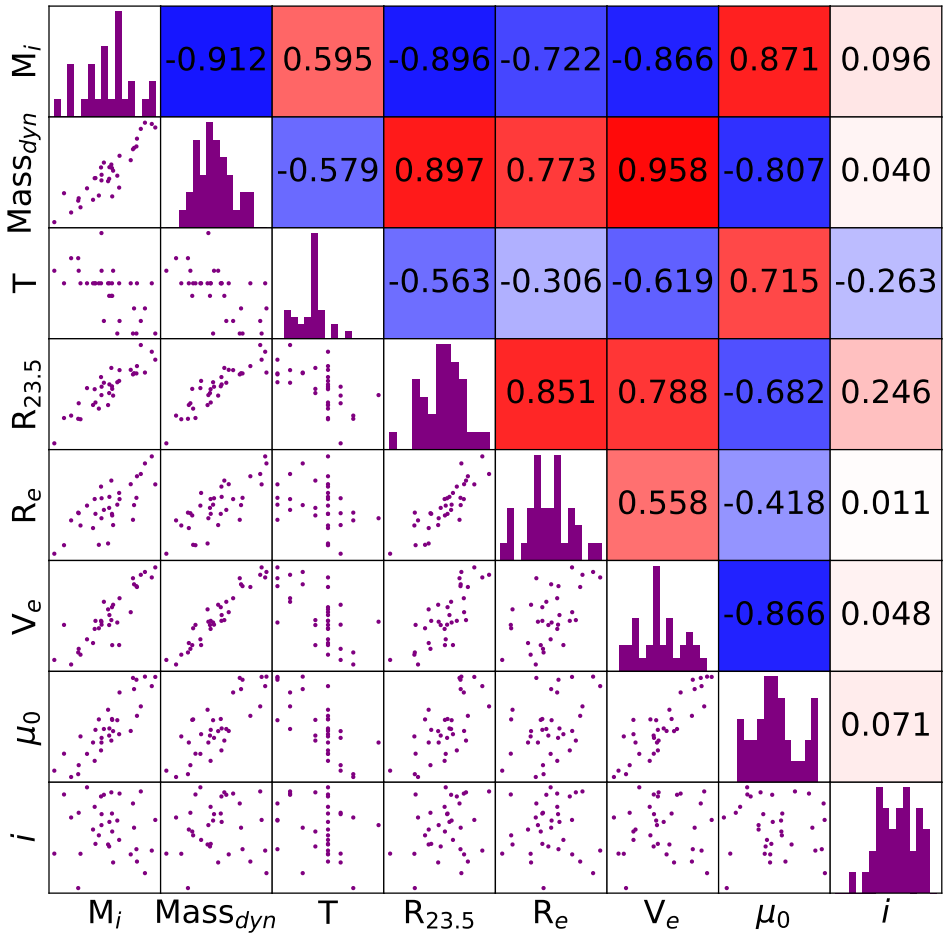

Figure B.1: Plot matrix for SV showing relationship between various measured quantities.

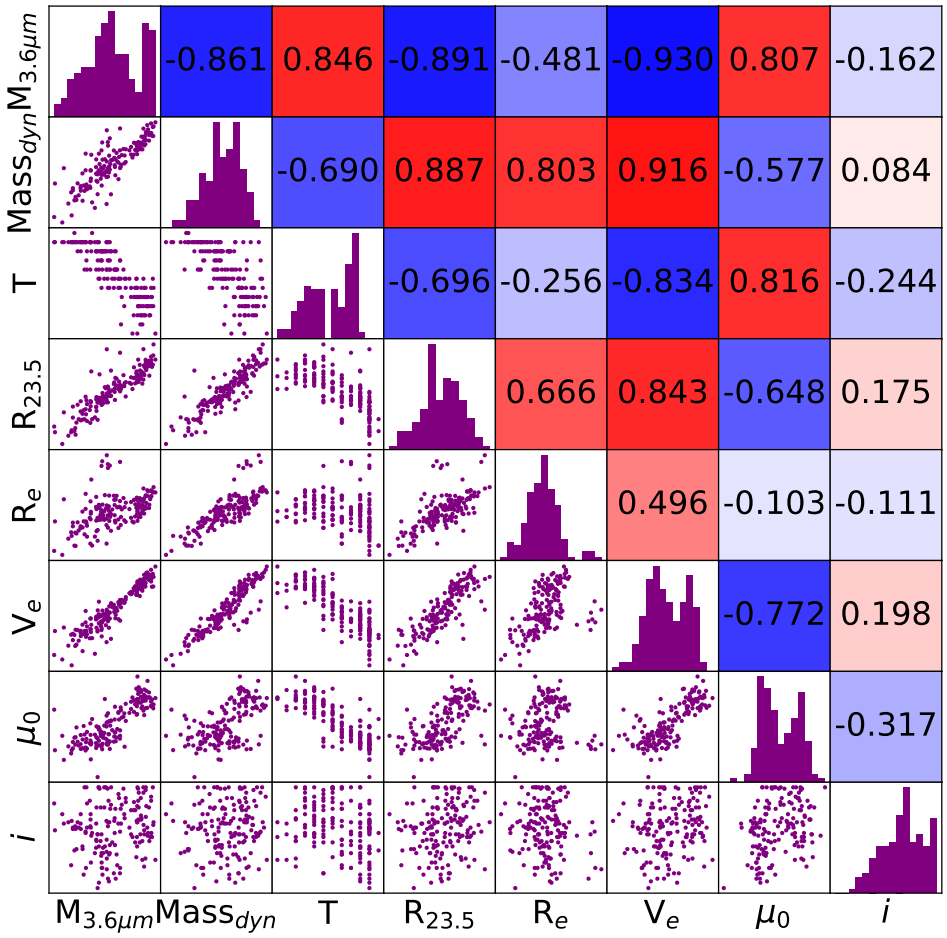

Figure B.2: Plot matrix for SP showing relationship between various measured quantities.

