Probing the nature of dark matter with small-scale cosmology: A new estimate of the satellite galaxy complement of the Milky Way

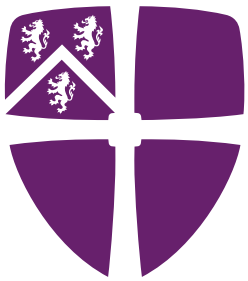
NEWTON, OLIVER,JAMES

Probing the nature of dark matter with small-scale cosmology

A new estimate of the satellite galaxy complement of the Milky Way

Oliver Newton

A Thesis presented for the degree of
Doctor of Philosophy



Durham University

Institute for Computational Cosmology
Ogden Centre for Fundamental Physics
Department of Physics
Durham University
United Kingdom

August 2019

Probing the nature of dark matter with small-scale cosmology

A new estimate of the satellite galaxy complement of the Milky Way

Oliver Newton

Submitted for the degree of Doctor of Philosophy

August 2019

Abstract: The standard model of cosmology has been enormously successful both at reproducing many observed properties of the Universe, and at predicting others. Despite this success one of its key components, a dark matter particle, has not been observed in targeted searches or indirect detection experiments. In addition, a number of discrepancies have arisen between observational proxies of DM structure and the small-scale predictions of this leading cosmological model, challenging its status as the standard paradigm. In this thesis we focus on two distinct but related lines of enquiry. In the first, we address the gap in observational capability concerning the satellite galaxies of the Milky Way. These objects are sensitive probes of the underlying distribution of dark matter, which is determined by the properties of the dark matter particle itself. Using partial observations of the Galactic satellite population, we introduce and use a Bayesian approach to infer the total luminosity function of these objects. We predict that there are 124^{+40}_{-27} (68 per cent confidence level) satellite galaxies brighter than $M_V = 0$ within 300 kpc of the Milky Way, and that half of this population should, in principle, be detectable by the forthcoming Large Synoptic Survey Telescope. In the second strand we use these estimates to test the predictions of alternative models to the standard paradigm and place robust lower limits on their allowed properties. We focus on two models of warm dark matter: thermal relics,

and sterile neutrinos in the Neutrino Minimal Standard Model (νMSM). For the former we obtain a robust lower limit on the mass of the dark matter particle, ruling out with 95 per cent confidence models with particle mass, $m_{\rm th} \leq 1.95$ keV, which is competitive with existing constraints. In the latter case the model depends on the size of the primordial lepton asymmetry, which we parametrize as L_6. Assuming a particle mass of $m_{\rm s} = 7$ keV—motivated by observations of an unexplained 3.55 keV line in X-ray spectra of galaxy clusters—we exclude values of $L_6 \geq 50$, in agreement with other work.

Declaration

The work described in this thesis was undertaken between October 2015 and August 2019 while the author was a research student under the supervision of Prof. Adrian Jenkins and Prof. Carlos Frenk in the Institute for Computational Cosmology at Durham University, England. No part of this thesis has been submitted for any degree or qualification at Durham University or elsewhere.

Chapters 2 and 3 have been published as a paper in Monthly Notices of the Royal Astronomical Society (MNRAS):

Newton O., Cautun M., Jenkins A., Frenk C. S., Helly J. C., 2018, MNRAS, **479**, 2853

All figures in this thesis were prepared by the author, or have been properly attributed in the figure caption.

Acknowledgements

This thesis is the culmination of a long journey. I do not think it an overstatement to say that without the support of a number of people over many years, it is unlikely that I would have embarked on and continued down this path. It is hard to do justice to them all in a few short sentences, and I do not have the space to include everyone I would like to. So, in the space that I *do* have, I would like to pay tribute to those who have been especially present through the ups and downs of PhD life.

I want to begin with a word of thanks to my supervisors, Adrian Jenkins and Carlos Frenk, for their steadfast support and words of wisdom as I (re-)embarked on academic life. It was a winning combination: Adrian's calm demeanour provided a ready tonic to Carlos' occasionally rambunctious disposition, and our weekly meetings were a highlight of my time in Durham. It has been a thoroughly enjoyable experience to work with both of you, and I shall miss our regular discussions when I leave the Institute. At this point, it would be remiss of me to neglect the contributions of Marius Cautun, who was not my supervisor officially but has almost certainly invested a similar amount of time! You have been a valuable source of insight and support for several years, and I am grateful for the sage advice you so readily offered when the going got tough. While I may be leaving Durham for pastures new, I look forward to continuing to work with you, Adrian, and Carlos for many more years to come.

Over my time in Durham, I have been privileged to meet many interesting people, both within and outside of astronomy. Marina, Carolin, Tobit, and Moisés, you all ensured I did not work *too* hard during my first year at Durham and I have many fond memories of

our time living in St. Mary's College. In subsequent years I benefited from the company of Antonia and Adam as housemates and friends. I enjoyed our many theological and political discussions—and the natural conclusion that the world would (obviously) be a better place if we were in charge! Within astronomy, I want to pay tribute to Piotr, Behzad, Stuart S, Jan, Jacob, Dan, Griffin, James, Chris, Anna, Steve & Will, Josh, Tom C, Calvin, Aidan, Ra'ad, Ash, Ugne, Raj, Lizelke, Sownak, Azi, Paddy, Ruari, Stu, Jaime, Bitten, Julie W, Lizzie, Emma, Dave R, Alex E, Mark, Matteo L, Mathilde, Matthieu, Birkin, Jake, and Vicky. Thank you all for your companionship and support over the years, both at work and outside it. Special thanks must also go to my office-mates past and present: Dave C, Louise, Andrew R, Stefan, Alex S, Jack, and Tom R; variously for helping me settle into PhD life and for the mischievous antics that lighten the mental load of day-to-day research. I also want to make special mention of Tom and Jack, one for being an excellent housemate during my at times stressful final year, and both of whom generously agreed to proofread some of the thesis text to enable me to rectify several grammatical defects plaguing the final drafts.

Music-making has played a central and significant role in my life from an early age, and through it, I have benefited from many unique opportunities and made many firm friends. It was through music that I was fortunate enough to meet Birdie, Butcher, Daisy, Ian, James, Liam, Neil, Pony, and Rachel—and more besides—during my undergraduate degree. I am pleased we have been able to continue our regular adventures together and look forward to more, even as we all gradually diffuse across the globe! In Durham, I enjoyed spending many Sunday afternoons and Monday evenings rehearsing with DUCB and dubb, and also in post-rehearsal socializing, heading away on tours and for contests, and performing in numerous concerts. Without a doubt, these groups enriched my postgraduate experience and kept me occupied when out of the office. In particular, Abbie, Andy, Clara, Kit, Liz, Lydia, and Thomas, your close companionship extended far beyond music-making and I will carry forward many warm memories of the times we shared, and our weekly Bake Off nights. Outside of University, I would like to thank Fishburn and Reg Vardy brass bands—the former for getting me back into form, and the latter for taking me under their

wing and developing my abilities, affording me many opportunities along the way. I also extend my thanks to the members of both bands for their friendship over these last few years. Special thanks must go to Duncan, Ed, and Jonny, for their camaraderie and good humour in the kitchen sink department.

I would like to end this now-lengthy section by thanking the most important people of all: my parents. Thank you for raising me to be the person I am today, and supporting my various interests—sometimes, I now know, to the detriment of your own. Your unwavering support and encouragement mean more than I have perhaps ever been able to express. Thank you for being there with me, every step of the way.

“But wait. . . something’s fishy. . . ”

— M. Cautun

Dedicated to

Mum and Dad

and

my grandparents,
three of whom saw this journey begin;
two of whom will see it conclude

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 A model of cosmology

The modern field of cosmology can trace its origins to one of the most famous scientific theories of the last century: General Relativity (Einstein, 1915, 1917, translated by W. Perrett and G. B. Jeffery). The Newtonian law of universal gravitation, which had for over 200 years been the accepted model to describe the interactions between massive objects, was subsumed into this new theory which described gravitational effects as a geometric property of space and time. The first astrophysical predictions of General Relativity proved its merits: the advance of the perihelion of the planet Mercury could be explained without recourse to additional, arbitrary parameters (Einstein, 1915); and its predictions of the deflection of light by massive objects were confirmed spectacularly by observations of the total solar eclipse of 29 May 1919 (Dyson et al., 1920). In 1917, Einstein took a broader astrophysical view and applied General Relativity to study all matter in the Universe as a whole, under the assumption that spatially it is both homogeneous and isotropic. Under the action of gravitational attraction, such a model universe would if released from dynamical equilibrium, collapse in on itself. Dissatisfied that the theory was unable to produce a temporally infinite universe—this was the prevailing cosmological view of the

time—Einstein revised the field equations of General Relativity to include a constant term, Λ (Einstein, 1917). This extra term effectively imbues the vacuum of space with an intrinsic energy density and negative pressure, counterbalancing the attractive force of gravity and permitting Einstein's desired spatially finite, 'static' universe. However, this was later proved to be an unstable configuration which could be nudged easily into a state of permanent expansion or contraction by minute fluctuations in either the value of Λ, the matter density, or in the geometric curvature of the universe.

The first exact solutions to the field equations of General Relativity in a homogeneous and isotropic universe were derived by Friedman (1922), and several years later independently by Lemaître (1927), and provided for an expanding (or contracting) universe. These solutions depend upon two key ingredients: the total mass–energy density of the universe and the curvature of spacetime. The latter underpins the fundamental tenet of General Relativity that gravity is the geometric manifestation of the curvature of spacetime due to a massive object that alters the paths of other objects travelling within that region. In the same way, the total matter content of the universe also affects its geometry which can be categorized into three regimes:

Flat If the mass–energy density is equal to the 'critical' density, spacetime is said to be 'flat' and the universe has zero curvature—analogous to a flat sheet of paper with infinite spatial extent.

Closed A total mass–energy density *greater* than the critical density produces a 'closed' universe in which the gravitational attraction of its matter content overcomes the expansion of space and eventually causes the universe to collapse back on itself. A universe with such 'positive' curvature can be imagined as similar to the surface of a sphere.

Open A universe with a density less than the critical density describes an 'open' universe with negative curvature, analogous to the surface of a saddle which continues to expand forever.

All of these scenarios can be realized *without* the inclusion of a non-zero cosmological constant term in the field equations of General Relativity, prompting subsequent cosmological models to assume Λ=0. This assumption would come to be reconsidered towards the end of the 20$^{\text{th}}$ century in light of new observational measurements; we will return to this in Section 1.2.

The concept of an expanding universe was met with resistance from the astronomical community until 1929, when Edwin Hubble published observational measurements of the recession of 'faint nebulae', which only recently had been confirmed to be extragalactic objects (Hubble, 1924, 1925). These demonstrated that the 'nebulae', now known to be nearby galaxies, were all receding with velocities, v, that were proportional to their distance from the Milky Way (MW), r, such that

$$v = H_0\, r, \tag{1.1.1}$$

where the constant of proportionality, H_0—the 'Hubble constant'—describes the present rate of expansion of the Universe (Hubble, 1929). The most recent observational measurements of this quantity place its value at $H_0 \approx 70$ km s^{-1} Mpc^{-1} (Ade et al., 2016; Riess et al., 2016; Planck Collaboration et al., 2018). The relationship that is given in equation (1.1.1) had been demonstrated previously by Lemaître in his 1927 paper using measurements of the 'faint nebulae', although this received scant attention at the time (Lemaître, 1927). The observational verification of an expanding Universe prompted Lemaître to extrapolate the philosophical and mathematical connotations of such a model backwards in time. From this, Lemaître concluded that at some finite time in the past all mass was concentrated into a single point, or 'primaeval atom', from which all of space and time came into existence (Lemaître, 1931). Thus, the Universe had a beginning, and this cosmological view would later come to be known as the 'Big Bang' model.∗

Many of Lemaître's contemporaries found the idea that the Universe had a beginning to be distasteful, and alternative models were soon advocated (e.g. Zwicky, 1929; Einstein, 1930; Milne, 1935). By the 1950s only two cosmological models seemed promising: the

∗This term was actually coined by Fred Hoyle during a BBC Radio interview broadcast in 1949

steady-state model of Bondi & Gold (1948) and Hoyle (1948), and the newly christened Big Bang theory of Lemaître. The former acknowledged the observed expansion of the Universe but required that its appearance does not change over time, obviating the need for a 'beginning' or 'end' by continually creating matter in order to maintain the matter density of the universe. The latter model provided for a universe in which time and space have a definite beginning and, having been developed further in the intervening years, could now produce predictions for the formation of the nuclei of light chemical elements in a $\sim 10^9$ K plasma shortly after the Big Bang (Alpher et al., 1948). Extending these predictions further, approximately 380 000 years after this the expansion of the Universe cools this plasma sufficiently to enable the light nuclei to combine with free electrons to form stable atoms. The disappearance of free electrons decouples radiation from matter and the Universe becomes transparent to this black-body radiation, enabling photons to traverse the Universe unimpeded. It was anticipated that the subsequent expansion of space would redshift this black-body emission to the microwave regime by the present day (Alpher & Herman, 1948a,b; Gamow, 1948a,b).

Fifteen years later, this emission was discovered by the serendipitous detection of isotropic microwave radiation by Penzias & Wilson (1965) in a spectacular confirmation of the predictions of the Big Bang model. Today, this model forms a fundamental and accepted part of our understanding of the Universe. Measurements of this cosmic microwave background (CMB) are now used to constrain many cosmological parameters such as the curvature of the Universe and the baryon density with exquisite precision (e.g. Smoot et al., 1992; Hinshaw et al., 2013; Planck Collaboration et al., 2018).

1.2 The development of a standard cosmological paradigm

The current standard cosmological model, known as 'Λ Cold Dark Matter' (ΛCDM), is the culmination of almost 85 years of theoretical and observational endeavour to understand

the creation and evolution of the Universe. At its foundation is the conceptually simple Big Bang model, discussed in the preceding section, that describes a Universe that began in an extremely hot, dense state before rapidly expanding and cooling to form the rich structure that we observe today (Lemaître, 1931). This expansion, it is theorized, was driven initially by a period of exponential cosmic inflation that amplified quantum fluctuations in the primordial plasma to macroscopic scales (Guth, 1981; Linde, 1982a). These seeded the early Universe with density fluctuations that, in the later gravitationally-dominated epoch, would be responsible for the formation of cosmic structure (Guth & Pi, 1982; Hawking, 1982; Linde, 1982b; Starobinsky, 1982).

1.2.1 Dark matter

Contemporaneous with the proposition of the Big Bang model, an observational study of the Coma cluster of galaxies was carried out by Fritz Zwicky. This showed that there was hundreds of times more mass in the cluster than was observable from the stellar component alone (Zwicky, 1933, 1937) and was the first unambiguous evidence for what he dubbed '*dunkle Materie*', or 'dark matter' (DM)—taken to mean any astrophysical substance that is too faint to be detected. It also supplied the answer to other puzzling observations dating from as early as 1884 that had also identified the need for additional, 'non-luminous' matter to make sense of their measurements (Kelvin, 1904; Poincaré, 1906; Poincaré & Vergne, 1911; Kapteyn, 1922; Oort, 1932). The existence of this mysterious substance was further reinforced by observations of the rotation curve of M31 that indicated that most of its mass was at large radii, beyond the stellar component (Babcock, 1939); and by the dynamical arguments of Kahn & Woltjer (1959) that required the masses of the MW and M31 to be much larger than are inferred from their luminous components in order to explain their present approach towards one another.

Despite a growing base of observational evidence in favour of DM (e.g. Roberts, 1966; Freeman, 1970; Rubin & Ford, 1970; Rogstad & Shostak, 1972; Roberts & Rots, 1973), its importance for galaxy formation and cosmological models was not established until

the mid-1970s, when Ostriker & Peebles (1973) demonstrated with numerical simulations that galaxies require a massive halo in order to stabilize their discs. Support for this followed in a series of observational papers that endorsed the existence of massive haloes of non-luminous—or exceptionally faint—matter around galaxies (Einasto et al., 1974a,b; Ostriker et al., 1974; Rubin et al., 1980, see also Faber & Gallagher 1979). Quantifying the contribution of this unseen material to the matter density of the Universe was crucially important in order to understand its effect on cosmic evolution at late times; however, the identity of the DM remained entirely speculative (Gershtein & Zel'dovich, 1966a,b; Cowsik & McClelland, 1972, 1973; Szalay & Marx, 1976). Four years later, attempts to model its macroscopic effects on the evolution of cosmic structure resulted in a model of hierarchical structure formation that incorporated DM as a key component, driving the evolution of the Universe (White & Rees, 1978).

The notion that the DM might be particle-like in nature did not receive much consideration until Lubimov et al. (1980) constrained the mass of the electron anti-neutrino to be $\sim$30 eV by studying the $\beta-$decay spectrum of tritium. Shortly after the Big Bang when the temperature is greater than $\sim 10^{10}$ K, neutrinos and other fundamental particles are created in thermal equilibrium in a primordial plasma. The number density of particular particle species can be calculated from their momentum distributions in a black-body spectrum, which in turn can be used to determine the critical density for the closure of the Universe. This work had been carried out previously by Gershtein & Zel'dovich (1966a) and Cowsik & McClelland (1972), and had demonstrated that neutrinos with masses in the $\mathcal{O}(10$ eV) regime are sufficient to close the Universe and fulfil the role of the DM (Szalay & Marx, 1976). Although the mass of the electron anti-neutrino claimed by Lubimov et al. (1980) would later be refuted, their work helped to popularize the concept that DM was composed not of macroscopic astrophysical objects but rather, microscopic fundamental particles.

From a cosmological perspective, the precise identity of particle-like DM is fairly unimportant, save for the strength of self-interaction the DM might experience. All that is required is that the choice of particle species does not introduce a dependence on additional long-range forces besides gravity on large astrophysical scales, and that it exists in sufficient

quantity to influence astrophysical objects. Of more consequence for structure formation is the initial velocity distribution of the DM (Bond et al., 1980; Doroshkevich et al., 1980; Bond & Szalay, 1983). DM particles with a mass in the range proposed by Lubimov et al. (1980) would decouple from thermal equilibrium early in the evolutionary history of the Universe. The temperature at this time far exceeds the rest mass of the DM, imparting the particles with relativistic velocities which are retained throughout much of the formation of cosmic structure. Such 'hot' DM (HDM) particles would have sufficient velocity to free-stream out of small-scale density fluctuations, suppressing the formation of structures below a characteristic free-streaming scale that is inversely proportional to the DM particle mass (Schramm & Steigman, 1981; Peebles, 1982a). This scale was characterized with one of the first HDM simulations to be carried out within an inflationary cosmology which predicted significant clustering of galaxies on supercluster scales but very little structure below this (White et al., 1983). This was in significant disagreement with the first 3D survey of galaxies in the local Universe that had been carried out by the Harvard Center for Astrophysics (CfA) a year previously. This showed less clustering on large scales and more structure on smaller ones, ruling out such HDM models (Davis et al., 1982).

While HDM had proved to be an unsuccessful model of the Universe, it helped to establish a general template with which to investigate the desired behaviour of DM on astrophysical scales to find agreement with observations. In contrast with HDM, the class of models that came to be known as 'cold' DM (CDM) preferred DM particles that were non-relativistic at early times, enabling the formation of much smaller structures (Blumenthal et al., 1984). The first numerical simulation of CDM in an inflationary cosmology was carried out by Davis et al. (1985) and obtained excellent agreement with the CfA survey observations. While other DM models were also considered (Bond et al., 1982; Olive & Turner, 1982; Pagels & Primack, 1982; Peebles, 1982b), the superb agreement of the CDM model with observations firmly established CDM as a fundamental component of the Big Bang cosmogony that appears to govern our Universe.