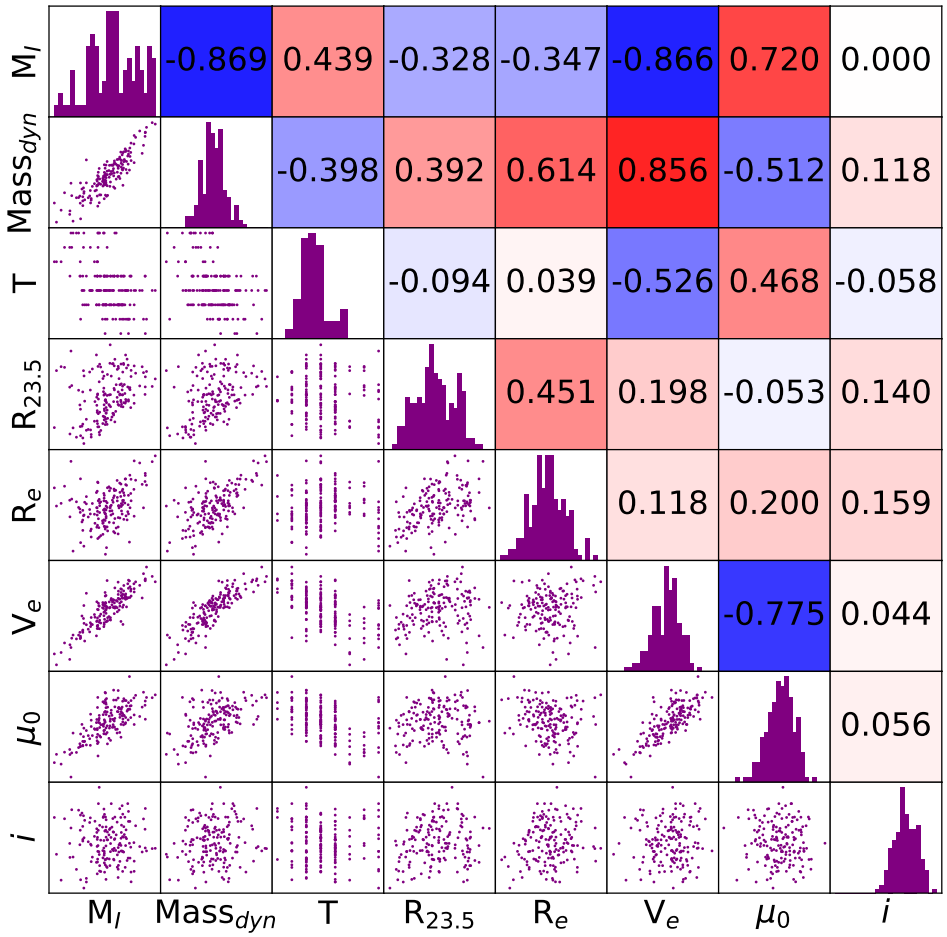

Figure B.3: Plot matrix for SF showing relationship between various measured quantities.

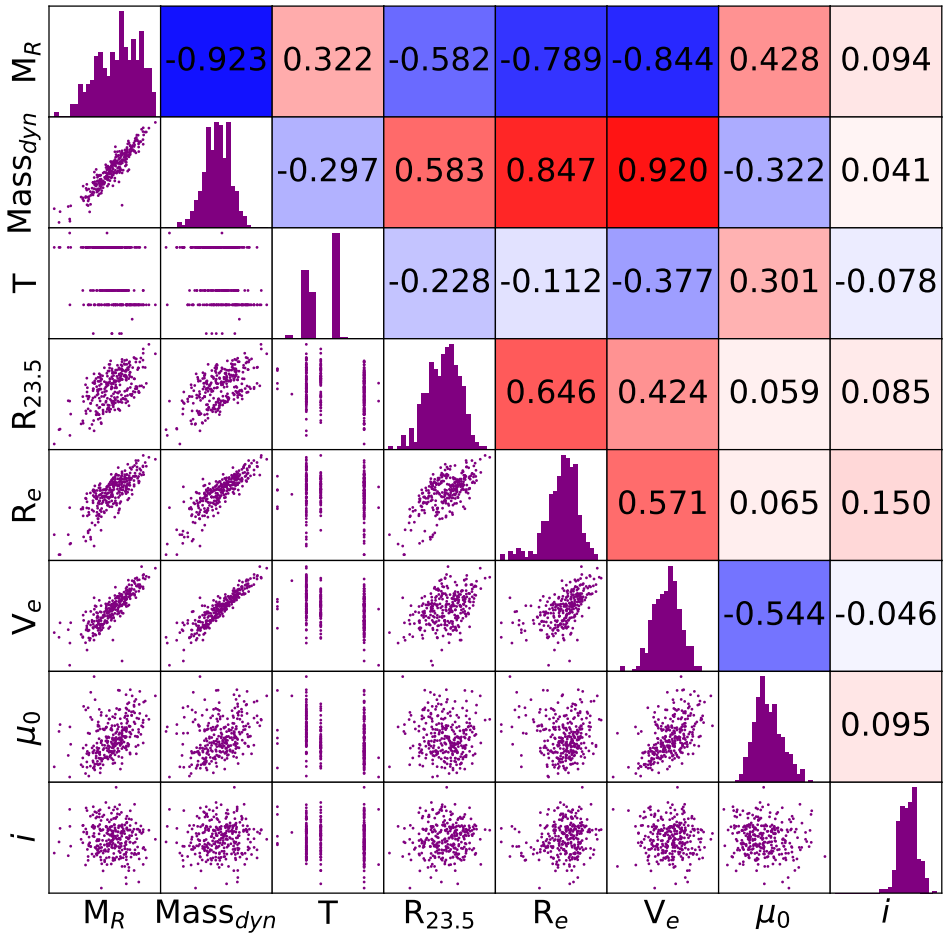

Figure B.4: Plot matrix for C97 showing relationship between various measured quantities.