1.2.2 Dark energy

The dominance of the CDM model extended to the end of the decade until improved observational measurements determined that galaxy correlations on large scales are stronger than predicted by CDM (Efstathiou et al., 1990), and that the baryon fractions predicted by CDM+Big Bang nucleosynthesis models are too low to be consistent with observations of galaxy clusters (White et al., 1993). The discovery of anisotropies in the CMB by the COsmic Background Explorer (COBE) satellite challenged further the applicability of purely CDM models, compelling the introduction of large 'biasing' factors that required visible galaxies to preferentially trace the most massive DM structures (Efstathiou et al., 1992; Smoot et al., 1992; Wright et al., 1992). This helped to instigate the serious exploration of several alternative models, including mixed DM models that incorporated two types of DM component: CDM and HDM (e.g. Davis et al., 1992), and models with a non-zero cosmological constant term.

As we discussed in Section 1.1, a non-zero, positive cosmological constant imparts a positive energy density to the vacuum of space, counteracting the gravitational attraction of ordinary matter. This 'dark energy' was originally introduced by Einstein into his theory of General Relativity to achieve a 'static universe': one in which space is finite and neither expands nor contracts, while time is infinite in extent (Einstein, 1917). After the discovery of the expansion of the Universe (Hubble, 1929) a non-zero constant term was no longer considered necessary in the cosmological models that followed, until the difficulties encountered in reconciling these with the observed anisotropies of the CMB. Finally, at the end of the millennium, new observational results provided compelling evidence in support of a non-zero cosmological constant: measurements of Type Ia supernovae in distant galaxies confirmed that the Universe was expanding at an *accelerating* rate (Riess et al., 1998; Schmidt et al., 1998; Perlmutter et al., 1999). This, together with improved measurements of the CMB demonstrating that the Universe has a flat geometry (de Bernardis et al., 2000; Hanany et al., 2000), encouraged the adoption of $\Lambda>0$ as a key component of the standard cosmological description of the Universe.

The current cosmological paradigm, ΛCDM, performs remarkably well at reproducing a large number of observable properties of the Universe and forms the basis of our present understanding of cosmic evolutionary history from the first moments of existence to the present. This rests on three key components: first, a model for the expansion of time and space that is grounded in the framework of General Relativity and incorporates a period of exponential inflation that amplifies quantum fluctuations to macroscopic scales, seeding the formation of cosmic structure. Secondly, the inclusion of a significant matter component dominated by CDM; a massive, non-baryonic, collisionless and weakly interacting form of matter that is the primary architect of the observed large-scale structure of the Universe through its gravitational interactions with ordinary matter. Third, a non-zero, positive cosmological constant, Λ, in the field equations of General Relativity which permeates all of space and constitutes over two-thirds of the total mass–energy content of the Universe at the present day, and is responsible for the accelerated expansion of the Universe. This model is the simplest and most widely-accepted description of our present understanding of the Universe. However, several discrepancies have arisen in the highly non-linear regime in which small haloes and dwarf galaxies form that could provide an important test of this cosmological paradigm.

1.3 Small scale challenges to ΛCDM

ΛCDM is born of a class of cosmological models that assume the 'cosmological principle' of homogeneity and isotropy throughout the Universe. On large scales, this predicts structure that accords with observations of the Universe. On 'small scales', defined as the non-linear regime where the variance of density fluctuations becomes large, predictions are more difficult to make, necessitating an increasing reliance on N–body simulations to probe these scales. This has attracted considerable attention in recent years as increasing computational power has delivered access to ever more exquisite and detailed simulations to study the predictions of cosmological models across an enormous range of scales. Alongside this, developments in instrument design and observational technique have

improved the observational data used to test the results of these advanced numerical simulations.

Over the last twenty years, the MW and its immediate environs have provided an ideal laboratory to probe the outcome of structure formation on the smallest scales. It is difficult to make theoretical predictions in this non-linear regime, and observations have uncovered several puzzling deviations from theoretical expectations, spurring considerable efforts to reconcile theoretical predictions with observational realities. Deficiencies in modelling the physics of galaxy formation could offer an appealingly simple explanation for some 'small-scale challenges' to the standard cosmological paradigm; however, it is not yet clear that these offer the entire explanation. What is clear is that 'small-scale cosmology' provides a powerful probe of fundamental theories that can help us to develop our theoretical understanding and possibly provide hints at more fundamental physics which will lead to deeper insights on the nature of the cosmos.

1.3.1 'Missing' satellite galaxies

Early simulations of the small scale structure inside individual MW-like ΛCDM haloes demonstrated that the DM accumulates into clumps spanning a wide range of masses, at least down to the resolution limit of numerical simulations (e.g. Dubinski & Carlberg, 1991; Warren et al., 1992, but see also Ghigna et al., 1998; Moore et al., 1999). Within a hierarchically assembled Universe, small clumps are the first to form from the primordial density fluctuations before merging into larger haloes, occasionally surviving this process to exist at the present day as small *sub*haloes. In a typical MW-mass halo, there are expected to be thousands of such objects, many of which should be capable of hosting a visible galaxy.

One of the first investigations to compare in detail the abundance of ΛCDM substructure in MW haloes with the number of observed satellites was carried out by Kauffmann et al. (1993). Using semi-analytic models of galaxy formation they found that typical MW-like haloes are capable of hosting ∼100 dwarf galaxies at least as bright as the classical

satellites, which was a factor of ∼10 times more than had been observed. A few years later a similar study on this question was carried out by Klypin et al. (1999), this time with higher-resolution dissipationless DM-only simulations. They compared the number of subhaloes that formed in their simulations as a function of their maximum circular velocities with the number of dwarf galaxies observed around both the MW and M31. The maximum circular velocity of a subhalo measures the depth of the potential well, providing a reasonable indication of the mass contained within the bound substructure and importantly, can be compared directly with observational measurements of the same quantity. Klypin et al. found when carrying out this comparison that their higher-resolution simulations also predicted an overabundance of DM subhaloes at fixed circular velocity, similar to that found by Kauffmann et al. (1993). This discrepancy seemingly worsened further when Moore et al. (1999) found in their numerical simulations that the MW should contain ∼500 DM subhaloes capable of hosting dwarf galaxies at least as faint as the faintest known classical satellites.

That such a discrepancy might exist is not too surprising, although the scale of the discovered disparity had not been anticipated. Most early simulations consisted entirely of DM simulation particles and could not be expected to account for the various physical processes that govern the growth and formation of the luminous component of galaxy haloes. One such process is the expulsion of gas from nascent galaxies by the intense radiation of young stars and the injection of kinetic energy by supernovae (Larson, 1974). Another is the condensation of the gas needed for star formation into DM haloes, controlled by the reionization of the surrounding intergalactic neutral hydrogen (White & Rees, 1978, see also Efstathiou 1992). While these baryonic feedback mechanisms have been known about for many years, even today their effects on galaxy formation are still not understood fully—although the parameter space is constrained much better.

In the late 1990s, the first fully hydrodynamic simulations that self-consistently traced the evolution of gas and stars at sufficient resolution to model reliably the formation of dwarf satellite galaxies were still several years away. Moreover, strong observational constraints on processes such as the reionization of hydrogen at early times in the history of the

Universe were also lacking. The best way of quickly exploring the vast and relatively unconstrained parameter space of baryonic feedback mechanisms was therefore to use semi-analytic models. This work provided a good qualitative understanding of the importance and relative contribution of different baryonic feedback processes on the formation and evolution of dwarf galaxies, and also predicted the existence of a large population of fainter satellites around the MW waiting to be discovered (Bullock et al., 2000; Benson et al., 2002a,b; Somerville, 2002). However, additional observations and better simulations were still needed to verify many of the assumptions.

The first high-resolution hydrodynamic cosmological simulations of the MW *and* the Local Group to address this gap in theoretical capability helped to confirm the early semi-analytic results in more detail, providing firmer ground for additional theoretical developments on various aspects of baryonic feedback (Okamoto et al., 2005; Macciò et al., 2007; Okamoto et al., 2008). In particular, some of this work suggested that gas accretion and galaxy formation is strongly suppressed by the UV background in haloes with masses $\lesssim 10^9$ $M_\odot$, potentially offering a natural solution to the 'missing satellites problem'. Improvements in the observational constraints on galaxy formation processes have provided a complimentary benchmark against which to compare progressively more detailed numerical simulations (e.g. Fan et al., 2006; Bolton et al., 2011; Caruana et al., 2012, 2014; Becker et al., 2015; Greig et al., 2017; Planck Collaboration et al., 2018). This has culminated with the current generation of state-of-the-art hydrodynamic simulations in ΛCDM that are capable of reproducing a number of observed properties of the Local Group. In particular, these have produced excellent agreement with observations of the bright end of the satellite galaxy luminosity function around MW like haloes, and in the Local Group more widely (Shen et al., 2014; Sawala et al., 2015, 2016b; Grand et al., 2016; Wetzel et al., 2016).

Alongside these advancements, further discoveries of Local Group dwarf galaxies have continued apace. A decade of searches using the Sloan Digital Sky Survey (SDSS), Dark Energy Survey (DES), and other Galactic surveys have now increased the total population of known satellite galaxies around the MW to ~60 such systems (e.g. Koposov et al., 2008; McConnachie, 2012; Bechtol et al., 2015; Drlica-Wagner et al., 2015; Koposov et al.,

2015a). Many of these are very faint and diffuse, vindicating the earlier predictions of semi-analytic galaxy formation models and further diminishing the status of the 'missing satellites problem' as a failure of ΛCDM. These surveys cover only a small fraction of the total virial volume of the MW, leaving open the possibility for a significant number of future discoveries of nearby ultrafaint dwarf galaxies and motivating work to estimate the size and properties of the total satellite galaxy population of the MW (Koposov et al., 2008; Tollerud et al., 2008; Hargis et al., 2014). We explore our own approach to this question in Chapter 2 and discuss the results in Chapter 3.

The steady development of the theoretical underpinnings of galaxy formation processes, set in the context of the ΛCDM framework for the formation and evolution of cosmological structure, seems to provide a ready solution to the 'missing satellites problem'. The broad confirmation of these theoretical insights with observational breakthroughs in the detection of ultrafaint, diffuse, low surface brightness dwarf galaxies, has further strengthened this argument. The strongest test will come as the resolution of hydrodynamic simulations pushes into the regime of the lowest mass subhaloes capable of hosting galaxies. When new surveys such as the LSST come online over the next decade they will see some of the faintest dwarf galaxies in the Universe and will provide some of the most detailed insights into the small scale regime.

1.3.2 Density profiles of DM haloes

Early theoretical calculations suggested that DM haloes experience 'violent relaxation' during their collapse, producing an isothermal density distribution of DM in the halo at late times (Lynden-Bell, 1967; Shu, 1978). This process occurs in collisionless systems that experience rapid changes in their gravitational potential (e.g. due to merging with other haloes), leading swiftly to a state of equilibrium that is independent of the details of the initial state. Later studies of massive DM haloes using early N–body simulations found that the density profiles scaled as $\rho(r) \propto r^{-2}$ in support of this conclusion, although they lacked sufficient resolution to probe the central regions of these haloes (Frenk et al.,

1985; Quinn et al., 1986).

It was not until the advent of the high-resolution simulations of Dubinski & Carlberg (1991) and Warren et al. (1992) that the small-scale structure in the centre of DM haloes was able to be investigated. These simulations suggested that a Hernquist profile (Hernquist, 1990), which scales as $\rho(r) \propto r^{-1}$ in the central regions of haloes, provides a better fit to the DM distribution in low-mass haloes down to $\sim$ kpc scales. This profile produces what is known as a 'cusp', where the density profile increases extremely rapidly with decreasing radius such that $\rho(r) \rightarrow \infty$ as $r \rightarrow 0$ and contrasts with the constant value at $r = 0$ that is obtained by an isothermal profile.

The nature of DM precludes direct observational measurements of the density profiles of haloes. However, the rotation curves of galaxies provide reasonable proxies to them. Measurements of this quantity are complicated somewhat by uncertainties in the mass-to-light ratio of the galaxy disc, which describes the mass associated with the baryonic material. We represent this mass-to-light ratio as Υ (Moore, 1994). A $\Upsilon \sim 1$ implies that the baryonic material dominates the mass profile in the central regions of the DM halo, while DM becomes more dominant at larger radii. In effect, this assumed ratio and the spatial extent of the galaxy defines a radial scale that marks the transition between these two regimes. Increasing Υ effectively decreases the contribution of DM to the central matter density, thereby increasing the scale radius that marks the regime transition from baryon-dominated to DM-dominated.

Dwarf galaxies are small and highly DM-dominated, making them ideal as probes of the inner density profile of haloes. The first attempt to fit the theoretically-determined DM density profiles to an observed galaxy rotation curve was made by Flores & Primack (1994), who obtained best agreement using an isothermal profile with a constant density core, such that inside some characteristic radius, r_c, the density $\rho(r < r_c) \propto r^0$. Later the same year this procedure was repeated by Moore (1994) using several other dwarf galaxy rotation curves, who found the same result: the rotation curves appeared to favour DM density profiles with constant density cores, in direct contradiction to the predictions of CDM N–body simulations.

Over the subsequent twenty years the search for solutions to this 'cusp–core problem', sometimes also known as the 'core catastrophe', spurred considerable efforts to understand the effect of baryons on the DM. It had been appreciated some years earlier that baryon-poor dwarf galaxies form via processes such as stellar evolution and supernovae feedback that lead to the ejection of baryonic material from the system (Dekel & Silk, 1986). In addition to the destruction of the baryonic disc, the expulsion of mass in this way would also affect the distribution of DM at the centre of the halo and might lead to the formation of a core (e.g. Navarro et al., 1996a; Gelato & Sommer-Larsen, 1999; Read & Gilmore, 2005; Pontzen & Governato, 2012). Another possible mechanism to produce cores from initially 'cuspy' DM profiles is from resonant effects induced by a stellar bar in the baryonic disc that transfers angular momentum to the halo. To be relevant for dwarf galaxies these bars would have had to form early in the galaxy assembly history, and it was argued that such structures could actually help to drive the evolution of the system towards its final diffuse, low surface brightness state (Weinberg & Katz, 2002). The effect of dynamical friction on infalling clumps of material was also shown to provide a potential means of transferring energy to the halo and flattening the inner DM cusp. This provided a plausible explanation for the survival of the large population of globular clusters in the Fornax dwarf galaxy, the orbits of which would otherwise have been expected to decay such that they had merged into the centre by the present day (El-Zant et al., 2001; Goerdt et al., 2006; Sánchez-Salcedo et al., 2006; Mashchenko et al., 2008; Cole et al., 2012).

At approximately the same time, new techniques to measure the inner slopes of density profiles of nearby dwarf galaxies seemed to find consistency with a wider range of values than had previously been allowed. One technique made use of dwarf spheroidal galaxies that have two kinematically and spatially distinct stellar subpopulations to place constraints on the inner density profile. Initial analyses using measurements of the stellar populations of the nearby dwarf galaxy, Sculptor, found that both cores and cusps are consistent with the new data (Battaglia et al., 2008; Amorisco & Evans, 2012), and later work even showed a preference for *cuspy* density profiles, rather than cores (Breddels et al., 2013; Richardson & Fairbairn, 2014; Strigari et al., 2017). This was disputed strongly by Walker

& Peñarrubia (2011), who developed a method to measure the mass profile slope using distinct stellar populations independently of assumptions about the particular DM halo model. Using spectroscopic measurements of two distinct metallicity subpopulations their results appeared to rule out robustly cuspy density profiles in both the Sculptor and Fornax dwarf spheroidals at high statistical significance. However, later work by Kowalczyk et al. (2013) and Genina et al. (2018) demonstrated that assuming that dwarf galaxies are spherical when they are in fact aspherical introduces a strong line-of-sight dependent bias into the result that can make dwarf galaxies appear to have a core even when the underlying density profile actually has a cusp.

By the time the new measurements of dwarf galaxies with apparently cuspy density profiles were made, a number of advanced hydrodynamic simulations had developed a clear preference for the formation of cores in haloes with masses above some threshold (Governato et al., 2012; Munshi et al., 2013; Madau et al., 2014; Chan et al., 2015; Oñorbe et al., 2015; Tollet et al., 2016; Fitts et al., 2017), while others such as APOSTLE (Sawala et al., 2016b) and Auriga (Grand et al., 2016) produced only cusps (Bose et al., 2019). The cause of the discrepancy between the simulations was not entirely clear, although suspicions fell primarily on differences in the subgrid prescriptions employed by different codes to model physical processes occurring on scales below the resolution limit of the simulation. A recent comparison study with the EAGLE simulations (Crain et al., 2015; Schaye et al., 2015) carried out by Benítez-Llambay et al. (2019) suggests that the choice of threshold density at which gas is converted into stars in such simulations is the source of the different behaviours and *not* the 'burstiness' or strength of supernovae feedback, as had previously been thought. This parameter is used to tune the subgrid prescriptions that control star formation in hydrodynamic simulations; as such it has no physical meaning and therefore cannot be constrained observationally. This serves to illustrate the difficulties inherent in performing comparisons of observations with such simulations and in attributing discrepancies to an underlying physical cause.

It is clear that a complete, physically-motivated treatment of baryonic processes is a crucial ingredient to understand the formation of structure inside DM haloes in a ΛCDM

cosmological framework. While theoretical advances and improvements in simulation technique have unquestionably furthered our understanding of these complex interactions, it seems that current state-of-the-art cosmological hydrodynamic simulations cannot provide reliable predictions of the inner DM density profile. In addition, while a number of measurements of the density profiles of local dwarf galaxies using a variety of techniques appear to favour cuspy profiles, the most recent observations of the Carina, Draco, Eridanus II and Fornax galaxies contradict this (e.g. Contenta et al., 2018; Hayashi et al., 2018; Pascale et al., 2018; Read et al., 2018; Boldrini et al., 2019). Measurements of the transverse velocity will eventually help but these are a number of years away, so a solution to the cusp–core problem will depend heavily on the progress of hydrodynamic simulations. In particular, it would seem that significant advances in the self-consistent modelling of star formation processes are the necessary next step if we are to make robust comparisons with observational measurements.

1.3.3 Planes of satellite galaxies

In the mid-1970s, it was noted almost simultaneously by two independent groups that the Magellanic Clouds and several of the then-known 'classical' satellites of the MW appeared to lie in a thin plane (Kunkel & Demers, 1976; Lynden-Bell, 1976). The kinematic properties of these objects are dictated by their accretion onto the MW, and although there was some disagreement over the exact orientation of the plane and precisely which satellites were members of this structure, it offers the potential to reveal important insights about the formation of the MW.

A key question raised by this configuration of satellite galaxies is whether such anisotropic accretion is expected in CDM models, or whether the MW is a 'less-than-typical' galaxy. An answer to this question depends fundamentally on the nature of structure formation in the Universe and more specifically, the particularities of the environment around the nascent MW. Unfortunately, the development of ΛCDM was not to be completed until the early 1990s, and a further decade of work was required before the predictions of structure

formation on the scales of the MW could be explored (see Section 1.2). Thus, from its first notice, the anisotropic distribution of satellites around the MW received little further attention for nearly thirty years.