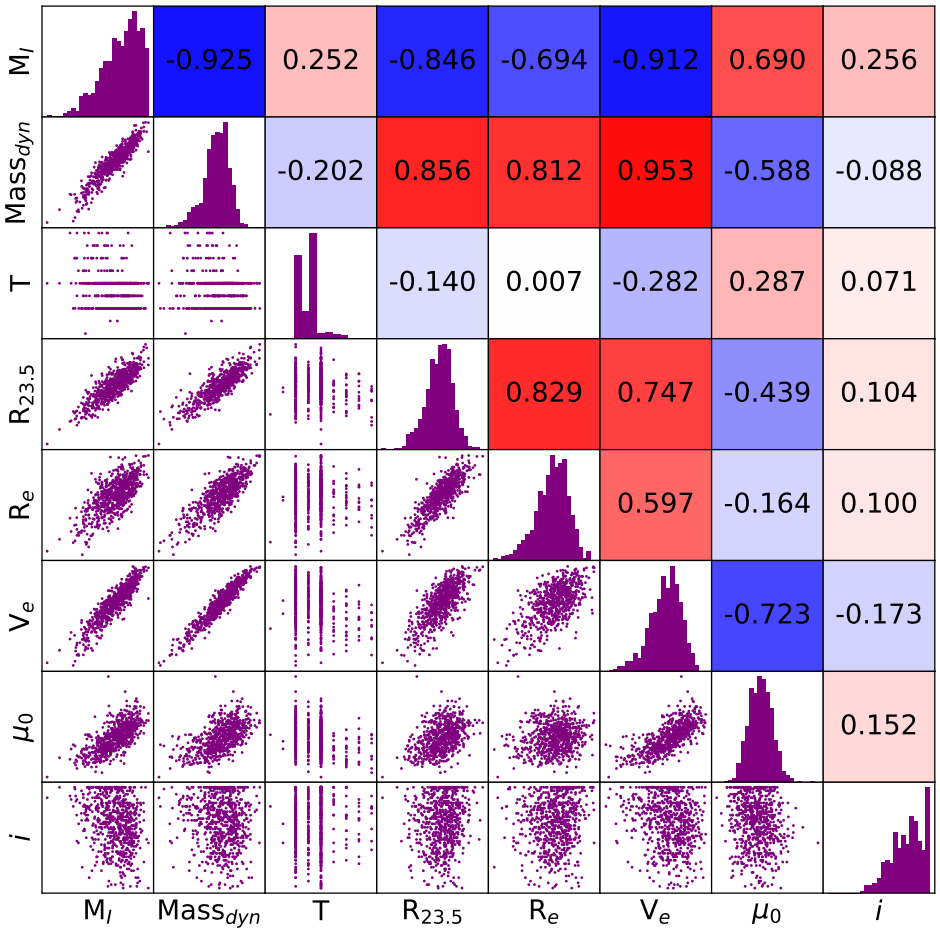

Figure B.5: Plot matrix for M92 showing relationship between various measured quantities.

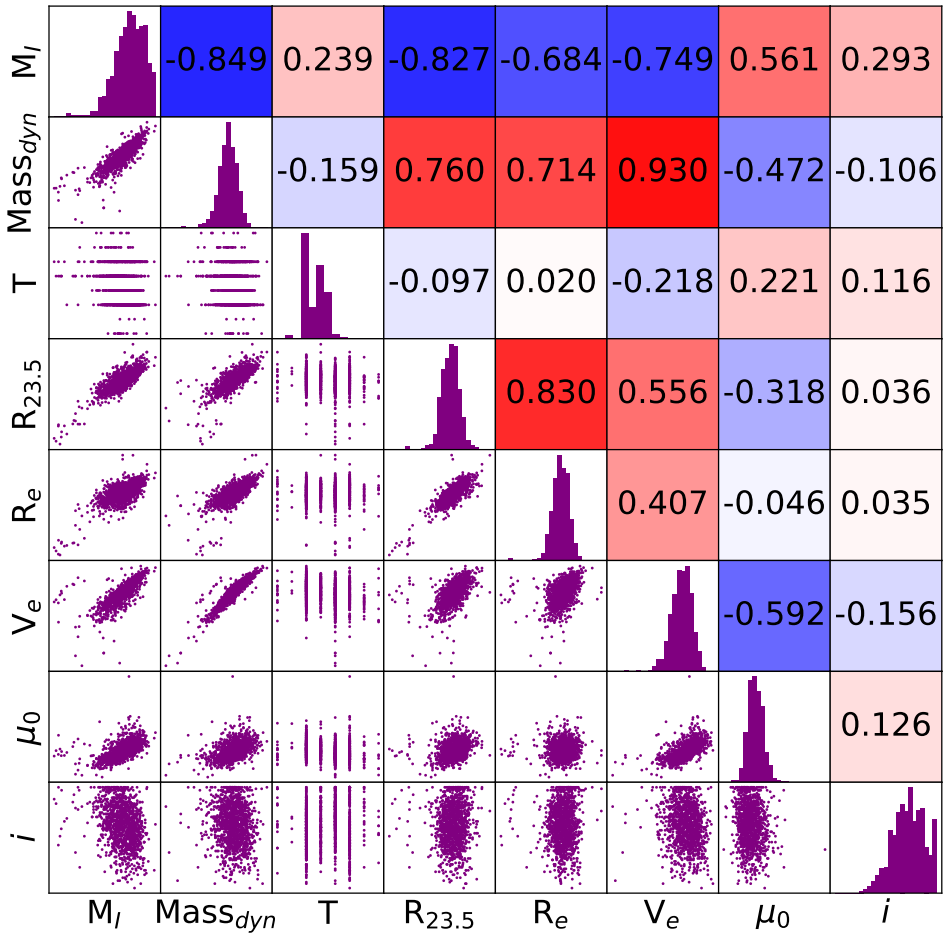

Figure B.6: Plot matrix for M96 showing relationship between various measured quantities.

Appendix C

The Residuals in the RAR Against Individual Galaxy Parameters

Presented here are the residuals of the RAR and how they relate to various galaxy parameters. Each plot has six panels with one for each dataset as labelled. The panels have 2D histograms with the x-axis as a chosen galaxy parameter and the y-axis as the residual. Overlaid on the histogram is a median line in orange and a 68.3 % interval as two blue lines. The residuals are broadly observed to be flat, with the strongest correlation being a slight dependence on velocity.

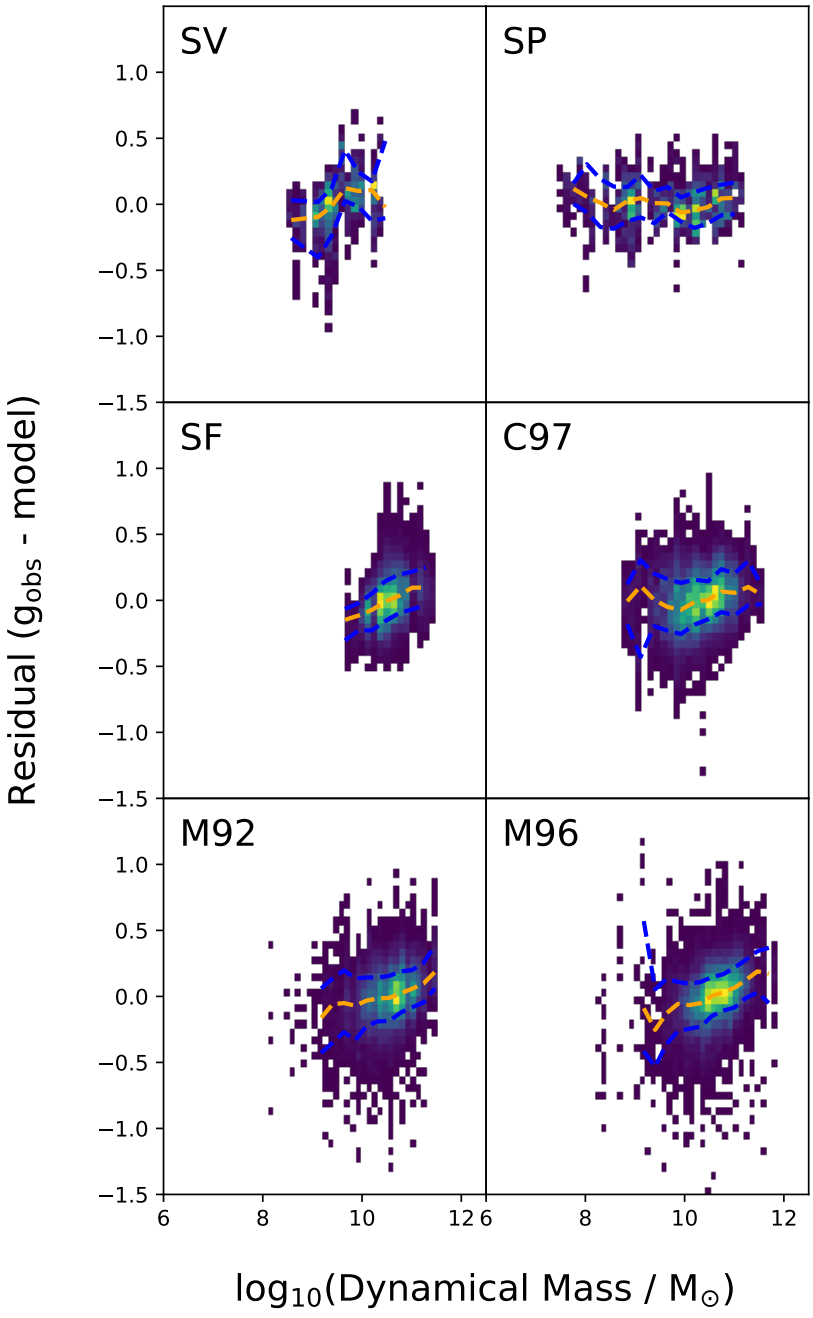

Figure C.1: Residuals from the fit to Eq. 10.2 as a function of dynamical mass evaluated at the effective radius.

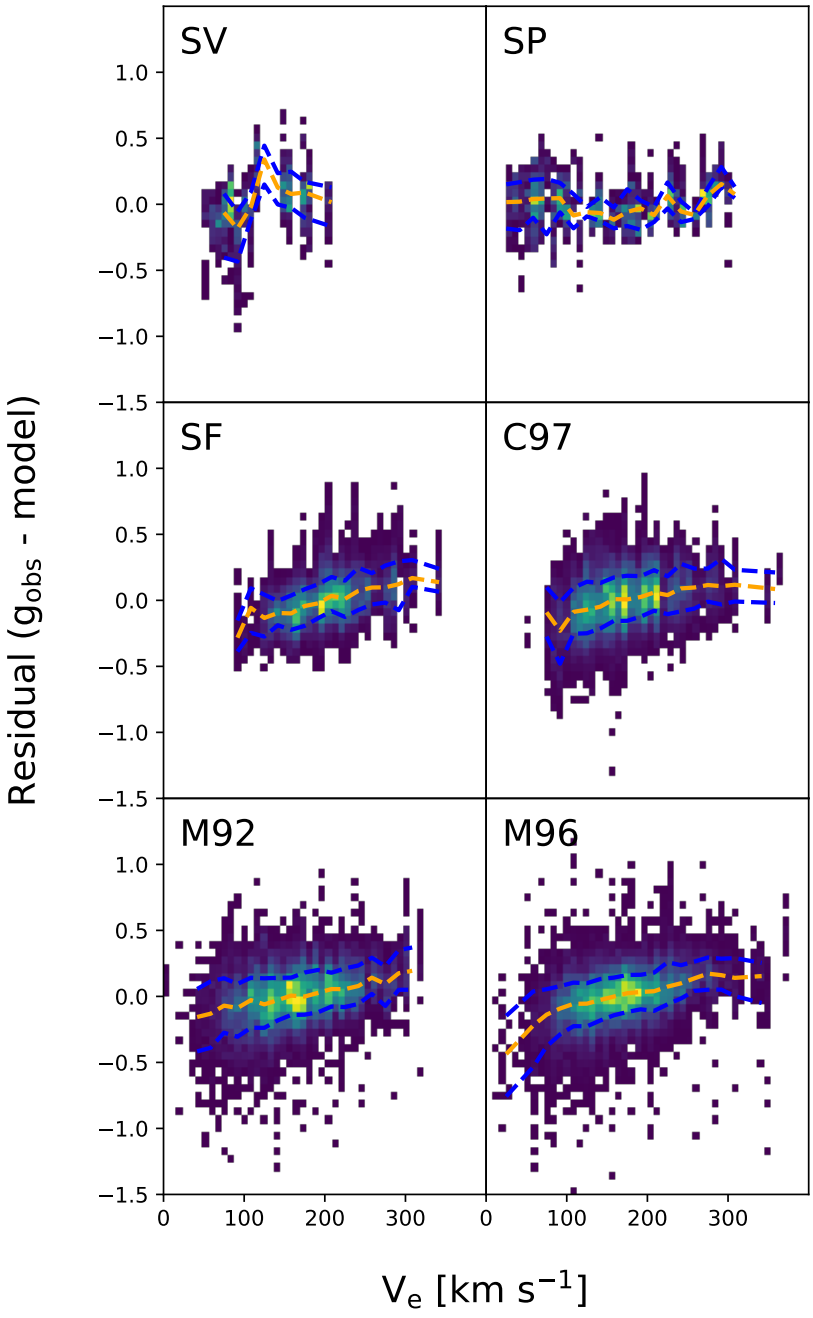

Figure C.2: Residuals from the fit to Eq. 10.2 as a function of V_{rot} evaluated at the effective radius.