After the turn of the 21st century, the alignment of the classical satellites was re-examined by Kroupa et al. (2005), who noted that in the standard paradigm substructures fall into DM haloes from cosmological filaments. This scenario implies the existence of preferential points of entry for accreting material into the host halo, which could affect the spatial distribution of the DM. The subhaloes were thought to follow a similar distribution so, under the assumption that the substructure provides a vehicle for the formation of most galaxies, this offers a possible explanation for the anisotropic distribution of the classical satellites. However, the observational evidence available at the time suggested that the spatial distribution of DM around the MW is not especially aspherical at large distances, and it could be approximated as isotropic (e.g. Olling & Merrifield, 2000; Ibata et al., 2001; Martínez-Delgado et al., 2004). Kroupa et al. (2005) found that the likelihood of drawing a disc-like plane of satellites, as is observed around the MW, from an approximately isotropic underlying DM subhalo distribution was less than 0.5 per cent. From this they concluded that the MW dwarf galaxies cannot reside within the DM substructures that simulations predicted compose most of the structure of the MW halo, posing a direct challenge to ΛCDM.

Shortly thereafter, Kang et al. (2005) and Zentner et al. (2005) considered the issue using *N*–body simulations of CDM haloes and, in the latter case, also with the application of semi-analytic galaxy formation models. Rather than *assuming* a completely isotropic parent distribution as Kroupa et al. (2005) had done, Kang et al. (2005) found that drawing the MW satellite galaxies from the slightly oblate, tri-axial spatial distribution of DM predicted by simulations improved the likelihood of the formation of a 'great disc' of satellites. In contrast, the subhaloes in the new *N*–body simulations are preferentially dispersed along the major axes of their hosts in a very flattened spatial configuration. Such an arrangement of the substructure around the MW might be expected to yield an improvement in the likelihood of forming a plane of satellites similar to that observed;

however, this was not to be. Both Kang et al. and Zentner et al. found that drawing the satellites from this population produced less favourable results.

A better outcome was achieved by assuming that the brightest satellites inhabited the most massive subhaloes. Zentner et al. (2005) found in both their DM-only and semi-analytic models that sampling from this population produced satellite galaxies with planar distributions similar to that seen around the MW. This was refined further by the extensively tested semi-analytic models employed by Libeskind et al. (2005). They demonstrated that the most important subhaloes for satellite galaxy formation are not the most massive at the *present* day but rather, those that had the most massive progenitor prior to accretion into the host halo. This is because the present-day mass of the subhalo can be influenced by processes such as tidal stripping, in which material is lost from the outer parts of the subhalo due to tidal interactions with the host. This process has a much smaller impact on the satellite galaxy itself, as most star formation takes place close to the centre of haloes before they fall into larger hosts. The spatial distribution of subhaloes selected by progenitor mass is highly biased and provides a good match to the observations of the classical satellite distribution, and indeed to later discoveries of MW satellites observed in the SDSS (Willman et al., 2005; Belokurov et al., 2006a,b; Zucker et al., 2006a,b; Irwin et al., 2007; Walsh et al., 2007; Wang et al., 2013).

This work demonstrated that the ΛCDM framework is not only capable of producing planes of satellites but also planes whose present-day configurations can resemble that seen in observations of the companions of the MW. What was less clear was whether the 'great disc' was a chance alignment of satellites on otherwise random orbits, or if they had a common origin. Pawlowski et al. (2012) provided the first tantalising hints at an answer to this question by uncovering a 'vast polar structure' of satellite galaxies, globular clusters, and streams of stars and gas around the MW. From this, they argued that the likelihood of a chance alignment of so many systems in such a way as to produce the thin plane is sufficiently remote that such a scenario could be ruled out. Instead, they favoured the explanation that the satellites and stellar systems formed at the same time from tidal debris thrown off in the past during a major interaction between the MW and another galaxy.

These conclusions seemed to be strengthened further a year later when measurements of the 6D motions of the classical satellites established that 60–80 per cent of them orbit in a seemingly rotationally supported, coherent structure (Pawlowski & Kroupa, 2013). This solution was appealing because it offered a natural explanation for the distribution and kinematics of the MW satellites, although it also implies that most of the MW satellite galaxies are actually tidal dwarf galaxies containing very little DM. If this is the case, the MW is not representative of typical MW-mass haloes that might be found elsewhere in the Universe.

Concurrent with these developments, observations of the satellite population of M31 suggested that a large fraction of its satellites also inhabited a thin plane (Conn et al., 2013; Ibata et al., 2013). Prior to this discovery, the existence of such a structure around the MW could be explained broadly as a very unusual but possible outcome of ΛCDM models. However, two such objects in close proximity are significantly less common. In the specific case of the MW and M31, this can be redeemed somewhat as they are within the same group. In this scenario, the filamentary accretion of satellites onto hosts during hierarchical growth provides a plausible explanation for the existence of coherent planes of satellites around *both* the MW and M31, while allowing for the expected rarity of satellite planes in general (e.g. Libeskind et al., 2009; Cautun et al., 2015). However, now that observations have confirmed the existence of a similar such kinematically coherent plane of satellites in the Centaurus A system (Tully et al., 2015; Müller et al., 2018), reconciling ΛCDM predictions with observations has become much more difficult.

The planes of satellite galaxies around the MW, M31 and Centaurus A, have provided fertile ground to test detailed predictions of the prevailing cosmological paradigm on small scales. While the ΛCDM model has demonstrated an ability to produce kinematically coherent systems, they are not expected to be common, and the existence of three such structures around hosts in the local Universe has proved difficult to reconcile with the standard paradigm. Additional observations of a statistically significant sample of host galaxies and measurements of the 6D position and velocity phase space of their satellite complements will enable more robust observational tests of the prevalence of these satellite

systems to be carried out. Improvements in the resolution and modelling of the physical processes in simulations will also assist in understanding environmental effects on these structures as well as any connection with halo properties. Today, the 'planes of satellites' problem remains an open challenge to the ΛCDM cosmological paradigm.

1.3.4 Too big to fail

A relatively recent addition to the catalogue of small-scale issues for ΛCDM emerged from a proposed solution to both the missing satellites and cusp–core problems (see sections 1.3.1 and 1.3.2). Early analytic calculations, supported by later hydrodynamic simulations, had shown that various baryonic processes play a significant role in the assembly of a luminous component inside subhaloes (e.g. Larson, 1974; White & Rees, 1978; Bullock et al., 2000; Benson et al., 2002a,b; Somerville, 2002). If the interplay of these complicated mechanisms conspires to preclude the formation of galaxies inside the majority of low-mass subhaloes, then the missing satellites problem would no longer present an issue. Rather, a large population of low-mass, 'dark' subhaloes accompanied by a handful of bright satellites that reside in more massive substructure appears to be a natural outcome of galaxy formation within the ΛCDM cosmological framework. Semi-analytic models of galaxy formation soon began to indicate that the most massive subhaloes in a system are most likely to host visible galaxies (Zentner et al., 2005). In particular, Libeskind et al. (2005) showed that subhaloes with higher peak masses at earlier times are the best predictor of the presence of a luminous baryonic component. This does not correlate perfectly with high mass at z=0 as various dynamical processes strip mass from subhaloes during their accretion onto the host, producing considerable scatter. The present-day mass of a given subhalo is, therefore, a function of its particular evolutionary history after accretion into the parent halo, and of its mass at the time of infall. Verifying this observationally is difficult as we cannot yet observe galaxy formation in such low-mass haloes at high redshift. This is where hydrodynamic simulations become extremely useful.

Boylan-Kolchin et al. (2011) were one of the first groups to investigate the preferred sites of galaxy formation around the MW using observations of nearby satellite galaxies. They compared the masses inferred from kinematic measurements of the centres of the satellites to the central masses of the most massive subhaloes in DM-only simulations of MW-like haloes. This avoids the necessity of accounting for the dynamical processes that affect the subhalo during accretion as the central regions are least likely to have been perturbed by such mechanisms up to the present time. These comparisons showed that the most massive simulated DM subhaloes are too centrally dense to host the MW satellite galaxies, implying that the satellites have formed preferentially in less massive subhaloes while the most massive structures failed to form galaxies at all (Boylan-Kolchin et al., 2011, 2012). This is at odds with galaxy formation theory in which the most massive, dense haloes are sufficiently dynamically 'hot' that gas that has been ionized by the UV background is still cool enough to sink to the bottom of the halo's gravitational potential well, where it is able to cool further to form stars. Such haloes should therefore be 'too big to fail' at forming a galaxy.

One possible solution to the 'too big to fail' (TBTF) problem concerns the estimated mass of the MW halo which is highly uncertain, with typical estimates in the range $(0.5–2.0) \times 10^{12}$ $M_\odot$ (Cautun et al., 2014b; Piffl et al., 2014; Wang et al., 2015). A MW halo mass at the lower end of this would have fewer massive subhaloes than suggested by the simulations and would offer a natural solution to the TBTF problem (Boylan-Kolchin et al., 2012; Wang & White, 2012; Vera-Ciro et al., 2013). However, such a low mass is unlikely based on a variety of measures such as the Local Group timing argument (Li & White, 2008), the analysis of the positions, line-of-sight velocities and proper motions of the MW satellites (Watkins et al., 2010; Li et al., 2017), the radial dependence of the Galactic escape speed (Piffl et al., 2014), and from the modelling of stellar streams in the Galactic halo (Küpper et al., 2015).

Another possibility is that the MW might simply be somewhat unusual and not representative of a 'typical' MW-mass halo as might be obtained from large cosmological simulations (Boylan-Kolchin et al., 2011). This idea was tested observationally using the satellites of

M31, and later on with the field galaxies of the Local Group which lie outside the virial radius of either M31 or the MW. Studies of these populations found that both groups of dwarf galaxies *also* appear to possess TBTF problems of their own, suggesting that the TBTF problem cannot be explained entirely by the peculiarities of the evolutionary history of the MW (Garrison-Kimmel et al., 2014; Kirby et al., 2014; Tollerud et al., 2014).

Perhaps the most compelling solution to the TBTF problem might be supplied by attempts to model the complex processes that influence the evolution of the baryonic components of small DM haloes. A number of comparative studies of field dwarf galaxies and MW-like systems in both DM-only and hydrodynamic simulations have suggested that the TBTF problem is peculiar only to the DM-only realizations (Brooks & Zolotov, 2014; Madau et al., 2014; Sawala et al., 2015; Dutton et al., 2016; Wetzel et al., 2016). Including in models a complete treatment of baryonic physics that encompasses feedback from supernovae and the destruction of subhaloes by the disc of the host galaxy appears to alleviate the discrepancies that originally gave rise to the TBTF problem. However, confirming this via comparisons with observations of dwarf galaxies is difficult, and is complicated by uncertainties in the assumptions required in order to interpret the observational measurements. An important example of this is the measurement of the kinematics of HI gas in dwarf galaxies, which is typically used to probe the DM halo at larger radii than the stellar component is able to. Observational results tend to support the existence of the TBTF problem even when compared with hydrodynamic simulations that attempt to model baryonic processes (Papastergis & Shankar, 2016). However, studies with synthetic radio data soon found that HI measurements systematically underestimate the mass enclosed, likely due to turbulence in the interstellar medium induced by supernovae (Verbeke et al., 2017; Oman et al., 2019). Such measurements therefore no longer offer a compelling test of the TBTF problem.

Numerical hydrodynamic simulations have helped to establish the importance of baryonic processes on both the formation of galaxies and their effect on the DM itself, and will no doubt play a continuing important role in the identification of a solution to the TBTF problem (e.g. Arraki et al., 2014; Brooks & Zolotov, 2014; Brook & Di Cintio, 2015;

Dutton et al., 2016; Sawala et al., 2016b; Tomozeiu et al., 2016; Wetzel et al., 2016, see also Section 1.3.2). However, relative to their DM-only counterparts the resolution that current simulations can achieve is poor. The TBTF problem manifests on scales on the order of 300–500 pc in dwarf galaxies, far below the scales that hydrodynamic simulations can resolve adequately. Many of the relevant baryonic processes that these simulations try to capture are therefore consigned to various subgrid prescriptions that vary enormously between different hydrodynamic schemes. While it has been demonstrated that such processes have a significant influence on the DM, it would be premature to claim that they offer a complete solution to the TBTF problem until this regime can be tested more thoroughly and self-consistently with more advanced simulations. The TBTF problem still potentially remains an outstanding and unsolved challenge to ΛCDM.

1.3.5 Summary of small-scale challenges to ΛCDM

The CDM model produced highly successful predictions of the structure on large scales which were enhanced further by the inclusion of $\Lambda>0$ in the field equations of General Relativity. The theoretical and numerical predictions of ΛCDM on large scales have now been verified by a diverse number of observational measurements, with such remarkable success that it has now become the standard cosmological paradigm. On smaller scales, the first hints at additional complexity beyond that predicted by theoretical calculations of its predecessor, CDM, came from observations of the spatial distribution of satellite galaxies around the MW (see Section 1.3.3). At the time it was difficult to read much into this as the Local Group had yet to be modelled well, and in any case predictions for particular DM haloes remained out of reach. This changed in the following decades with continual improvements in computational power and the availability of additional computational resources, enabling the development of more detailed numerical simulations of structure formation in a CDM universe.

By the early 1990s, DM-only simulations were finally able to resolve the internal structure of DM haloes of a similar scale to the MW. Almost immediately, discrepancies between

the CDM simulations and observations of nearby dwarf galaxies uncovered the cusp–core problem (see Section 1.3.2), and shortly thereafter comparisons of the number of dwarf galaxies around the MW with the number of subhaloes in simulations revealed the missing satellites problem (see Section 1.3.1). As these problems materialized in the highly non-linear regime of structure formation, the later transition to a cosmological model with a non-zero cosmological constant was unable to alleviate them. Therefore, these challenges apply equally to the ΛCDM model as well. This galvanized significant effort to understand better, and to model, the baryonic processes that drive star formation and stellar feedback and shape the evolution of DM, in the pursuit of a solution to these problems. While these cannot yet solve the cusp–core problem with any degree of certainty, it is now broadly accepted that the missing satellites problem can be explained entirely by these baryonic mechanisms.

Along the way, improvements to simulations of the Local Group have allowed the unusual spatial distribution of MW dwarf galaxies to be probed in more detail. These showed that while planes of satellites around host galaxies are somewhat unusual, they are not inconsistent with the predictions of ΛCDM (see Section 1.3.3). However, the existence of three such planes around objects in the local Universe is more challenging to explain and will require the accumulation of more observational measurements of a larger sample of hosts to understand fully. The same requirement can also be placed on one of the most recent additions to the catalogue of challenges to ΛCDM: the TBTF problem (see Section 1.3.4). While hydrodynamic simulations *appear* to show that the problem is mostly alleviated by the inclusion of baryonic physics, they are unable to probe the scales necessary to confirm this. In addition, systematics in observational measurements complicate comparisons with theoretical predictions, making TBTF an unsolved problem that is difficult to reconcile within the framework of ΛCDM.

In this section, we have considered several so-called ‘small-scale challenges’ to the standard cosmological paradigm. Such discrepancies indicate deficiencies in our theoretical understanding or modelling of physical processes, or more fundamentally, they could highlight deficiencies in the standard cosmological model itself. In our consideration of

possible solutions, we have focused primarily on reconciling these problems by modelling additional baryonic processes within the framework of ΛCDM. This has achieved varying degrees of success but some problems remain. If instead, we take these issues to be indicative of more fundamental structural issues in the ΛCDM framework itself, then we must look for an alternative cosmological model. In the next section, we will consider one particular class of cosmological models that might offer a solution to some of the outstanding small-scale challenges.

1.4 Warm dark matter

During the development of ΛCDM several models of DM were proposed that can be broadly categorized into three families: 'hot', 'cold', and 'warm' DM (the first two of these were introduced in Section 1.2). These are distinguished primarily by the velocity distribution of their DM particles which, in models of DM that produce the particle in thermal equilibrium with the primordial plasma, is closely tied to the DM particle mass. Less massive DM particles generally achieve higher thermal velocities at early times in the evolution of the Universe (i.e. are 'hotter'), free-streaming out of small density fluctuations and suppressing the formation of structure below a characteristic free-streaming scale. In the simplest DM models, this scale depends inversely on the mass of the DM particle, as illustrated by the late-time linear theory power spectra for examples of each DM model family in Fig. 1.1. In the hottest models, the suppression is so severe that only objects on the scale of galaxy clusters can form; such models were swiftly ruled out by comparisons of N–body simulations with observations of the large-scale structure in our Universe (White et al., 1983).

The 'cold' and 'warm' DM (WDM) models held more promise because they permitted the formation of smaller DM objects. Initially, CDM found most favour as it performed extremely well at reproducing objects on $\sim$ Mpc scales and was relatively easy to motivate from particle physics considerations. WDM models were also capable of achieving similar results on large scales but at the time lacked a physically motivated particle candidate

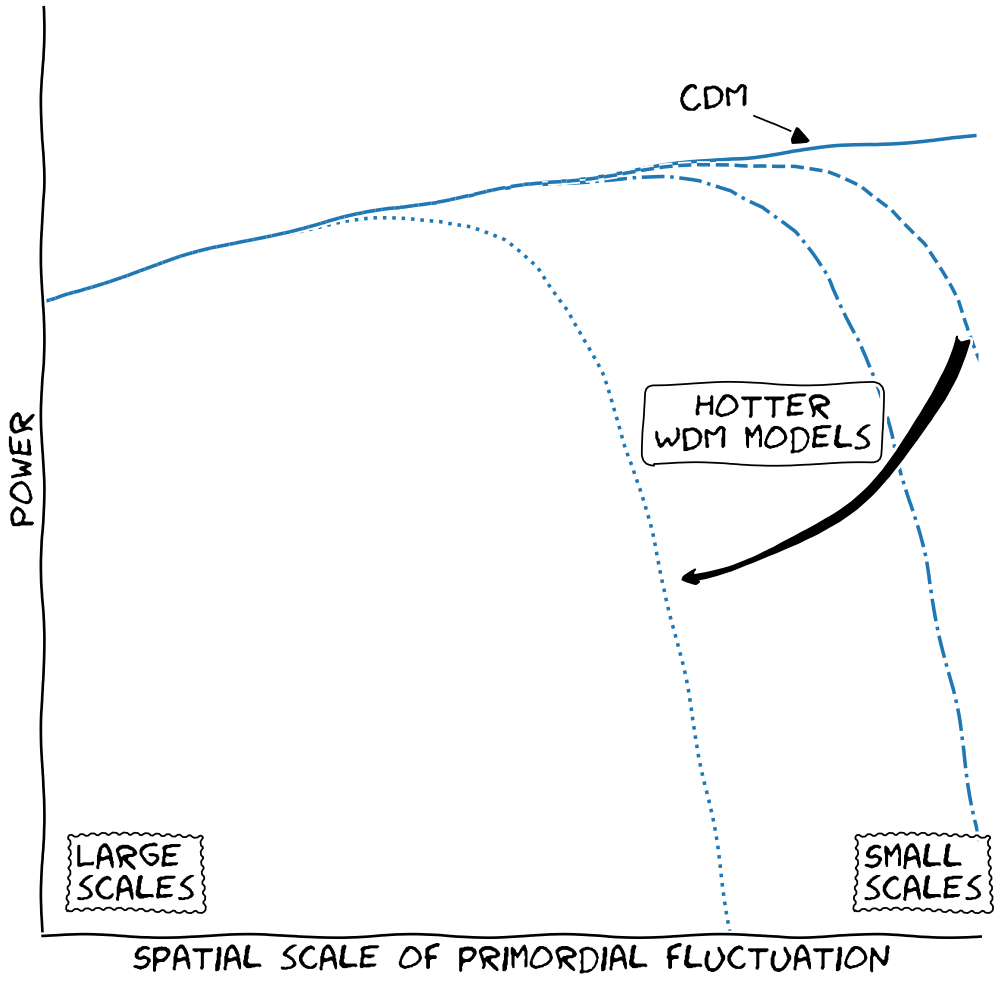


Figure 1.1
Diagram of the power spectra of different classes of DM model. Models that produce DM particles with higher velocity distributions suppress the formation of structures on progressively larger scales in comparison with 'cold', non-relativistic DM. Structures with spatial scales larger than that of the suppression scale are unaffected by the choice of DM particle and produce the same results as CDM.

(Bond et al., 1982; Olive & Turner, 1982; Pagels & Primack, 1982; Peebles, 1982b). As we discussed in Section 1.2, the CDM model continued to perform well as numerical *N*–body simulations improved; however, eventually, discrepancies emerged on both large (Efstathiou et al., 1990) and small (Davis & Peebles, 1983; Davis et al., 1985; Efstathiou, 1992) scales, which encouraged the exploration and reconsideration of other cosmological models.

One alternative that gained traction in the 1980s was a two-component model of 'cold + hot dark matter' (C+HDM; Shafi & Stecker, 1984; Davis et al., 1992). From a cosmological perspective, this was appealing as the combination of both components approximated to WDM models and circumvented the lack of an identified WDM particle candidate in the expected region of parameter space, which had been explored thoroughly by particle detectors. This also solved the problems produced by modelling either DM component as a single-component DM cosmology; specifically, that HDM models tended to suppress the formation of structure on scales below $\sim$7 Mpc in clear disagreement with observations, while CDM models alone could not supply enough power on large scales to account for that seen in surveys at the time (Efstathiou et al., 1990). However, this solution was generally disfavoured by the particle physics community because of the difficulty of extending the Standard Model of particle physics to incorporate it in such manner.

The introduction of a positive cosmological constant alleviated a number of discrepancies between theory and observation. However, no known particle has properties that enable it to fulfil the role of CDM. Today, several mechanisms have been proposed to generate WDM particles, either initially in equilibrium with the primordial plasma or through other means (e.g. Colombi et al., 1996; Bode et al., 2001). In the continued absence of the direct detection of a DM particle, ΛWDM models offer a compelling alternative to ΛCDM. Their viability can be assessed in an astrophysical context by comparing their predictions of the formation of small DM structures with the visible counterparts of these observed in the Universe today. We describe one approach to achieve this using the abundance of satellite galaxies of the MW in Chapter 4, choosing to focus on two of the simplest models of WDM: thermal relics and sterile neutrinos, both of which we introduce briefly in the

following subsections.

1.4.1 Thermal relics

Thermal relic models represent a generic, non-baryonic WDM particle that is produced initially in equilibrium with the primordial plasma. As the Universe cools and these particles decouple from radiation they stream away with relativistic velocities, becoming non-relativistic at later times and before matter–radiation equality (Avila-Reese et al., 2001; Bode et al., 2001). As discussed above, the relativistic speed, v, of thermal relic WDM allows the particles to free-stream out of density fluctuations, smoothing out inhomogeneities on scales below the free-streaming scale,

$$\lambda_{\rm FS}(t) = \int_0^t {\rm d}t' \frac{v(t')}{a(t')}, \tag{1.4.1}$$

where a is the scale factor, a measure of the relative expansion of the Universe over cosmic time, t. The streaming speed is a function of the momentum imparted to the DM particle at early times. Consequently it, and hence $\lambda_{\rm FS}$, is inversely proportional to the rest mass of the WDM particles. The free-streaming scale imprints onto the power spectrum of density fluctuations a suppression of large k-modes in Fourier space, while leaving smaller k-modes unchanged (see Fig. 1.1).

This behaviour was confirmed by comparisons of CDM N–body simulations with numerical simulations of structure formation in various thermal relic WDM models (Colombi et al., 1996). The exploration of the thermal relic WDM mass parameter space in large-scale structure simulations suggested that WDM was unable to match observational data of, for example, the clustering of galaxies and galaxy clusters when normalized to these scales, without choosing a very low particle mass akin to HDM. Thereafter, models of WDM generally fell out of favour; however, measurements of the accelerating expansion of the Universe using Type Ia supernovae provided a new context in which to consider them (Riess et al., 1998; Schmidt et al., 1998; Perlmutter et al., 1999). Much like the introduction of $\Lambda{\neq}0$ to CDM models, many of the large-scale problems encountered by solely WDM

models no longer presented an issue in ΛWDM and thermal relic particle masses that had previously been ruled out erroneously, were viable once again. In Chapter 4, we will revise the constraints on the thermal relic particle mass parameter space by drawing comparisons between small-scale structure formation in ΛWDM models and the observed population of MW satellite galaxies.

1.4.2 Sterile neutrinos

Sterile neutrino WDM was first proposed by Dodelson & Widrow (1994) as a simple extension to the family of 'ordinary' neutrinos in the Standard Model of particle physics. In this, fermions possess a fundamental property called 'chirality', which can be either 'left' or 'right'. All fermions of the Standard Model have both left- and right-chiral components, with the notable exception of the neutrinos, all of which are left-chiral. As gauge bosons only interact with left-chiral fermions and right-chiral anti-fermions, introducing right-chiral components (a.k.a. 'sterile' neutrinos) to the 'ordinary' neutrinos provides a mechanism to produce matter which does not couple to the fundamental forces and primarily interacts gravitationally. Such a mechanism also offers a natural explanation for neutrino flavour oscillations that were discovered later by Super-Kamiokande and the Sudbury Neutrino Observatory collaborations (Super-Kamiokande Collaboration et al., 1998; SNO Collaboration et al., 2001, 2002).

Unlike thermal relics, these particles are not produced in equilibrium with the primordial plasma. Instead, the most efficient way to produce sterile neutrinos is via mixing with the Standard Model neutrinos (e.g. Manohar, 1987). This process is controlled by the 'mixing angle', θ_M, that parametrizes the evolution in time of neutrino quantum states and relates to the abundance of different types of neutrino in the Universe. Under the simplest extension to the Standard Model of particle physics known as the Neutrino Minimal Standard Model (νMSM), three right-chiral neutrinos are introduced, all of which interact extremely weakly via the fundamental forces (Asaka & Shaposhnikov, 2005; Asaka

et al., 2005; Canetti et al., 2013a,b). Two of these have masses between ∼100 MeV∧ and ∼245 GeV (the electroweak scale), while the third could have a mass of $\mathcal{O}$(1 keV) and could act as the WDM.

The two most massive sterile neutrino species rapidly decay into leptons. This produces a slight overabundance of leptons compared with anti-leptons which can explain the asymmetry between the density of baryons and anti-baryons in the Universe. The underlying reason for this dominance of matter over anti-matter has preoccupied astroparticle physicists for over half a century (e.g. Sakharov, 1967a,b, 1991; Canetti et al., 2012), and this mechanism of the νMSM offers a natural solution to this unsolved question. The lepton asymmetry can be encapsulated by the lepton asymmetry parameter, L_6, and relates to the sterile neutrino mixing angle, θ_M, providing a potential observational proxy of sterile neutrino properties. Unfortunately, L_6 is not yet constrained well observationally, so it is effectively a free parameter. Variations in θ_M, represented by changes in L_6, can affect the power spectrum of a sterile neutrino of fixed mass, m_s, in non-trivial ways, allowing such WDM candidates to circumvent a number of astrophysical constraints on the mass of thermal DM particles. Sterile neutrino models can, therefore, be fully parametrized by a combination of m_s and L_6. In Chapter 5 we discuss the constraints that we obtain on this parameter space by comparing the suppression of the formation of small-scale DM structure with the observed abundance of satellite galaxies around the MW.

1.5 Thesis outline

In this thesis, we pursue two primary lines of enquiry. In the first strand, we seek to address the current gap in observational capability concerning the nearby dwarf galaxies of the Local Group. Observations of this population of objects now span a significant fraction of the sky but struggle to detect the faintest objects beyond a few tens of kpc from the MW. In Chapter 2, we develop and test a Bayesian approach to infer the luminosity function of

∧When discussing the masses of fundamental particles we choose to follow the convention established by the particle physics community and use natural units, setting the speed of light, c=1.

the satellite galaxies of the MW using high-resolution DM-only simulations of MW-mass host haloes and partial observations of this population by the SDSS and DES. We consider estimates of the luminosity function in Chapter 3 and explore a number of theoretical dependencies that could influence the final result, such as the congruence of theoretical and observational 'tracers' of the underlying DM structure, the assumed mass of the MW halo, and the volume considered in the calculation. We conclude this strand with a look to future observations of the local Universe, making predictions for the satellite population that might be observed by the forthcoming Large Synoptic Survey Telescope (LSST).

In the second strand, we turn our attention to alternative models to ΛCDM, choosing to focus on the WDM class of cosmological models and their predictions of structure formation on small scales. In Chapter 4, we develop an improved method to constrain the viable parameter spaces of WDM models by comparing their predictions of the abundance of small-scale structure with the satellite population of the MW inferred in the preceding chapter. We demonstrate the efficacy of this methodology by placing constraints on the generic class of thermal relic WDM models and consider further the effect on these of different prescriptions of baryonic physics. We carry out the same procedure for νMSM sterile neutrino WDM models in Chapter 5, motivated by the recent observations of an unexplained ∼3.5 keV line, setting our consideration of the continued viability of the standard cosmological paradigm in the context of this ongoing discussion.

Finally, in Chapter 6, we summarize our findings and consider future lines of enquiry that could test further our present understanding of the Universe.

CHAPTER 2

A Bayesian approach to infer the satellite galaxy luminosity function of the Milky Way

---

2.1 Introduction

Proposed in the 1980s (e.g. Peebles, 1982b; Blumenthal et al., 1984; Davis et al., 1985), the ΛCDM model has proved remarkably successful at predicting numerous observable properties of the Universe and their evolution over time; as a result, it has become the 'standard model' of cosmology (see Frenk & White, 2012; Weinberg et al., 2015, for recent reviews). Hierarchical structure formation is fundamental to this model, which predicts that DM haloes form by mergers of smaller haloes and smooth mass accretion. Merged (sub)haloes that are not completely disrupted are detectable today as satellite galaxies and, potentially, as non-luminous substructures.

The MW halo and its associated satellite galaxies offer an ideal environment in which to probe hierarchical growth which, in turn, can be used to constrain the faint end of galaxy formation and the properties of the DM. However, the current census of MW satellite galaxies is highly incomplete. The most recent surveys—such as the Sloan Digital Sky Survey (SDSS; Alam et al., 2015) and the Dark Energy Survey (DES; Bechtol et al., 2015;

Drlica-Wagner et al., 2015)—do not cover the entirety of the sky and are also subject to detectability limits that depend on the surface brightness of, and distance to the satellite galaxies. The goal of this chapter is to overcome some of these limitations and, using theoretical priors based on cosmological simulations of MW-like haloes, to estimate the expected total number of MW satellite galaxies.

In the 1990s, DM-only CDM simulations showed that many more subhaloes survive within MW-like haloes than there are visible satellites orbiting the MW (Klypin et al., 1999; Moore et al., 1999; Springel et al., 2008). This disparity is often referred to as the 'missing satellites problem for cold dark matter' (see Section 1.3.1). This rather unfortunate nomenclature is very misleading if, as is common usage, the word 'satellite' is taken to mean a visible galaxy: DM-only simulations have, of course, nothing to say about visible galaxies. Simple processes, at the heart of galaxy formation theory, such as the reionization of hydrogen in the early Universe and supernovae feedback, make it impossible for visible galaxies to form in the vast majority of CDM haloes. Such processes were first discussed and calculated in this context using semi-analytic techniques with different approximations in the early 2000s (Bullock et al., 2000; Benson et al., 2002a,b; Somerville, 2002). For example, Benson et al. (2002a) showed how the abundance and stellar content of dwarf galaxies are driven by reionization and supernovae feedback. Their model produced an excellent match to the luminosity function of the (11 'classical'—the only known at the time) satellites of the MW and predicted that the MW halo should host a large population of fainter satellites. Just such a population was discovered several years later in the SDSS (Koposov et al., 2008, and references therein).

The early semi-analytic results have been confirmed using full hydrodynamic simulations (e.g. Okamoto et al., 2005; Macciò et al., 2007). For example, the most recent such simulations have confirmed that below a certain halo mass, typically $\sim 10^{10}$ $M_\odot$, dwarf galaxy formation is strongly suppressed, and that the majority of haloes with masses $\lesssim 10^9$ $M_\odot$, should not host a luminous component (stellar mass greater than 10^4 $M_\odot$) (Shen et al., 2014; Sawala et al., 2015, 2016a; Wheeler et al., 2015).

In recent years, alternatives to CDM have elicited considerable interest. Some of these,

such as Warm Dark Matter (WDM, Avila-Reese et al., 2001; Bode et al., 2001), models with interactions besides gravity between DM particles and photons or neutrinos (Bœhm et al., 2014), and axionic DM (Marsh, 2016), predict a cut-off in the primordial matter power spectrum on astrophysically relevant scales, which would suppress the formation of small galaxies (Bode et al., 2001; Polisensky & Ricotti, 2011; Lovell et al., 2012; Schewtschenko et al., 2015). The abundance of the faintest galaxies can thus, in principle, reveal or rule out the presence of a power spectrum cut-off. By requiring that WDM models should produce at least enough substructures to match the observed Galactic satellite count, constraints on the mass and properties of the DM particle can be derived (Macciò & Fontanot, 2010; Lovell et al., 2014; Kennedy et al., 2014; Schneider, 2016; Bose et al., 2017; Lovell et al., 2017).

Past and current surveys have now discovered a plethora of satellites around the MW, with the count currently standing at 56: 11 classical satellites, 17 discovered in each of the SDSS and DES surveys, and 11 found in other surveys. Despite this relatively large number of known satellites, current estimates suggest that there could be at least a factor of 3–5 times more still waiting to be discovered (Koposov et al., 2008; Tollerud et al., 2008; Hargis et al., 2014). These estimates were made prior to the DES and are based only on SDSS data. These predictions start from an assumed radial profile for the distribution of Galactic satellites: either that it follows the DM density profile—as in Koposov et al. (2008, hereafter K08), which is not a good assumption—or that it follows the subhalo number density profile (as in the other studies cited above). Then, for each observed satellite, they calculate the number of satellites in the entire fiducial volume that must be present in order to have, on average, one object with the corresponding properties within the survey volume.

This and the subsequent chapter improve upon previous estimates of the Galactic satellite count in three major ways. First, while previous studies were based on SDSS data alone, our result makes use of the combined SDSS and DES data, which together cover an area equivalent to nearly half of the sky. Secondly, to properly account for stochastic effects, we introduce a new Bayesian approach for estimating the total satellite count. Stochastic

effects—which we find to be the leading cause of uncertainty—have been overlooked in previous studies, resulting in a significant underestimation of their errors. Finally, we make use of a set of five high-resolution simulated host haloes—taken from the Aquarius Project (Springel et al., 2008)—to characterize uncertainties arising from host-to-host variation. In December 2016, Jethwa et al. (2018) presented a Bayesian estimate of the total number of Galactic satellites. Their result is the outcome of applying abundance matching to the SDSS observations and, while it properly accounts for stochastic effects, it depends on more and uncertain assumptions (mostly related to abundance matching) than the result presented here.

We organize this chapter as follows. Section 2.2 introduces the observational data set used in this analysis and Section 2.3 describes, tests, and compares our Bayesian technique with previous works. We present a summary of the methodology in Section 2.4. We present our main results in Chapter 3, detailing their sensitivity to the assumed MW halo mass and the radial dependence of the satellite count.

2.2 Observational Data

Very few of the current set of MW satellites were known prior to the start of the 21st century. Discoveries made after this time, using a multitude of techniques, together with data from SDSS data release 2 (DR2) and the Two Micron All-Sky Survey (2MASS)—before a major advance with SDSS DR5 (Adelman-McCarthy et al., 2007)—brought the total to 23 dwarf galaxies. Since then, the SDSS survey area has nearly doubled and DES is now electronically available. Combining the two surveys produces a sky coverage area of 47 per cent, with SDSS and DES contributing 14 555 and 5000 square degrees, respectively. An analysis of DES data added a further 17 dwarf galaxies to the running total (Bechtol et al., 2015; Drlica-Wagner et al., 2015; Kim et al., 2015; Koposov et al., 2015a), which, together with other discoveries, brings the total number of dwarf galaxies, as of February 2018, to 56. These are listed in Tables 2.1 and 2.2.

Table 2.1

Known MW satellite galaxies identified in surveys used in this analysis, grouped according to the survey in which they were detected. For each satellite we provide its absolute $V-$band magnitude, M_V, heliocentric distance, $D_\odot$, and—for DES satellites—its probability of association with the LMC.

Satellite	M_V	$D_\odot$ (kpc)	p^a_{LMC}	Reference[e]
		Classical		
Carina	-9.1	105		
Draco I	-8.8	76		
Fornax	-13.4	147		
Leo I	-12.0	254		
Leo II	-9.8	233		
LMC	-18.1	51		
Ursa Minor	-8.8	76		
SMC	-16.8	64		
Sculptor	-11.1	86		
Sextans	-9.3	86		
Sagittarius I	-13.5	26		
		SDSS DR9		
Boötes I	-6.3	66		
Boötes II	-2.7	42		
Canes Venatici I	-8.6	218		
Canes Venatici II	-4.9	160		
Coma	-4.1	44		
Hercules	-6.6	132		
Leo IV	-5.8	154		
Leo V	-5.2	178		
Leo T	-8.0	417		
Pegasus III	-3.4	215		(1)
Pisces I[b]	...	80		(2)
Pisces II	-5.0	182		
Segue I	-1.5	23		
Segue II	-2.5	35		
Ursa Major I	-5.5	97		
Ursa Major II	-4.2	32		
Willman I	-2.7	38		
		DES		
Cetus II[c]	0.0	30	0.00[d]	(3)
Columba I	-4.2	183	0.11	(4)
Eridanus II	-7.1	366	0.00[d]	(5)
Eridanus III[c]	-2.4	95	0.00[d]	(3)
Grus I[c]	-3.4	120	0.64	(3)

Table 2.1
Continued...