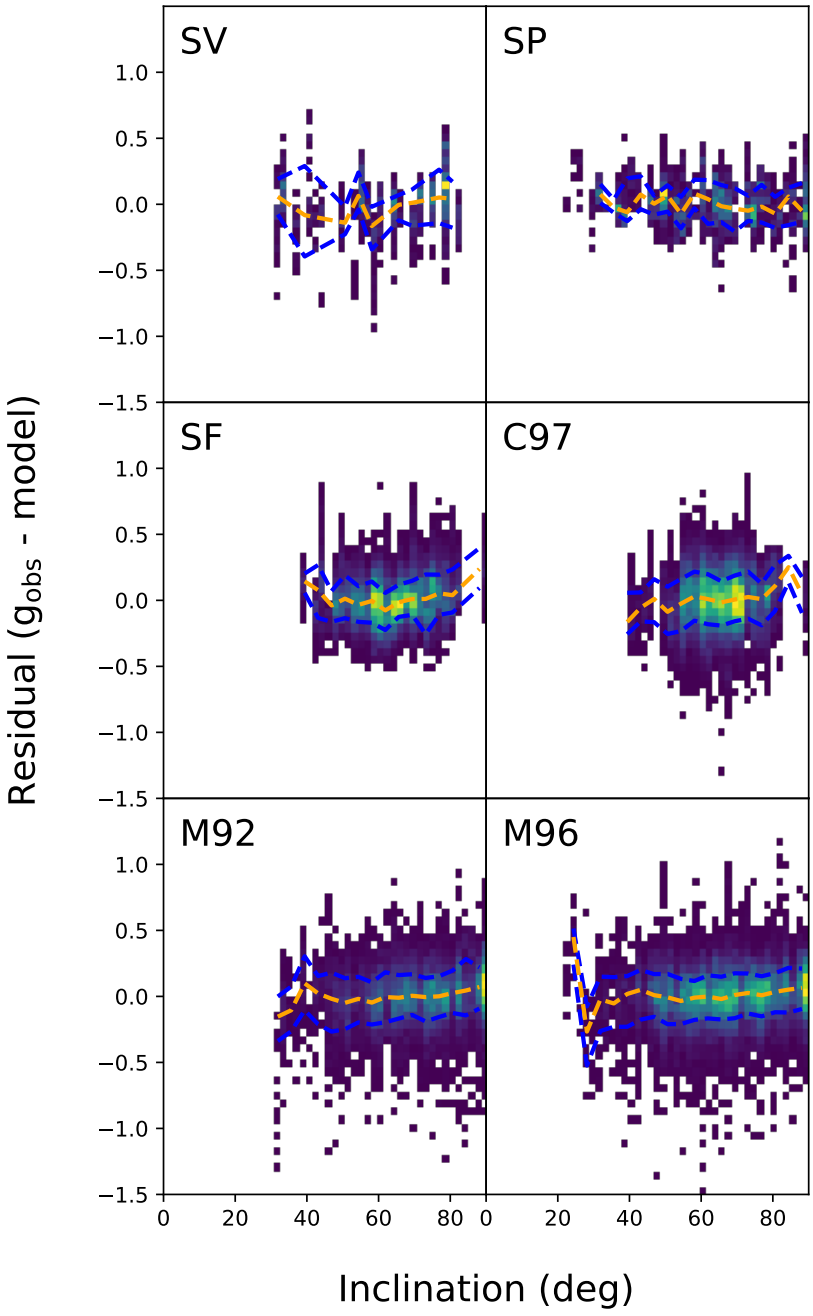

Figure C.3: Residuals from the fit to Eq. 10.2 as a function of inclination of the galaxy.

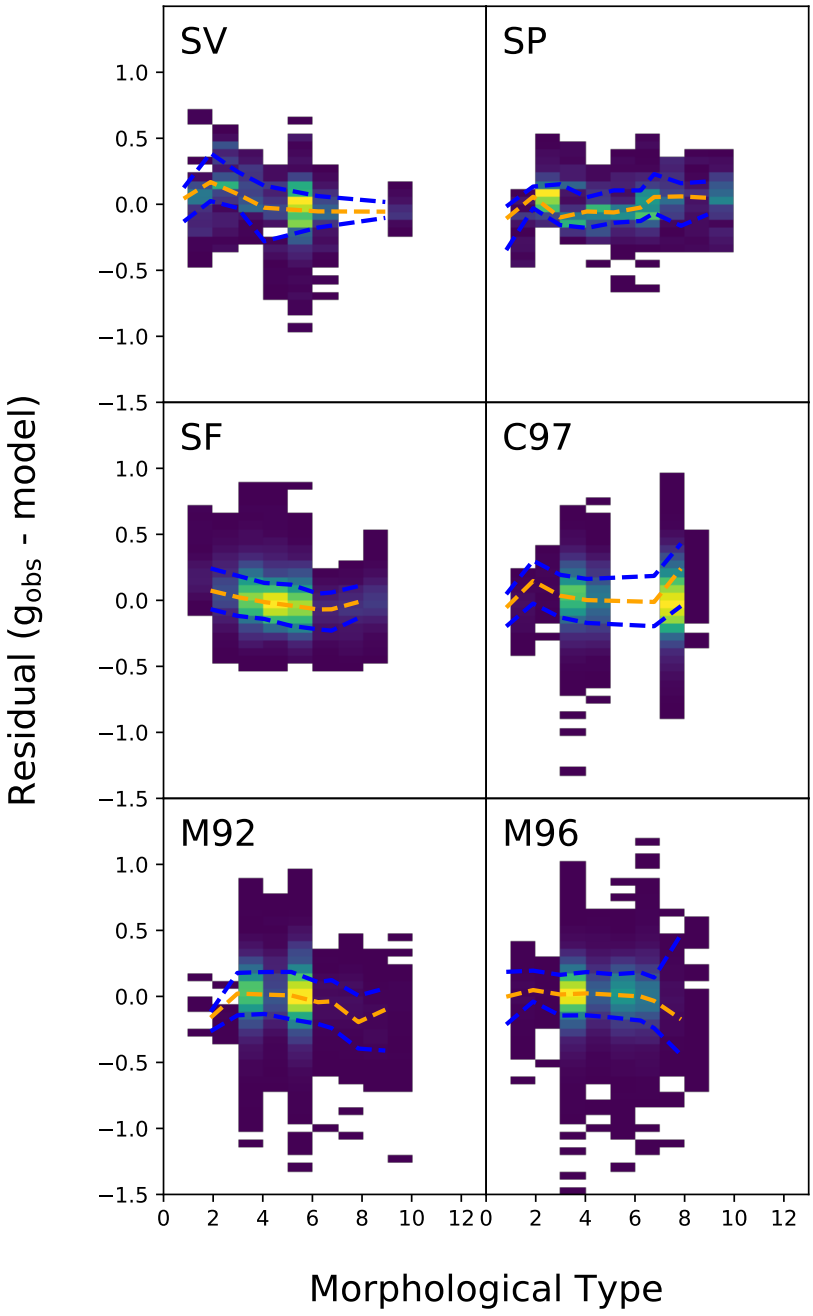

Figure C.4: Residuals from the fit to Eq. 10.2 as a function of Morphological Type.