Satellite	M_V	$D_\odot$ (kpc)	p^a_{LMC}	Reference[e]
Grus II[c]	-3.9	53	0.57	(3)
Horologium I	-3.5	87	0.79	(3, 6)
Horologium II[c]	-2.6	78	0.80	(3)
Indus II[c]	-4.3	214	0.19	(3)
Phoenix II[c]	-3.7	95	0.75	(3)
Pictoris[c]	-3.7	126	0.62	(3)
Reticulum II	-3.6	32	0.75	(3, 6)
Reticulum III[c]	-3.3	92	0.58	(3)
Tucana II	-3.9	58	0.75	(3, 7)
Tucana III[c]	-2.4	25	0.52	(3)
Tucana IV[c]	-3.5	48	0.79	(3)
Tucana V[c]	-1.6	55	0.81	(3)

[a] Obtained from Jethwa et al. (2016, Fig. 9).
[b] The method of detection was different to that applied to other satellites in the SDSS survey.
[c] Not spectroscopically confirmed.
[d] No probability of association with LMC provided.
[e] Data reproduced from McConnachie (2012, tables 2 and 3) unless indicated otherwise: (1) Kim et al. (2015, 2016), (2) Watkins et al. (2009), (3) Drlica-Wagner et al. (2015, Table 4), (4) Carlin et al. (2017), (5) Li et al. (2017), (6) Koposov et al. (2015b), (7) Walker et al. (2016).

Table 2.2

Known MW satellite galaxies identified in surveys *not* used in this analysis, grouped according to the survey in which they were detected. We provide the same data for each satellite as described in Table 2.1.

Satellite	M_V	$D_\odot$ (kpc)	Reference[b]
	VST ATLAS		
Aquarius II	-4.2	108	(1)
Crater II	-8.2	118	(2)
	Pan-STARRS		
Draco II	-2.9	20	(3)
Sagittarius II[a]	-5.2	67	(3)
Triangulum II	-1.2	28	(4)
	SMASH		
Hydra II	-4.8	134	(5)
	HSC		
Virgo I[a]	-0.3	91	(6)
Cetus III[a]	-2.4	251	(7)
	MagLiteS		
Carina II	-4.5	37	(8)
Carina III[a]	-2.4	28	(8)
Pictoris II[a]	-3.2	45	(9)

[a] Not spectroscopically confirmed.

[b] Data reproduced from: (1) Torrealba et al. (2016b), (2) Torrealba et al. (2016a), (3) Laevens et al. (2015), (4) Carlin et al. (2017), (5) Martin et al. (2015), (6) Homma et al. (2016), (7) Homma et al. (2018), (8) Torrealba et al. (2018), (9) Drlica-Wagner et al. (2016).

These discoveries resulted from the use of advanced search algorithms that comb through survey data and identify overdensities of stars which could signal the presence of a faint dwarf galaxy. For example, the SDSS has been analysed with two such search algorithms, by K08 and Walsh et al. (2009, hereafter W09), to find that both techniques recover the same number of dwarf galaxies—although the latter is sensitive to fainter objects. Each algorithm has a response function that—among other factors such as the survey surface brightness limits—is dependent on the absolute magnitude of the objects being searched for. Assuming isotropy, the number of observed satellites per unit absolute magnitude, $\mathrm{d}N_{\mathrm{sat}}/\mathrm{dM}_V$, is given by

$$\frac{\mathrm{d}N_{\mathrm{sat}}}{\mathrm{dM}_V} = \int_0^\infty \int_0^\infty \Omega r^2 \frac{\mathrm{d}^3 N_{\mathrm{sat}}}{\mathrm{d}r \ \mathrm{dM}_V \ \mathrm{d}r_{\mathrm{sat}}} \, \epsilon(r, \mathrm{M}_V, r_{\mathrm{sat}}) \, \mathrm{d}r \, \mathrm{d}r_{\mathrm{sat}} \,, \tag{2.2.1}$$

where the first integral is over the survey volume, with Ω the survey solid angle and r the radial distance from the Sun. The second integral is over the satellite size, r_{sat}; N is the distribution of satellites as a function of radial distance from the Sun, absolute magnitude, M_V, and size, r_{sat}. The last term, ϵ, denotes the efficiency of the search algorithm for identifying a satellite of magnitude, M_V, and size, r_{sat}, at distance, r, averaged over the survey's sky-footprint. At fixed absolute magnitude, most of the satellites detected in the SDSS have similar sizes and the detection efficiency, ϵ, is approximately equal for all objects (K08; W09). Thus, for the observed satellites, the dependence on r_{sat} in equation (2.2.1) can be approximated as a dependence on M_V alone.

The detection efficiency, ϵ, at fixed M_V, is a function of the radial distance and shows a rapid transition with radius from a 100 per cent to a 0 per cent chance of detection. We may therefore define an equivalent effective detection volume such that, on average, this effective volume includes the same number of satellites of magnitude M_V as predicted by equation (2.2.1). The effective radius, $R_{\mathrm{eff}}\,(\mathrm{M}_V)$, corresponding to this effective detection volume, is computed by solving the equation,

$$\frac{\mathrm{d}N_{\mathrm{sat}}}{\mathrm{dM}_V} = \int_0^{R_{\mathrm{eff}}(\mathrm{M}_V)} \Omega r^2 \, \mathrm{d}r \frac{\mathrm{d}^2 N_{\mathrm{sat}}}{\mathrm{d}r \ \mathrm{dM}_V} \,, \tag{2.2.2}$$

where the left-hand term is given by equation (2.2.1) and R_{eff} appears as the upper limit of

Table 2.3
The parameters of equation (2.2.3) quantifying the dependence on absolute V-band magnitude of the effective radius in the SDSS and DES surveys. The K08 parameters are taken from fits by W09.

Survey	Algorithm	a^*	b^*
SDSS	Koposov et al. (2008, K08)	0.205	1.72
	Walsh et al. (2009, W09)	0.187	1.58
DES	Jethwa et al. (2016, J16)	0.228	1.45

the integral. The value of $R_{\rm eff}$ depends on both the radial dependence of ϵ and on the radial distribution of satellites. As long as the radial distribution of satellites is nearly constant in the interval where the detection efficiency drops from 100 to 0 per cent, $R_{\rm eff}$ can be approximated as the radius at which the detection efficiency is 50 per cent, which is the value that we use in the rest of this and later chapters. This approximation is reasonable as ϵ decreases from 1 to 0 over a narrow radial range (e.g. see fig. 15 in W09). Making another choice for the effective radius, such as $\epsilon = 0.9$ (as used in Hargis et al. 2014), would underestimate the effective volume and thus overestimate the inferred satellite count. Both K08 and W09 show that, to good approximation, the effective detection radius, which corresponds to $\epsilon = 0.5$, is given by

$$R_{\rm eff}\,(\mathrm{M}_V) = 10^{(-a^*\mathrm{M}_V - b^*)}\ \mathrm{Mpc}\,, \tag{2.2.3}$$

where a^* and b^* are fitting parameters associated with the search algorithm response function. These values are provided in Table 2.3 for different algorithms.

The dependence of the effective radius on absolute V-band magnitude for the SDSS and DES surveys is shown in the upper panel of Fig. 2.1. For clarity, in the case of the SDSS we show only the W09 response function. For DES we give the Jethwa et al. (2016, hereafter J16) response function that was shown to give a good match to the actual detections. This is equal to the K08 response function as fitted by Tollerud et al. (2008, hereafter T08), but shifted to account for the additional depth of the DES compared to SDSS; however, this response function has not been verified at the same level of in-depth analysis as in e.g. W09. The figure shows that for the same absolute magnitude, DES is deeper and thus can detect satellites out to greater distances than SDSS. All bright dwarfs, i.e. $\mathrm{M}_V < -5.5$ for SDSS

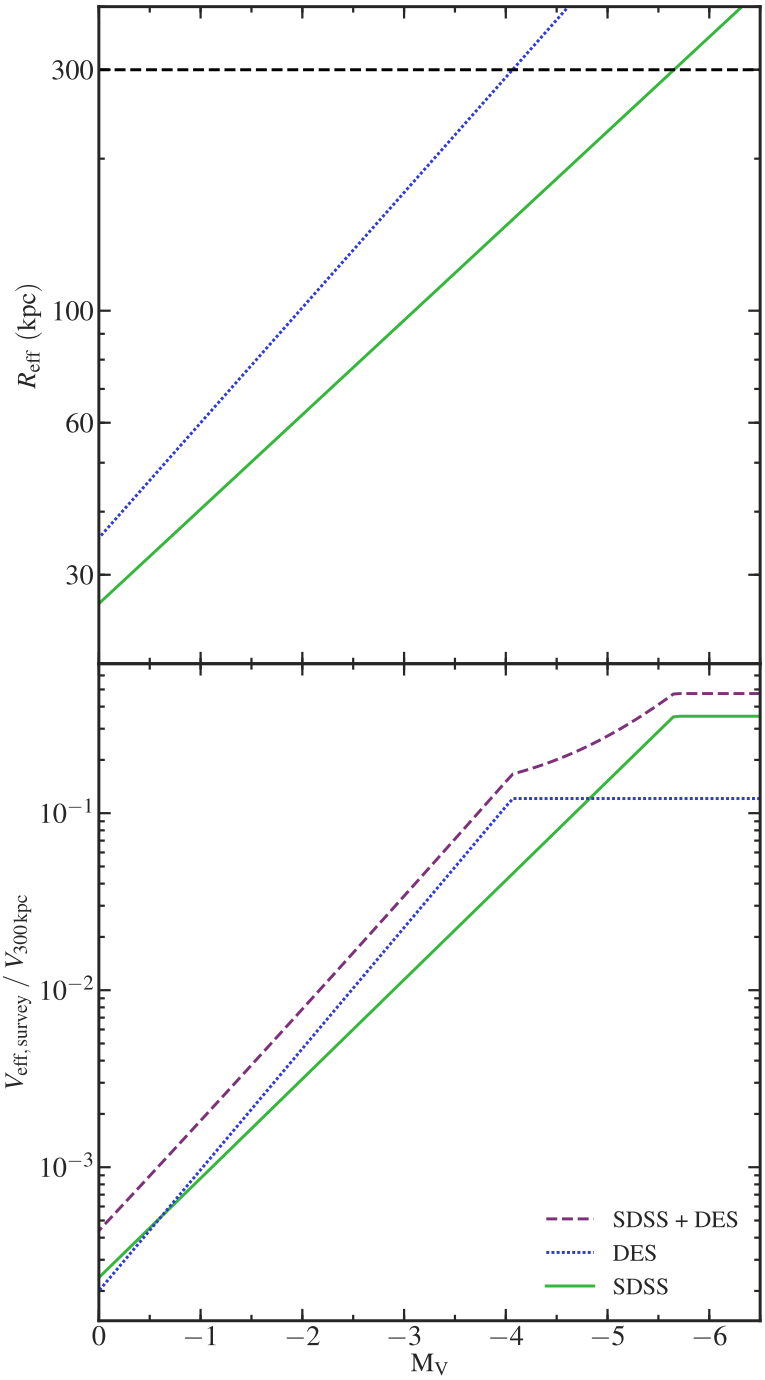


Figure 2.1

Upper panel: the effective detection radius, $R_{\rm eff}$, of satellites as a function of absolute magnitude, M_V, for the SDSS and DES surveys. The horizontal dashed line indicates our fiducial choice of outer radius, $R_{\rm out}$=300 kpc, for the MW satellite population. *Bottom panel:* the ratio of the effective volume surveyed by SDSS and DES, as a function of M_V, to the volume enclosed within 300 kpc. The dashed line shows the combined SDSS plus DES effective volumes. The two panels show the response functions of the W09 and J16 search algorithms, which are given in Table 2.3.

and $M_V < -4.0$ for DES, that are within the survey footprint and within our fiducial choice of outer radius, R_{out}=300 kpc, should have been detected within their respective surveys. Thus, the surveys may be considered 'complete'—for the purposes of this analysis—at the absolute magnitudes at which R_{eff} is greater than 300 kpc. Fainter objects can be detected only if they are closer than 300 kpc from the observer, with the faintest, M_V=0, dwarfs being detected only if they are within ∼30 kpc of the Sun. The survey response functions described here are a simplified representation of the response of the survey and the search algorithms applied to the survey data. In Appendix A, we consider the detailed effect of various parameters such as dwarf galaxy size and distance, and the characteristics of the observing scheme used. We discuss how these effects are incorporated into the method described in this Chapter, and any effects these might have on the results presented in Chapter 3.

To obtain a more informative perspective on the survey completeness, the bottom panel of Fig. 2.1 shows the ratio between the effective volume of each survey and the total volume enclosed within our fiducial radius of 300 kpc. The effective volumes of the independent conical survey regions of the SDSS and DES were calculated from the corresponding values of R_{eff} given by equation (2.2.3). The total volume probed by these surveys is given by the sum of the two survey volumes, and the fractional volume probed by these is indicated by the long dash line in Fig. 2.1. Even when combining the SDSS and DES footprints, the observations cover only ∼10 per cent of the fiducial volume at $M_V = -4$ and less than 0.1 per cent of the same volume at M_V=0.

2.3 Methods

We require two key ingredients to estimate the total population of satellite galaxies from a given survey of the MW. First, we need a prior for the radial distribution of satellites. For this we take the radial number density of subhaloes in simulations of MW analogues from the Aquarius Project, which, when subhaloes are selected by $v_{\rm peak}$—the highest maximum circular velocity achieved in the subhalo's history—is the same as the radial distribution of luminous satellites in hydrodynamic simulations and that of observed MW satellites (see Section 2.3.1). Secondly, we introduce and test our Bayesian framework used to infer the total number of satellites (Section 2.3.2). The need for a new methodology is motivated by several shortcomings of previous approaches, which we discuss in detail in Section 2.3.3.

We assume that the classical satellites, i.e. those with $M_V \leq -8.8$, are bright enough to have been observed by pre-SDSS surveys and that the observations are complete at these magnitudes (therefore ignoring the possible existence of concealed satellites in the Zone of Avoidance, the area of the sky that is obscured by the dust and stars in the plane of the MW). As such, the inferred luminosity function at the bright end will always match the observations, in line with previous studies (e.g. T08). The inference method is only applied to fainter satellites, that is, those with $M_V > -8.8$.

2.3.1 Tracer population

Any estimation of the total satellite count from incomplete observations needs a prior for the radial number density of these objects, which we estimate from N-body simulations. An ideal simulation from which to extract a tracer population should have high enough resolution for the density profile to be well sampled, and should also offer access to multiple realizations of MW-like haloes to account for host-to-host variations.

The Aquarius suite of simulations (Springel et al., 2008) achieves this. It consists of a set of six ΛCDM DM-only N-body simulations of isolated MW-like haloes which were run using the GADGET3 code and were labelled Aq-A to Aq-F. In this work we use the

Table 2.4
The DM particle mass, $m_{\rm p}$, softening length, ϵ, and host halo mass, $\rm M_{200}$, of the Aquarius simulations used in this work. Here, $\rm M_{200}$ denotes the mass inside the radius, R_{200}, within which the mean density equals 200 times the critical density.

Simulation	$m_{\rm p}$ ($\rm M_\odot$)	ϵ (pc)	$\rm M_{200}$ $\left(10^{12}\ \rm M_\odot\right)$
Aq-A1	1.712×10^3	20.5	1.839
Aq-A2	1.370×10^4	65.8	1.842
Aq-B2	6.447×10^3	65.8	0.819
Aq-C2	1.399×10^4	65.8	1.774
Aq-D2	1.397×10^4	65.8	1.774
Aq-E2	9.593×10^3	65.8	1.185

'level 2' simulations (L2, with a particle mass of $\sim 10^4$ $\rm M_\odot$), which corresponds to the highest resolution level available across all of the Aquarius haloes. Details of these simulations are provided in Table 2.4. The Aq-F halo experienced a late-time merger, making it unsuitable as representative of the MW halo; consequently, it is not used in this analysis. The cosmological parameters assumed for these simulations are derived from the WMAP first-year data release (Spergel et al., 2003): $H_0 = 73\ \rm km\,s^{-1}\,Mpc^{-1}$, $\Omega_{\rm M} = 0.25$, $\Omega_\Lambda = 0.75$, $n_s = 1.0$, $\sigma_8 = 0.9$.

Identifying subhaloes near the centre of simulated haloes using configuration space halo finders like SUBFIND can be difficult (Springel et al., 2008; Onions et al., 2012). In regions of high background density, these algorithms can struggle to identify substructures, a problem that is not addressed by improvements in simulation resolution. Subhalo finders are also affected by the resolution of the simulation to which they are applied; these effects can be assessed by comparing haloes which have been simulated at different resolution levels. One of the haloes in the Aquarius suite (Aq-A) was simulated at extremely high-resolution ('Level 1' or L1, with particle mass of $\sim 10^3$ $\rm M_\odot$). Even though the resolution of L2 is still very high, the abundance of subhaloes that are relevant to our analysis is suppressed relative to that at L1, particularly in the inner regions of the halo. The difference between the two levels is comparable to that seen across all other L2 profiles.

We can correct for these resolution effects in a relatively straightforward manner by using the Durham semi-analytic model GALFORM (Lacey et al., 2016; Simha & Cole, 2017) to populate the haloes and subhaloes in the Aquarius simulations with galaxies and track

their orbital evolution even after its halo is no longer resolved (the so-called 'orphan' galaxies). First, the semi-analytic model GALFORM described by Lacey et al. (2016), which is based on the same cosmology as the Aquarius simulation suite, is applied to each of the Aquarius DM haloes in turn. We use the Simha & Cole (2017) merging scheme to track the dynamical evolution of subhaloes over the course of cosmic time. Well-resolved subhaloes are tracked directly by the N-body simulation; however, those that fall below the resolution limit are lost. Simha & Cole recover this population by tracking the most bound particle in these subhaloes from the last epoch at which they were associated with a resolved subhalo. They then remove subhaloes from this population if one of the following criteria is satisfied:

(i) A time has elapsed after the last epoch at which the subhalo was resolved, which is equal to or greater than the dynamical friction timescale.

(ii) The subhalo passes within the halo tidal disruption radius at any time.

In both of the above cases the effects of tidal stripping on the subhalo are ignored, as are interactions between orbiting subhaloes.

In Fig. 2.2 we compare the normalized cumulative radial subhalo counts of the Aquarius A1 and A2 haloes with the $v_{\rm peak} \geq 10$ km s^{-1} selection threshold applied. Prior to the application of GALFORM the original normalized subhalo counts are highly discrepant in the inner regions of the haloes. The spread in the predicted counts at $M_V = 0$ in Aq-A1 and Aq-A2 is also wider than the spread in predictions from the other L2 haloes (B2–E2). When correcting for the 'orphan' population, which is very centrally concentrated, the discrepancy in the Aq-A1 and Aq-A2 normalized subhalo counts is almost completely eliminated. As a result the spread in the $M_V = 0$ predictions is also reduced such that it is much smaller than the spread in the predictions from the other 'L2 + orphans' haloes. The spread in these latter predictions is also significantly reduced by the correction, which shows that failing to account for this artificially inflates the halo-to-halo scatter.

A further factor that needs to be taken into account is the possible destruction of satellite galaxies by tidal interactions with the central galaxy in the halo. This effect has been

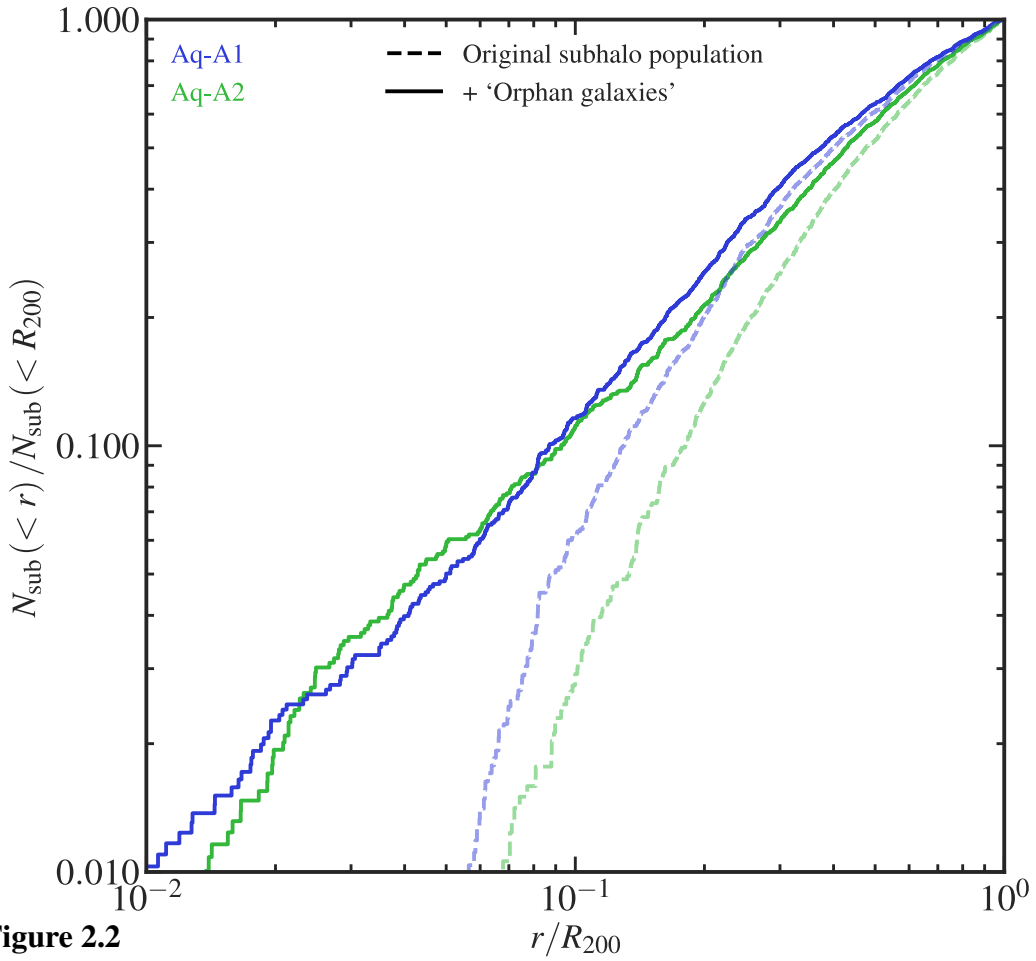


Figure 2.2

Normalized cumulative subhalo number counts for the Aq-A1 and Aq-A2 haloes. The dashed lines show the original, uncorrected number counts prior to the application of GALFORM. The solid lines show the number counts for each halo after adding 'orphan galaxies' to the original population. The subhalo populations before the correction are poorly sampled in the innermost regions, and are not well-converged between the two haloes.

calculated by Sawala et al. (2017, fig. 4, upper panel) using the APOSTLE hydrodynamic simulations that show that up to 40 per cent of satellites in the inner ∼30 kpc can be destroyed, although overall the destroyed fraction is much smaller (see also D'Onghia et al., 2010; Errani et al., 2017; Garrison-Kimmel et al., 2017). For our purposes this difference, which changes the radial subhalo distribution, is fairly important but it has the opposite effect to the omission of orphan galaxies and, as we discuss below, the two effects partially cancel out. To correct for these baryonic effects, we downsample the z=0 Aquarius subhaloes according to the value of the radius-dependent depletion rate derived by Sawala et al. (2017).† The radial dependence of the depletion factor and further details about this procedure are given in Section 2.3.1.1. We refer to this final population, which incorporates 'orphan galaxies' and baryonic effects, as our fiducial tracer population. Unless otherwise stated we use this subhalo population throughout the rest of this chapter.

We apply a selection cut to the fiducial Aquarius subhalo populations on the basis of their $v_{\rm peak}$ values, under the expectation that this will provide a stronger correlation with the likelihood of a galaxy forming within the subhalo (Sawala et al., 2016a) than, for example, selecting by present-day maximum circular velocity or present-day mass (Libeskind et al., 2005; Wang et al., 2013). This correlation has been shown to hold in the ΛCDM model, which is one of the priors in our analysis. In Fig. 2.3 we show the radial number density of subhaloes normalized by the mean subhalo density within R_{200}.‡ This is used to assess the appropriateness of applying a $v_{\rm peak}$ selection, and to determine the $v_{\rm peak}$ value down to which the profiles are consistent. We compare this against the radial distribution of luminous satellites selected from a set of high-resolution hydrodynamic simulations from the APOSTLE Project (Fattahi et al., 2016; Sawala et al., 2016b). This is a suite of 12 cosmological zoom resimulations of Local Group-like regions run with the GADGET3 code and EAGLE subgrid physics models (Crain et al., 2015; Schaye et al., 2015). Of these,

†There is an error in the values of the fitting parameters quoted by Sawala et al. (2017); see Section 2.3.1.1 for further details and the correct values of the parameters.

‡Throughout this and subsequent chapters, R_Δ denotes the radius of the spherical volume enclosing a density equal to Δ times the critical density for closure, $\rho_{\rm crit}(z) = 3H^2(z) \,/\, 8\pi G$; where $H(z)$ is the Hubble parameter, z is the redshift of interest, and G is the gravitational constant. M_Δ is the mass enclosed by this volume. Unless noted otherwise these quantities are evaluated at z=0.

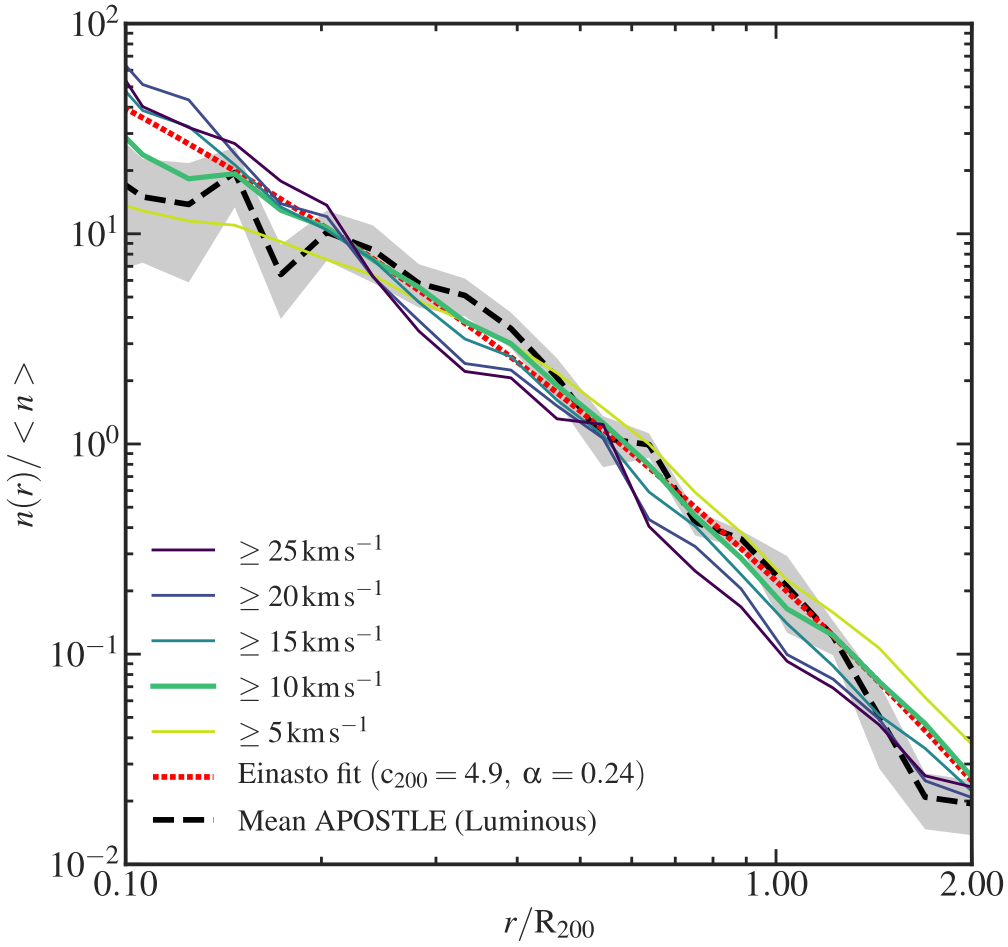


Figure 2.3

The radial number density of fiducial subhaloes normalized to the mean number density within R_{200}. The thin solid lines show the distributions for subhaloes with different $v_{\rm peak}$ cuts averaged over the five Aquarius haloes. The thick dashed line and associated shaded region show the radial distribution of *luminous* satellites and its associated 68 per cent scatter obtained using eight haloes from the APOSTLE high-resolution hydrodynamic simulations. The thick dotted line shows the best-fitting Einasto profile to the fiducial population. For ease of comparison the profile with our chosen selection criterion of $v_{\rm peak} \geq 10$ km s^{-1} is provided as a thick solid line.

four regions—which contain eight MW and M31 analogues—were re-run at much higher resolution and are used here. The APOSTLE data are not used beyond the provision of this reference profile as the simulation is unable to resolve ultrafaint luminous satellites at the magnitudes we are considering here.

Fig. 2.3 shows that the radial profile of subhaloes is largely independent of the value of $v_{\rm peak}$, except for values below 10 km s^{-1}, where resolution effects come into play. Most importantly, we find that the profiles of samples selected with thresholds above this value are in good agreement with the profile of the luminous APOSTLE satellites, and that of observed MW satellites (see Section 2.3.1.3), making this a good choice to model the radial distribution of satellites. We therefore only consider subhaloes with $v_{\rm peak} \geq 10$ km s^{-1} in the rest of our analysis.

2.3.1.1 Baryonic Effects

D'Onghia et al. (2010),Sawala et al. (2017), and Garrison-Kimmel et al. (2017) identify systematic differences in the subhalo radial number density profiles of haloes in DM-only and hydrodynamic simulations. The enhanced tidal stripping by the central baryonic disc leads to a reduction in the number of subhaloes in hydrodynamic simulations compared to their DM-only counterparts. The subhalo depletion is a radially varying function that peaks in the innermost regions of the host halo.

The subhalo number density profiles can be fit using a double power law functional form, which is given in Sawala et al. (2017, equation 2). With help from Till Sawala (private communication) we determined that some of the values stated for the fitting parameters of equation (2) in the published version of the paper are incorrect. Taking the raw data from Till Sawala we made our own fits, binning the data in units of $\chi = r \,/\, R_{200}$. Fig. 2.4 gives the averaged subhalo number density profiles of 4 MW-like haloes from the APOSTLE suite. To improve our statistics we also average over 5 Gyr of cosmic time, similar to Sawala et al.. To these profiles we fit a double power law of the form

$$\rho\,(r) = 2^{(\beta-\gamma)/\alpha}\,\rho_s\,(c\chi)^{-\gamma}\,(1+[c\chi]^{\alpha})^{(\gamma-\beta)/\alpha}\,, \tag{2.3.1}$$

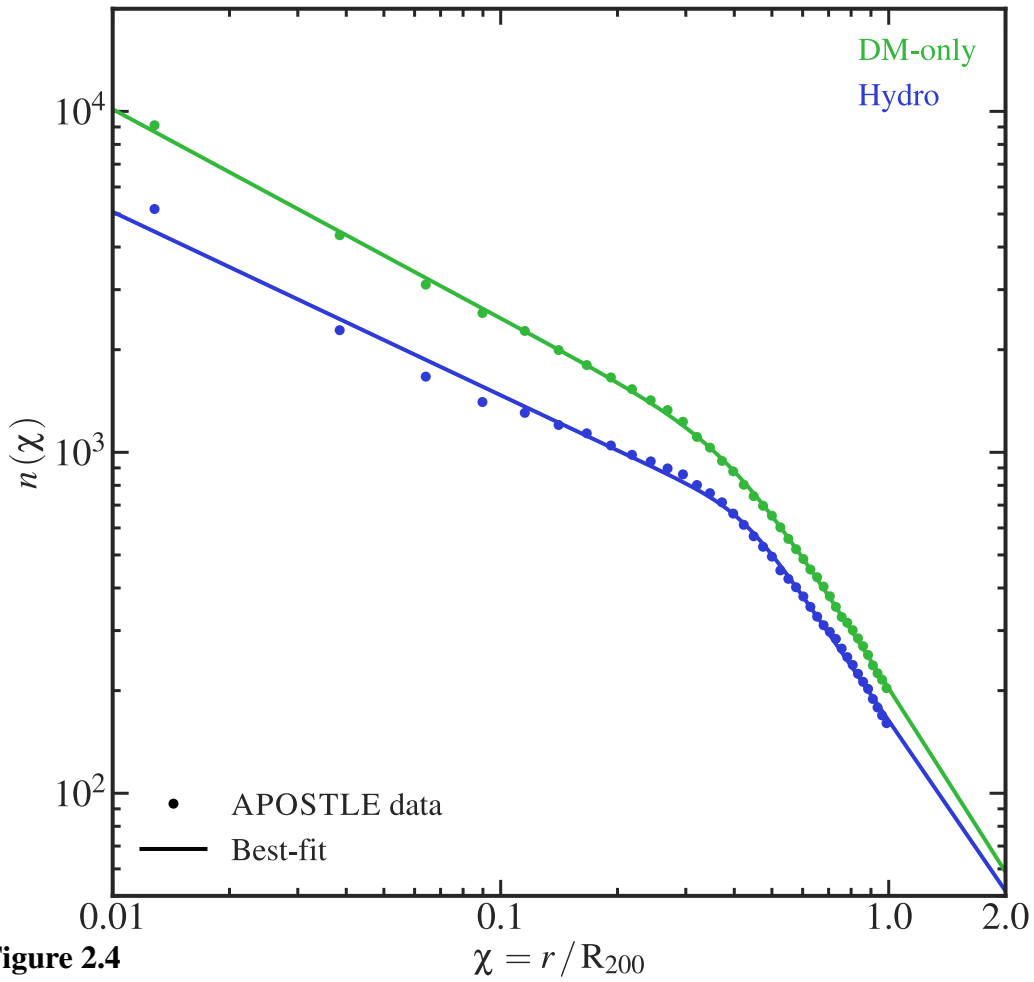


Figure 2.4

Fits to subhalo number density profiles in DM-only and hydrodynamic simulations. The points show averaged radial profiles for four APOSTLE haloes. To obtain better statistics, these points were also averaged over 5 Gyr of cosmic time; see Sawala et al. (2017) for details. The solid lines show the best-fitting double power laws (see main text for the best-fitting parameters).

which gives fitting parameters of

$$(c, \rho_s, \alpha, \beta, \gamma) = (2.50,\ 875,\ 4.41,\ 1.80,\ 0.613)$$

and

$$(c, \rho_s, \alpha, \beta, \gamma) = (2.35,\ 613,\ 8.35,\ 1.66,\ 0.537)$$

for the DM-only and hydrodynamic simulations, respectively.

These fits are only constrained in the radial range $\left[10^{-2}, 1.0\right]\ \chi$ but in practice we extrapolate the profiles over a slightly wider range of $\left[10^{-3}, 2.0\right]\ \chi$ to subsample our haloes. We find that only minimal extrapolation is required to achieve this, and that the ratio in this extended range is also slowly varying.

The subhalo depletion is given by the ratio between the hydrodynamic and DM-only subhalo number density profiles. We compute this using the best-fitting double power law fits given above. The ratio varies from ~0.5 for the inner halo to about ~0.8 at R_{200}. We correct the Aquarius subhalo distributions using this depletion value. For each subhalo, we compute the subhalo depletion value at its radial position and use a Monte Carlo approach to decide if this subhalo is retained or discarded. Only retained subhaloes are used as input to the Bayesian inference method.

2.3.1.2 Rescaling the Aquarius haloes to a fiducial MW halo mass

We would like to assess if the calculation of the total satellite count is sensitive to the mass of the MW halo. This is important given the large uncertainties in current estimates of the MW halo mass, with values typically in the range $(0.5 - 2.0) \times 10^{12}$ $M_\odot$ (e.g. Cautun et al., 2014b; Piffl et al., 2014; Wang et al., 2015). To do this, we rescale the Aquarius haloes to a fiducial MW halo mass, $M_{\rm MW,target}$, and apply our Bayesian method to these rescaled haloes. When expressed as a function of rescaled radial distances, $r\ /\ R_{200}$, the radial number density of subhaloes is largely independent of host mass (Springel et al., 2008; Han et al., 2016; Hellwing et al., 2016). Thus, we can rescale the original Aquarius haloes to different target masses by multiplying the radial distance of each subhalo by

the ratio $R_{200,\,\mathrm{target}}$ / $R_{200,\,\mathrm{original}}$. Unless specified otherwise, the results presented in this chapter and Chapter 3 are calculated for a fiducial MW halo mass, M_{MW}=1.0 × 10^{12} $\mathrm{M}_{\odot}$. The variation of these results with MW halo mass is analysed in Section 3.2.2.

2.3.1.3 Comparison to the MW satellite distribution

A further test of the appropriateness of a particular choice of tracer population can be obtained by comparing its radial distribution with that of the observed MW satellites. When calculating the latter, we need to correct for the radial incompleteness in the surveys: faint satellites can be detected only at small radial distances which, if unaccounted for, leads to a biased, more centrally concentrated satellite distribution. This radial profile, corrected for radial incompleteness, is given by

$$\frac{\mathrm{d}N\,(r)}{\mathrm{d}r} = \frac{\sum_i P_{\mathrm{MW,\,i}}\,\delta\,(r_i - r)}{\sum_i P_{\mathrm{MW,\,i}}\,\epsilon\left(r, M_{\mathrm{V,\,i}}\right)}\,, \tag{2.3.2}$$

where the sum is over all the observed classical, SDSS and DES satellites, r_i and $M_{\mathrm{V,\,i}}$ are the position and absolute magnitude of the i-th satellite, and $\delta\,(r_i - r)$ is the Dirac delta function. The quantity, $P_{\mathrm{MW,\,i}}$, denotes the probability that a satellite is associated with the MW, which we take to be 1 for all objects except the DES satellites. Many of these are likely to have fallen in as satellites of the Large Magellanic Cloud (LMC) and, being at first infall, are still concentrated near the position of the LMC which is adjacent to the relatively small region surveyed by the DES. For these objects we use the probabilities of association given by J16; we discuss this point in greater detail in Section 3.1.1 in the next chapter. The quantity, ϵ, is the detection efficiency (see Section 2.2) at distance, r, for satellites of magnitude, M_V, and accounts for radial incompleteness. The denominator of equation (2.3.2) is maximal for small r values, where all observed satellites have 100 per cent detection efficiency, and decreases at large r.

Fig. 2.5 shows that v_{peak}-selected subhaloes have the same radial distribution as the observed MW satellites, as predicted by theoretical arguments (Libeskind et al., 2005). This comparison demonstrates the validity of our fiducial choice for the radial distribution of satellites. The subhalo distribution given in Fig. 2.5 corresponds to a MW halo mass

of 1.0×10^{12} $M_\odot$ and using a slightly lower value for the MW halo mass leads to an even better agreement between the two radial distributions.