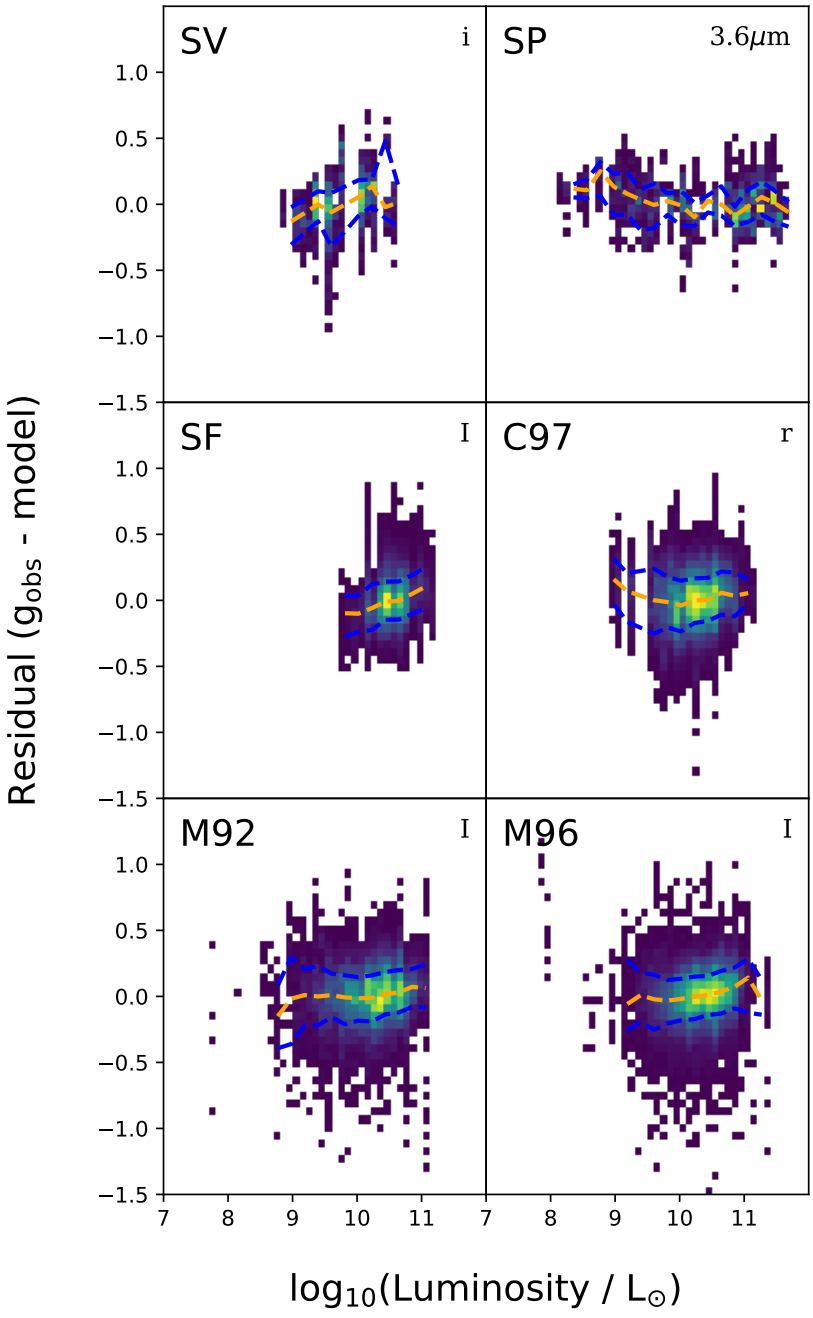

Figure C.5: Residuals from the fit to Eq. 10.2 as a function of Total Luminosity.