We also used equation (2.3.2) to compute the model-independent radial number density for three different observational subsamples: the classical, SDSS, and DES satellites. We find good agreement between the three subsamples (not shown), indicating that the data are consistent with the radial distribution being independent of satellite brightness. This is consistent with Fig. 2.3, where we find that the radial profile of $v_{\rm peak}$-selected objects is largely independent of the value of $v_{\rm peak}$.

2.3.1.4 A fit to the radial profile of subhaloes

In a later part of our analysis (Section 3.2.3), we will make use of a functional form for the radial profile of satellites in order to scale our results to different MW halo masses or fiducial volumes. For this, we fit an Einasto profile (Einasto, 1965; Navarro et al., 2004)§ to the $v_{\rm peak} \geq 10$ km s^{-1} curve shown in Fig. 2.3. The Einasto profile—or the very similar NFW profile (Navarro et al., 1995, 1996b, 1997)—provides a good description of the radial number density of substructures (Sales et al., 2007; Kuhlen et al., 2008; Springel et al., 2008; Han et al., 2016). We can parametrize the Einasto profile in terms of a shape parameter, α, and the concentration, $c = R_{200} \,/\, r_{-2}$, with r_{-2} the scale radius at which the logarithmic slope of the profile is -2. Using the scaled radial distance, $\chi = r \,/\, R_{200}$, the Einasto profile is given by

$$\frac{n(\chi)}{\langle \mathrm{n} \rangle} = \frac{\alpha c^3}{3 \left(\frac{\alpha}{2}\right)^{\frac{3}{\alpha}} \gamma\left(\frac{3}{\alpha}, \frac{2}{\alpha} c^\alpha\right)} \exp\left[-\frac{2}{\alpha} (c\chi)^\alpha\right] , \tag{2.3.3}$$

where $\langle \mathrm{n} \rangle$ is the mean number density within R_{200} and the lower incomplete Gamma function, γ, is defined as

$$\gamma(s, x) = \int_0^x t^{s-1} \, e^{-t} \, \mathrm{d}t \; . \tag{2.3.4}$$

§A fit to the DM density profile of this form was first introduced in Navarro et al. (2004) but only referred to as the "Einasto profile" in Merritt et al. (2006).

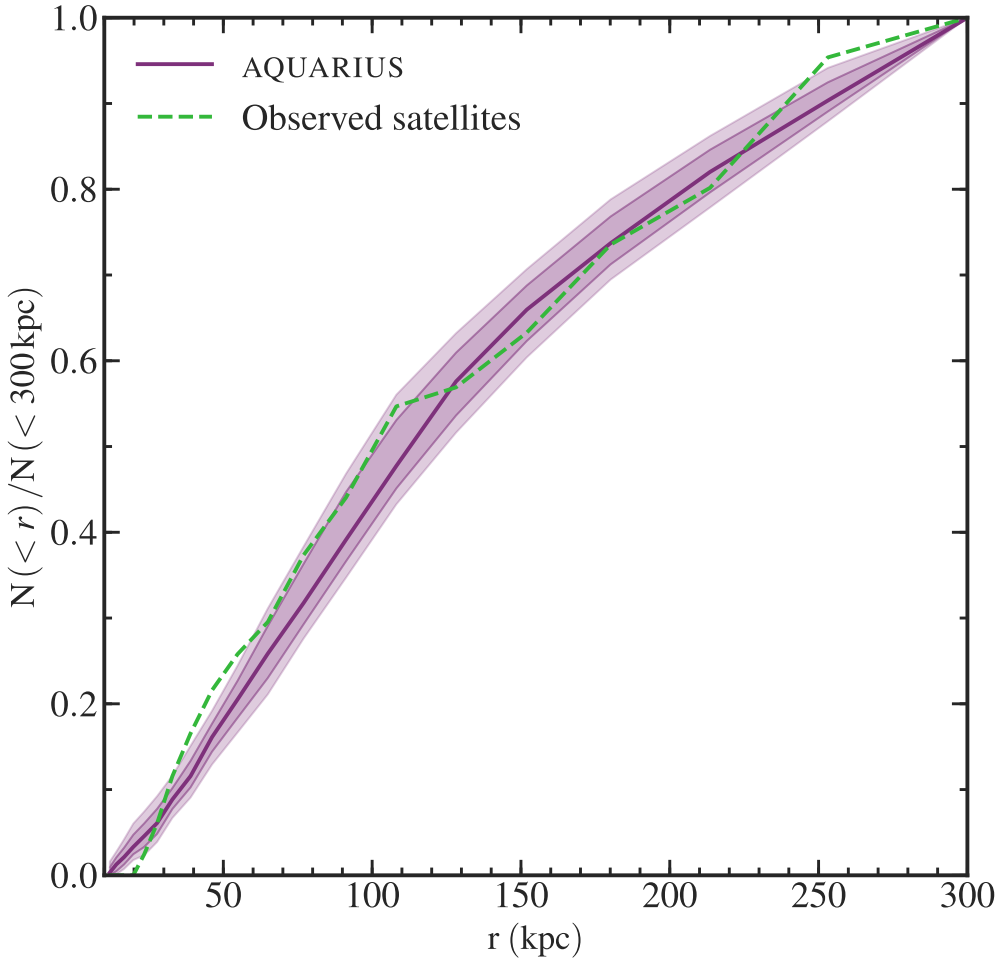


Figure 2.5

Comparison of the radial distribution of observed MW satellites (dashed line) with that of $v_{\rm peak}$-selected subhaloes from the five Aquarius haloes (solid line) rescaled to a host halo mass of 1.0×10^{12} $M_\odot$. The sample of observed satellites was corrected for survey radial incompleteness (see the text) and consists of the classical, SDSS, and DES satellites. We further accounted for the possibility that many of the DES satellites may have fallen in with the LMC by using the probabilities of association with the MW given by J16. The dark and light shaded regions represent the 68 per cent CL and 95 per cent CL (statistical error) bootstrapped error regions for the $v_{\rm peak}$-selected subhalo distribution, respectively.

We find that an Einasto profile with c=4.9 and α=0.24 provides a good match to the radial number density of subhaloes, as may be seen in Fig. 2.3.

2.3.2 The Bayesian inference method

We are interested in calculating the probability distribution function (PDF) of the total number of satellites, $N_{\rm tot}(< {\rm M}_V)$, if a survey with effective volume, $V_{\rm eff}({\rm M}_V)$, has detected $N_{\rm obs}(< {\rm M}_V)$ satellites. Note that both the effective volume and the number of satellites are functions of absolute magnitude; however, for ease of readability, we drop the explicit dependence on ${\rm M}_V$. Within the Bayesian formalism, the posterior probability of having a total of $N_{\rm tot}$ satellites given that we observe $N_{\rm obs}$ objects within a volume, $V_{\rm eff}$, is given by

$$P\left(N_{\rm tot}\,|N_{\rm obs}, V_{\rm eff}\right) = \frac{P\left(N_{\rm obs}\,|N_{\rm tot}, V_{\rm eff}\right) P\left(N_{\rm tot}\right)}{P\left(N_{\rm obs}, V_{\rm eff}\right)}\,, \tag{2.3.5}$$

where $P\left(N_{\rm obs}\,|N_{\rm tot}, V_{\rm eff}\right)$ is the likelihood of having $N_{\rm obs}$ objects within volume $V_{\rm eff}$ if there is a total of $N_{\rm tot}$ satellites. For the prior, $P\left(N_{\rm tot}\right)$, we take a flat distribution; the denominator is a normalization factor. Thus, we have

$$P\left(N_{\rm tot}\,|N_{\rm obs}, V_{\rm eff}\right) \propto P\left(N_{\rm obs}\,|N_{\rm tot}, V_{\rm eff}\right)\,. \tag{2.3.6}$$

The method needs two more ingredients: (i) a prior for the radial distribution of satellites, which we take as that of Aquarius $v_{\rm peak}$-selected subhaloes, and (ii) a sample of observed satellites, which we take as that of the SDSS and DES surveys. Thus, $N_{\rm tot}$ represents the inferred total number of MW satellites given these priors.

In practice, it is computationally prohibitive to evaluate the likelihood function over the full parameter space so we use Approximate Bayesian Computation (ABC). ABC methods approximate the likelihood by selecting model realizations that are consistent with the data. For our study, ABC is an accurate way to estimate the likelihood function because (i) we compare the realizations with the actual data rather than with summary statistics and (ii) our data set consists of a discrete number of satellites and our method selects realizations that exactly reproduce the observations.

The likelihood can be computed using a Monte Carlo method applied to each Aquarius halo. We start by selecting the satellite tracer population—i.e. the DM subhaloes—within our fiducial MW halo radius and organizing them into a randomly ordered list. Then, for each observed satellite, we estimate the required number of satellites of equal brightness such that there is only one such object inside the effective survey volume corresponding to that observed dwarf galaxy. Starting with the brightest observed satellite, we pick random numbers, $N_{\rm rand}$, until we find that only one of the top $N_{\rm rand}$ subhaloes is inside the corresponding effective survey volume. The resulting $N_{\rm rand}$ value corresponds to one possible realization of the total count of objects, $N_{\rm tot}(\mathrm{M}_V)$, of brightness equal to that of the observed satellite. We then remove the top $N_{\rm rand}$ subhaloes and repeat the same procedure for the next brightest observed satellite.

We considered ordering the subhalo list according to their $v_{\rm peak}$ values, which is equivalent to ordering them from brightest to faintest, assuming that $v_{\rm peak}$ is a luminosity indicator. This ordering would have the advantage of capturing correlations between the luminosity of spatially close satellites as would happen in the case of group accretion. For example, a massive satellite at first infall is likely to bring with it other luminous galaxies (Wang et al., 2013; Shao et al., 2016). In practice, we find that the effects of any such correlations are insignificant compared to the uncertainties introduced by host-to-host variability.

This Monte Carlo procedure generates one possible realization of the dependence of the total number of satellites on absolute magnitude, $N_{\rm tot}(< \mathrm{M}_V)$. To sample the full allowed space, the procedure must be repeated many times, for different locations of the survey volume, for different host haloes, and for new randomizations of the subhalo list. The details of how we achieve this are given in Section 2.3.2.1, together with a more computationally efficient implementation of the Monte Carlo algorithm just described.

Our Monte Carlo approach represents a discrete sampling of the effective volume, $V_{\rm eff}$, which is a smooth function of M_V. While in principle this may lead to biases, in practice there are enough observed satellites to sample densely the range of absolute magnitudes of interest; thus, any such effects are small, as may be seen in Section 2.3.2.2.

2.3.2.1 Practical implementation

For each Aquarius halo, we position an observer 8 kpc from the halo centre at one of six vertices of an octahedron, and select a spherical region of 300 kpc in radius centred on this point, similar to T08. All subhaloes within this region are sorted randomly and assigned an index. We then select a conical region with its apex at the observer position and its opening angle corresponding to the sky coverage of the survey from which the observational data are drawn. The maximum radial extent of the conical region, $R_{\rm eff}$, for an observed object of given magnitude is calculated using equation (2.2.3).

Starting with the brightest object in the survey, of magnitude $M_{\rm V,\,1}$, we sequentially select subhaloes from our sorted list until we identify one object within our mock survey volume. This sets the lower bound for $N_{\rm tot}(< M_{\rm V,\,1})$. To set the upper bound, we continue down the sorted list of subhaloes until we find the largest subhalo index which still corresponds to only one subhalo inside the mock survey volume. Every choice between the lower and upper bounds is equally consistent with the observation of one object of $M_{\rm V,\,1}$ within the survey volume; we therefore randomly select one number in this interval and remove this many subhaloes from the beginning of our ordered list. We then consider the next brightest object—of magnitude $M_{\rm V,\,2}$—and repeat the above procedure, using the updated list of subhaloes and the new effective survey volume, $V_{\rm eff}(M_{\rm V,\,2})$. We continue this process down to the faintest observed satellites in the survey.

The procedure is repeated for 1000 pointings evenly distributed across the simulated sky, and for six observer locations, creating 6000 realizations for each simulated halo. There are 5 Aquarius haloes so, in total, we obtain 3×10^4 realizations that are used to estimate the median and 68 per cent, 95 per cent, and 98 per cent uncertainties of the complete satellite luminosity function.

2.3.2.2 Validation

In order to validate the Bayesian inference method, we tested it on a set of 100 mock SDSS observations provided by Marius Cautun. The results of these tests, and a sample of 10 of the mocks, are shown in Fig. 2.6. The mock observations were generated from a 'blinded' luminosity function—indicated in the figure by the thick dotted line—and were obtained from the Aq-A1 halo distribution of subhaloes with $v_{\rm peak} \geq 10\ {\rm km\,s^{-1}}$ within 300 kpc. The selected subhaloes were then randomly assigned absolute magnitudes according to the input luminosity function. Mock observations were produced for 100 random pointings of a conical region analogous to the SDSS volume within the halo, taking into account the effective radius out to which satellites of different magnitudes could be identified. To model better the observations, mocks were generated using a radially dependent detection efficiency: for a given magnitude, using equation (2.2.3), we calculated $R_{\rm eff}$, which is the radius corresponding to a 50 per cent detection efficiency, and then assumed that the detection efficiency decreases from 1 to 0 linearly in the radial range $[0.5, 1.5]\,R_{\rm eff}$. Satellites found in regions where the detection efficiency is below unity were included in the mocks using a probabilistic approach by comparing a random number between 0 and 1 with the value of the detection efficiency. The luminosity functions for a sample of 10 of the 100 resulting mocks are shown as thin solid lines in Fig. 2.6. Even though all the mocks survey the same halo, we find a large spread in the number of observed satellites.

Taking each mock survey data set in turn, we apply the Bayesian inference method, producing 100 estimates of the total satellite luminosity function, 10 of which are shown in Fig. 2.6 as thick solid lines. To assess the method fully, we also illustrate the 68 per cent uncertainty region, taken from one of the mocks and shifted so that the centre of the region is aligned with the 'true' luminosity function. Most of the inferred satellite luminosity functions lie inside the 68 per cent uncertainty region, in line with statistical expectations, thus demonstrating the success of the method at reproducing the underlying true luminosity function. This uncertainty region, taken from one mock, is comparable to the 68 per cent confidence region obtained from the medians of all 100 mocks, which

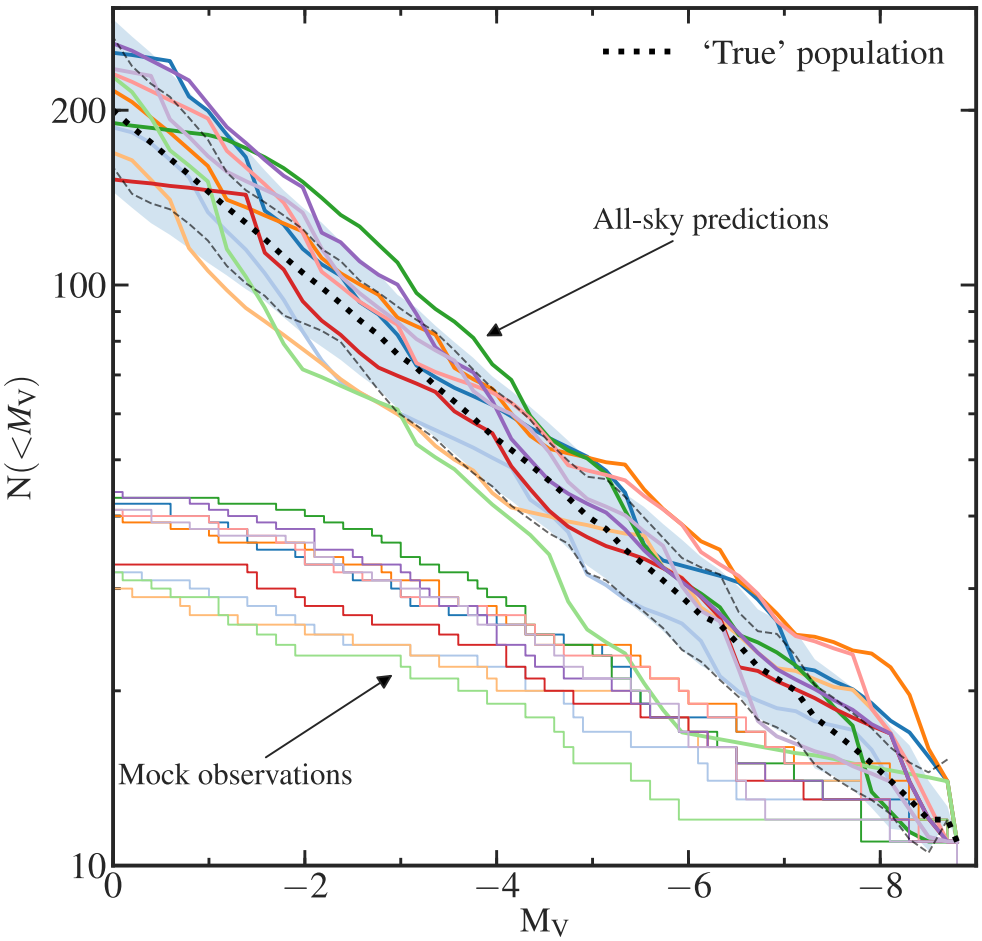


Figure 2.6

Tests of the Bayesian inference method using mock observations. The thick dotted line shows the input luminosity function used to create 100 SDSS mock observations. The luminosity functions of a sample of 10 of these are shown as thin solid lines. Each of the 10 mock observations was used, in turn, to predict a cumulative satellite luminosity function. The results are shown as thick solid lines. The shaded region represents the 68 per cent uncertainty from one of the mock predictions, shifted to lie on top of the input luminosity function. The dashed lines bound the 68 per cent confidence region over the medians of all 100 mock predictions.

further demonstrates that the method successfully estimates uncertainties. Note also that our inference method assumes that the detection efficiency is a step function at $R_{\rm eff}$, but the mocks were generated using a radially varying detection efficiency. Thus, this test also shows that assuming an effective detection radius is a good approximation and does not bias the inferred total luminosity function.

2.3.3 Comparison to previous inference methods

As we discussed briefly in Section 2.1, the previous method used for inferring the total satellite count has some drawbacks. The T08 method, which was also employed by Hargis et al. (2014), used a similar $v_{\rm peak}$-selected radial distribution of subhaloes as us (although not accounting for unresolved subhaloes or baryonic effects). However, the differences arise from the way in which these distributions are used. The T08 method employs a completeness volume, $V_{\rm comp}$, that is typically selected as the volume where the detection efficiency, $\epsilon(\mathrm{M}_V)$, has a given non-zero threshold value, e.g. $\epsilon(\mathrm{M}_V) = 0.9$. Note that the T08 completeness volume can be different from the effective volume used in our Bayesian method. To obtain an unbiased estimate, only observed satellites within that completeness volume, i.e. satellites with detection efficiencies above the threshold value, should be used for inferring the total satellite count. The T08 approach is based on calculating, for each observed satellite, the fraction of $v_{\rm peak}$-selected subhaloes inside the completeness survey volume associated with that satellite. This fraction, $\eta{=}N_{\rm sub}(< V_{\rm comp})/N_{\rm max\ sub}$, is the ratio of the number of subhaloes, $N_{\rm sub}(< V_{\rm comp})$, inside $V_{\rm comp}$ to the total number of subhaloes, $N_{\rm max\ sub}$, inside the halo. Then, for the i-th observed satellite, the fiducial halo volume contains

$$\frac{1}{\eta_i \epsilon_i} \tag{2.3.7}$$

satellites of absolute magnitude, $\mathrm{M_{V,\,i}}$, with ϵ_i the detection efficiency associated to the i-th observed satellite.