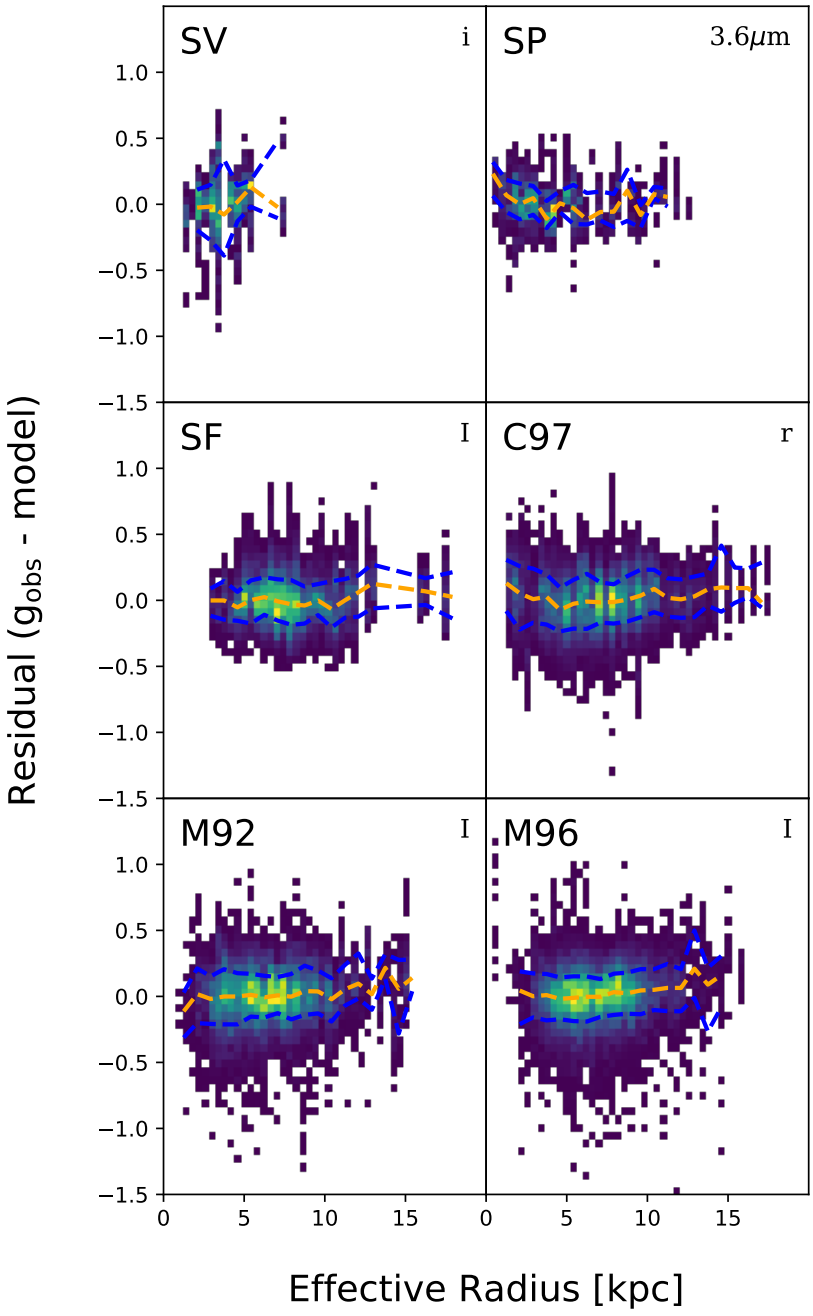

Figure C.6: Residuals from the fit to Eq. 10.2 as a function of the effective radius

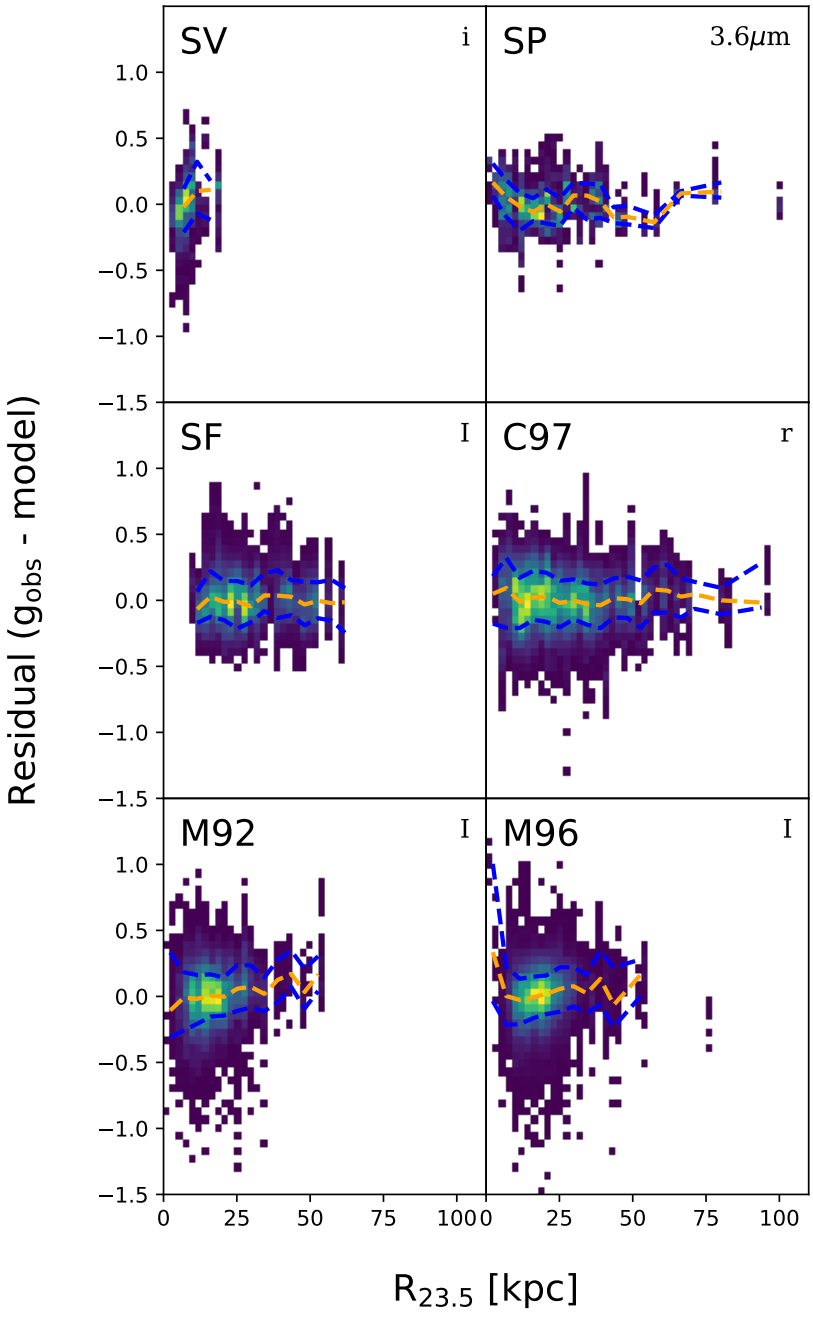

Figure C.7: Residuals from the fit to Eq. 10.2 as a function of $R_{23.5}$.

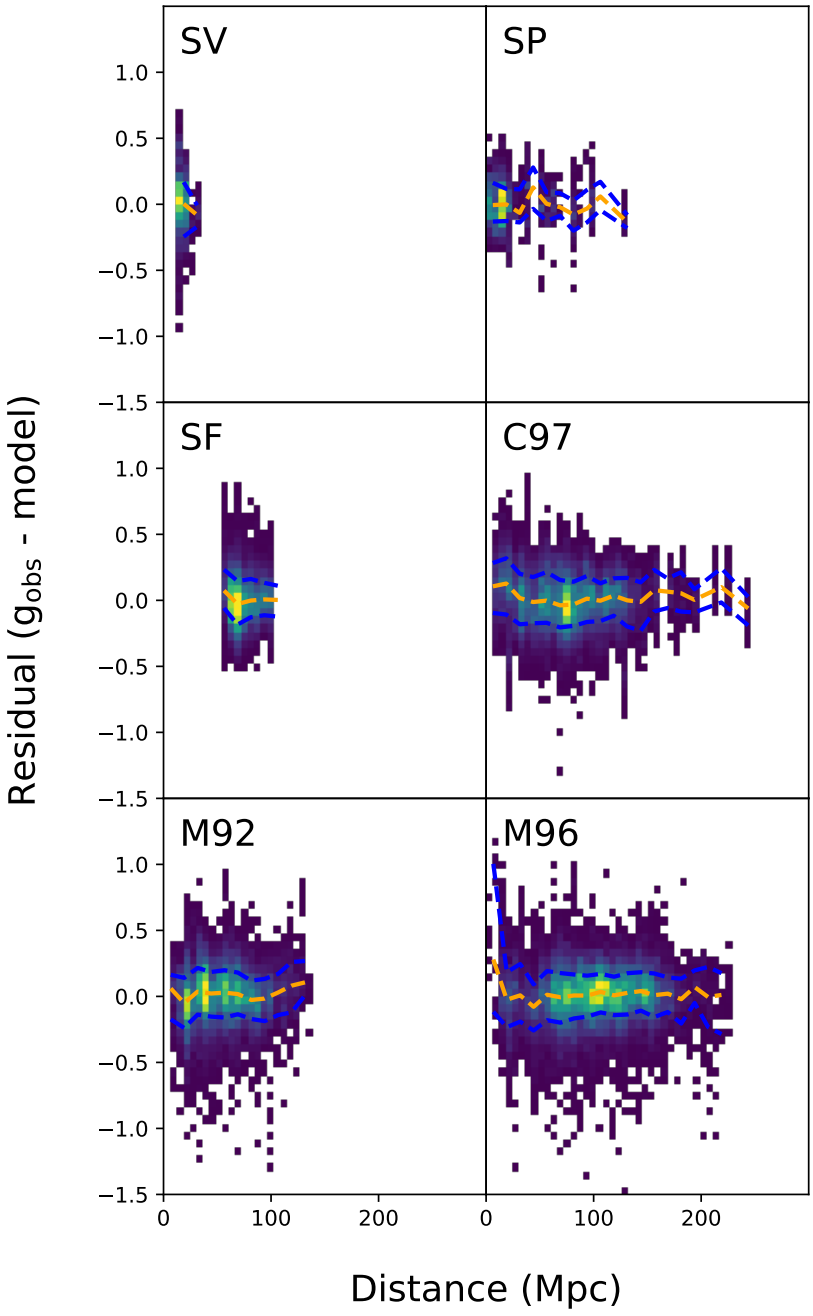

Figure C.8: Residuals from the fit to Eq. 10.2 as a function of distance.

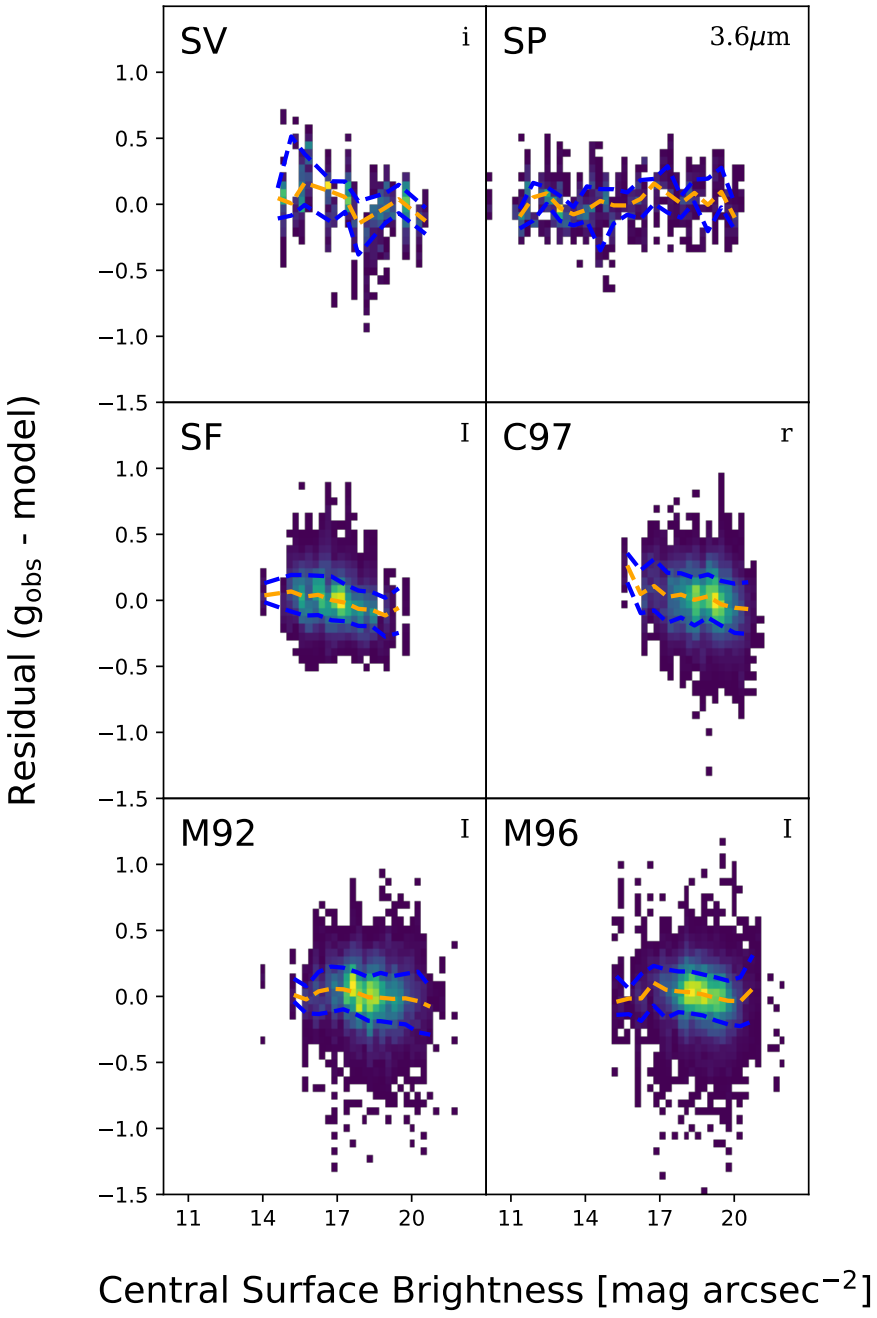

Figure C.9: Residuals from the fit to Eq. 10.2 as a function of central surface brightness.