Fig. 2.7 shows a comparison of the T08 approach, discussed above, with our Bayesian inference approach. These methods were applied to the same SDSS DR9 data set using

the W09 completeness function (see Table 2.3) and the subhalo distribution of a single simulated halo, Aq-A1, corrected for 'orphan galaxies' and baryonic effects. Here, when applying the T08 method, we choose a completeness radius corresponding to $\epsilon(M_V) = 0.5$, which is equal to the effective radius used by the Bayesian method, and only use observed satellites with detection efficiencies, $\epsilon \geq 0.5$. All the satellites detected by the W09 algorithm have $\epsilon > 0.5$ and thus pass this selection criterion. The median estimates produced by the T08 and Bayesian methods are similar. However, as we show in extensive tests detailed in Section 2.3.3.1, where we apply the T08 approach to mock observations similar to those in Fig. 2.6, the T08 method underestimates the uncertainties.

2.3.3.1 Testing previous methods

Here, we test the T08 method by applying it to a set of mock satellite observations. This is similar to the exercise in Section 2.3.2.2, where, using the same blind mock observations, we demonstrated that the Bayesian approach introduced in this paper successfully infers the input 'true' luminosity function used to generate the mock observations.

A set of 100 mock SDSS observations was generated from a 'true' population by Marius Cautun (see Section 2.3.2.2 for a description of the mocks) and supplied to us, and we applied the T08 method. In order to return an unbiased estimate, we applied the T08 approach using a completeness radius that corresponds to a detection efficiency, $\epsilon = 0.5$, and used as input only those observed satellites with detection efficiencies, $\epsilon \geq 0.5$. Using a random sample of 10 mock observations, we compare in Fig. 2.8 the scatter among the various mocks with the typical error of the T08 method. We find that the typical 68 per cent (statistical) uncertainty range estimated by the T08 method is too low: for most magnitude values, most of the 10 mocks are outside the 68 per cent (statistical) confidence interval. This was also demonstrated in Fig. 2.7 and arises because the T08 method does not incorporate the effects of stochasticity into its estimation of the uncertainties.

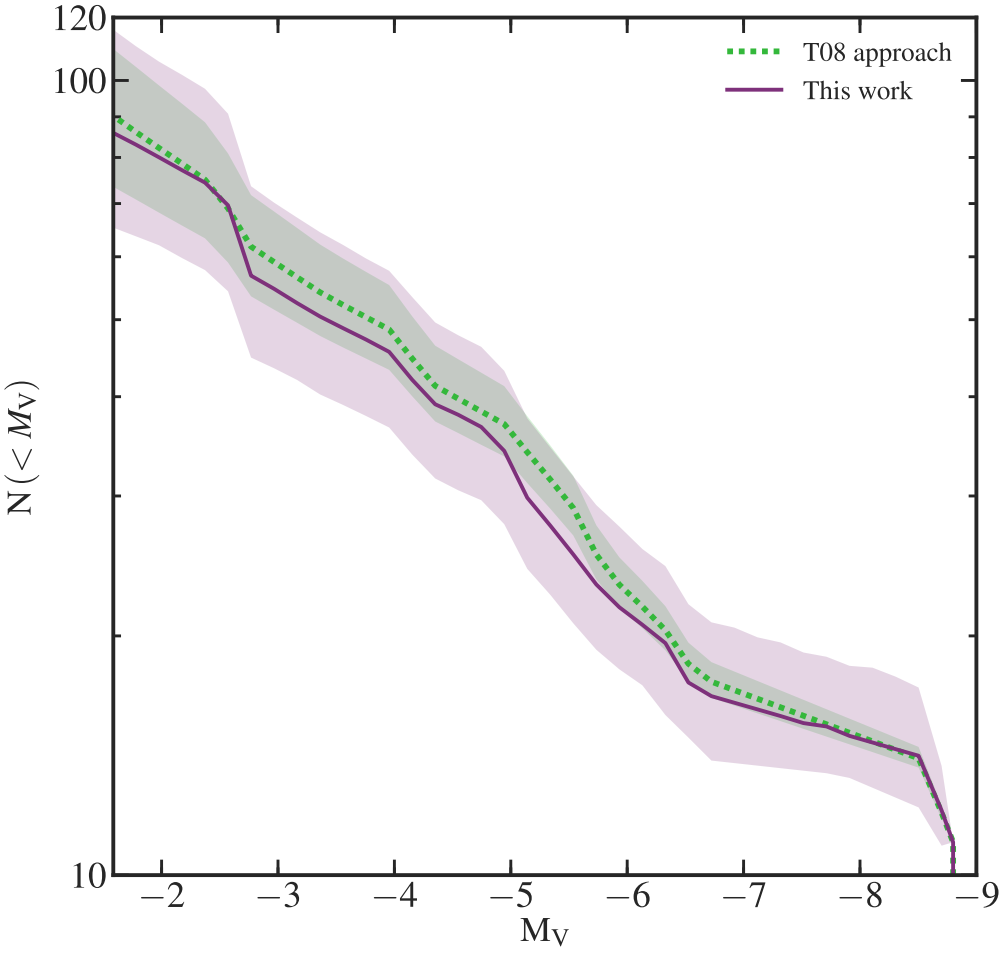


Figure 2.7

Comparison of two different inference methods for the total dwarf galaxy luminosity function: the T08 method and the Bayesian approach introduced here. Both methods were applied to the same data set, the SDSS. The median estimate (solid line) and associated 68 per cent uncertainties (shaded regions) for each method are shown. The T08 method does not account for stochastic effects, so it underpredicts the uncertainties.

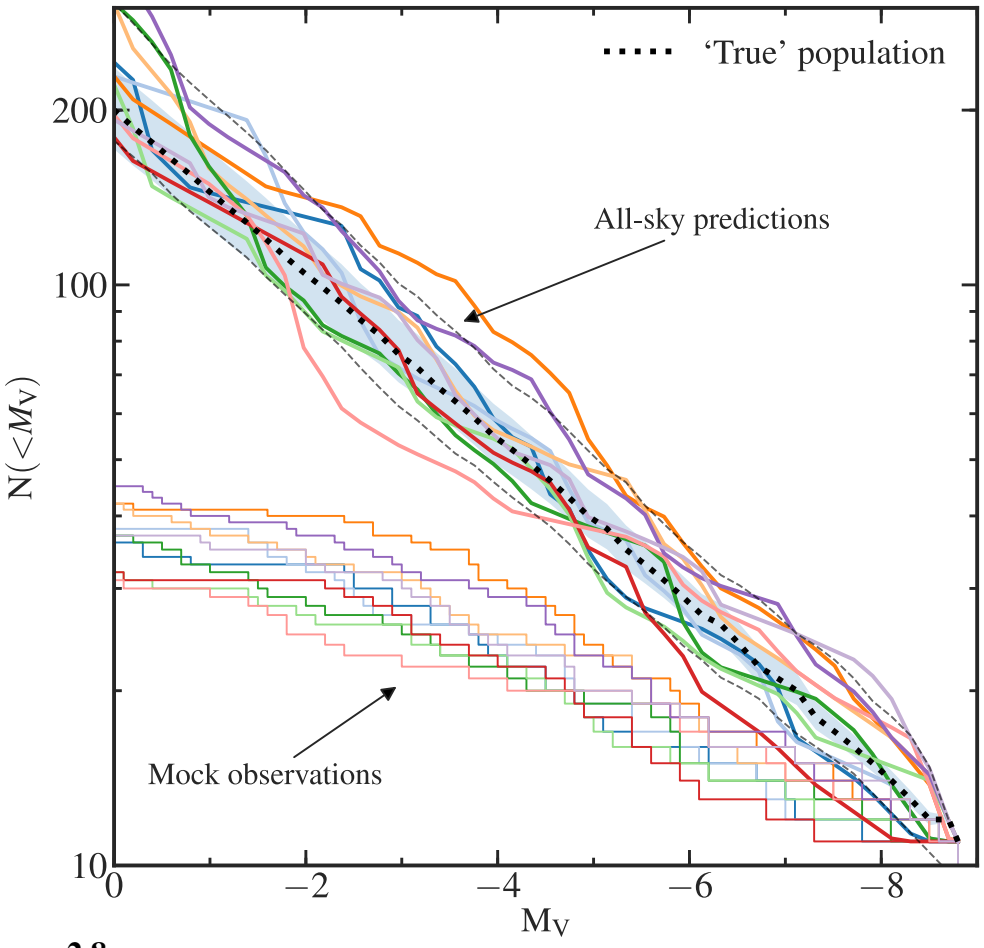


Figure 2.8

Test of the T08 method using mock observations. The thick dotted line shows the input luminosity function used to create the 10 SDSS mock observations, whose luminosity functions are shown as thin solid lines. Each of the mock observations was used, in turn, to predict a cumulative satellite luminosity function, with the corresponding results shown as thick solid lines. The shaded region represents the 68 per cent (statistical) uncertainty from one of the mocks, shifted to lie on top of the input luminosity function. The dashed lines bound the 68 per cent (statistical) confidence region over the medians of all 100 mock predictions.

2.3.3.2 Sources of uncertainty

There are two main factors that introduce uncertainties. First, the distribution of satellites is not isotropic but flattened. As a result, surveying different regions of the halo can introduce variations in the number of observed objects. Secondly, the presence or absence of satellites in the observed volume is a stochastic process. Given N satellites and the probability, η, of a satellite being inside the survey volume, then the number of observed satellites in the survey is a binomial distribution with parameters N and η. To determine which of the two effects is dominant, we applied the Bayesian inference method to the original subhalo distribution of the Aq-A1 halo and to many isotropized versions of it. These were generated keeping the same radial distances and isotropizing the angular coordinates by randomly sampling from uniform distributions of $1 - \cos\theta$ and ϕ in spherical coordinates. The results of this test, presented in Fig. 2.9, show that while anisotropy makes a noticeable contribution to the uncertainty at faint magnitudes, stochastic effects are the dominant source of uncertainty.

The T08 method accounts for anisotropy, but it does not account for stochastic effects, which leads to an underestimation of the errors. This underestimate is clearly seen in the mock observation tests detailed in Section 2.3.3.1, where we find that most of the T08 estimates lie further than the 68 per cent uncertainty interval from the input 'true' luminosity function. Given the probability, η, that a satellite is inside the volume $V_{\rm eff}$, the T08 method predicts η^{-1} satellites within the halo—see equation (2.3.7) without the ϵ term. While this is true on average, for any realization the number of satellites in the halo is given by a negative-binomial distribution with mean value η^{-1}. The width of this distribution, which characterizes the size of the stochastic effects, gives rise to an additional uncertainty that is not included in the T08 methodology.

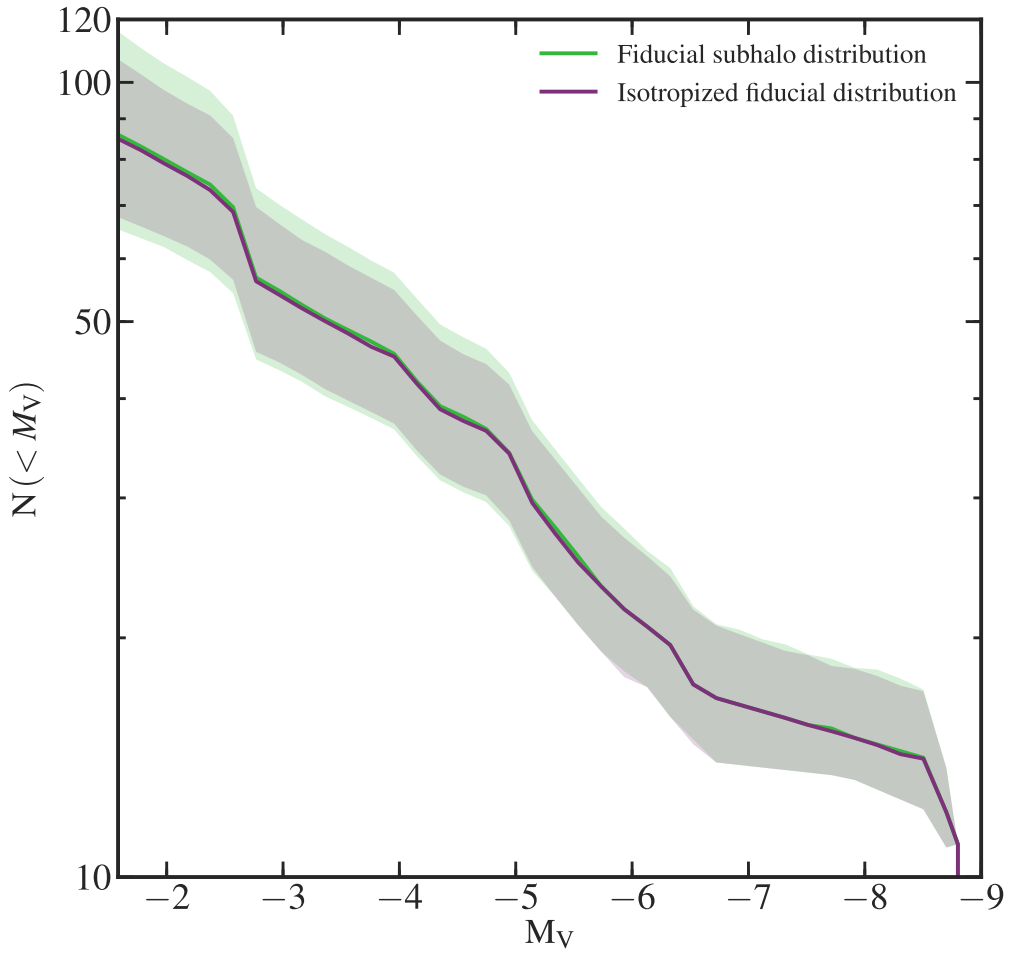


Figure 2.9

Comparison of the dominant sources of uncertainty in estimates of the total satellite luminosity function: the flattening of the subhalo distribution or the stochastic effects. The region labelled 'fiducial subhalo distribution' corresponds to applying our method to the fiducial subhalo population of the simulated halo, Aq-A1. This estimate is affected by both the shape of the tracer distribution and stochastic effects. The region labelled 'isotropized fiducial distribution' assumes the same radial distribution of subhaloes but with isotropized angular coordinates; this is affected only by stochastic effects. Both approaches have approximately the same median (solid line) and 68 per cent scatter (shaded region). Thus, stochastic effects are a major source of uncertainty.

2.4 Summary

In this chapter we developed a new method to infer several properties of the total satellite galaxy population of the MW from incomplete observations of these objects. Our approach depends upon two key ingredients: (i) observations of the MW satellite galaxy population using surveys in which the completeness is characterized well, and; (ii) a prior for the radial distribution of the satellite galaxy population. For the former, our observational data comprise catalogues of MW satellite galaxies identified by the SDSS and DES, which we assume are representative of the full satellite population (see Section 2.2). Our prior for the radial distribution is $v_{\rm peak}$-selected populations of subhaloes from high resolution DM-only haloes simulated in the Aquarius Project, which we find produce a good match to observations (see Section 2.3.1).

We described the Bayesian approach in Section 2.3.2 and carried out several tests with mock observational data to establish the reliability of the method. These verified that it successfully reproduces the total satellite galaxy population from partial observations, and models more completely the associated uncertainties in the estimate compared with previous approaches. These uncertainties are not affected strongly by anisotropy in the spatial distribution of the tracer population, and we find that stochastic effects are the dominant source of uncertainty.

In Chapter 3, we introduce the results of our analysis applied to the satellite galaxies of the MW. We consider the estimates obtained when using the SDSS and DES observations separately and together, and assess the effect of the mass of the MW halo—which is not constrained well—on the result. We set this in context of anticipated future developments in observational capability that will enable the satellite population to be surveyed to unprecedented depth and predict what discoveries these improvements will deliver.

CHAPTER 3

The total satellite population of the Milky Way

In the previous chapter, we introduced a Bayesian approach to infer the total population of MW satellite galaxies from observational surveys using DM-only numerical simulations as priors. We now provide the results of our analysis using the Aquarius haloes rescaled to a fiducial MW halo mass of 1.0×10^{12} $M_{\odot}$ and within a fiducial radius, $R_{\rm out}$=300 kpc. In Section 3.1.1, we perform our analysis for the SDSS and DES data separately. Each estimate requires extrapolations over large unobserved volumes which affects the luminosity function that is inferred. The surveys are also complete to different depths: the DES can probe fainter objects than the SDSS and consequently can estimate the satellite luminosity function to fainter magnitudes. We combine both surveys in the same analysis in Section 3.1.2, increasing the volume of the MW halo that is probed and enabling us to infer the total luminosity function across a larger range of magnitudes.

Our approach depends on several factors which can affect the luminosity function that is inferred. We consider these in Section 3.2, focusing on our choice of tracer population and the mass of the MW halo. We also consider how the luminosity function scales for different choices of the outer radius of interest. In Section 3.3, we look to forthcoming observational campaigns that will probe a larger volume of the MW halo at fainter magnitudes than extant surveys and make a prediction for what these will observe. We discuss the implications of

our results and consider some of the limitations of our method in Section 3.4. We present concluding remarks for both Chapter 2 and this chapter in Section 3.5.

3.1 Estimates of the luminosity function

3.1.1 Separate estimates from SDSS and DES

The results of applying our Bayesian inference method to the SDSS DR9 data set are displayed in the left-hand panel of Fig. 3.1. Also plotted here is the luminosity function of all satellite galaxies observed in the SDSS DR9 survey for which absolute magnitude measurements have been published to date; these data are provided in Table 2.1. We adopt the response functions of the two search algorithms detailed in Section 2.2, by K08 and W09. The counts inferred using the K08 function are systematically higher than those obtained using the W09 function at absolute magnitudes fainter than $M_V \approx -5.5$. This is expected and is a consequence of both algorithms detecting the same number of satellites, but the W09 algorithm probing deeper at fainter magnitudes. The larger scatter in the K08 estimate reflects the additional uncertainty introduced by requiring an extrapolation over larger volumes of the halo. In the remainder of this chapter we will use the results obtained using the W09 algorithm as it is able to detect—at least in principle—fainter objects.

Down to magnitude $M_V = -2.7$ (corresponding to the faintest satellite considered by Tollerud et al.), the SDSS data imply that there are at least 64^{+55}_{-26} (98 per cent CL, statistical error—note that the 68 per cent CL is shown in the figure) dwarf galaxies within a radial distance of 300 kpc. This is significantly lower than the estimate by Tollerud et al., who inferred 322^{+144}_{-76} at 98 per cent CL. The Tollerud et al. estimate is higher for two reasons. First, they adopted the K08 response function which is shallower than the W09 function. Secondly, their estimates were based on the SDSS DR5 data release that observed 10 satellites over a footprint of ~8000 square degrees. Since then, while SDSS DR9 has added an additional ~6500 square degrees of sky coverage, it has detected only four new satellites brighter than $M_V = -2.7$. Consequently, the number of observed satellites per

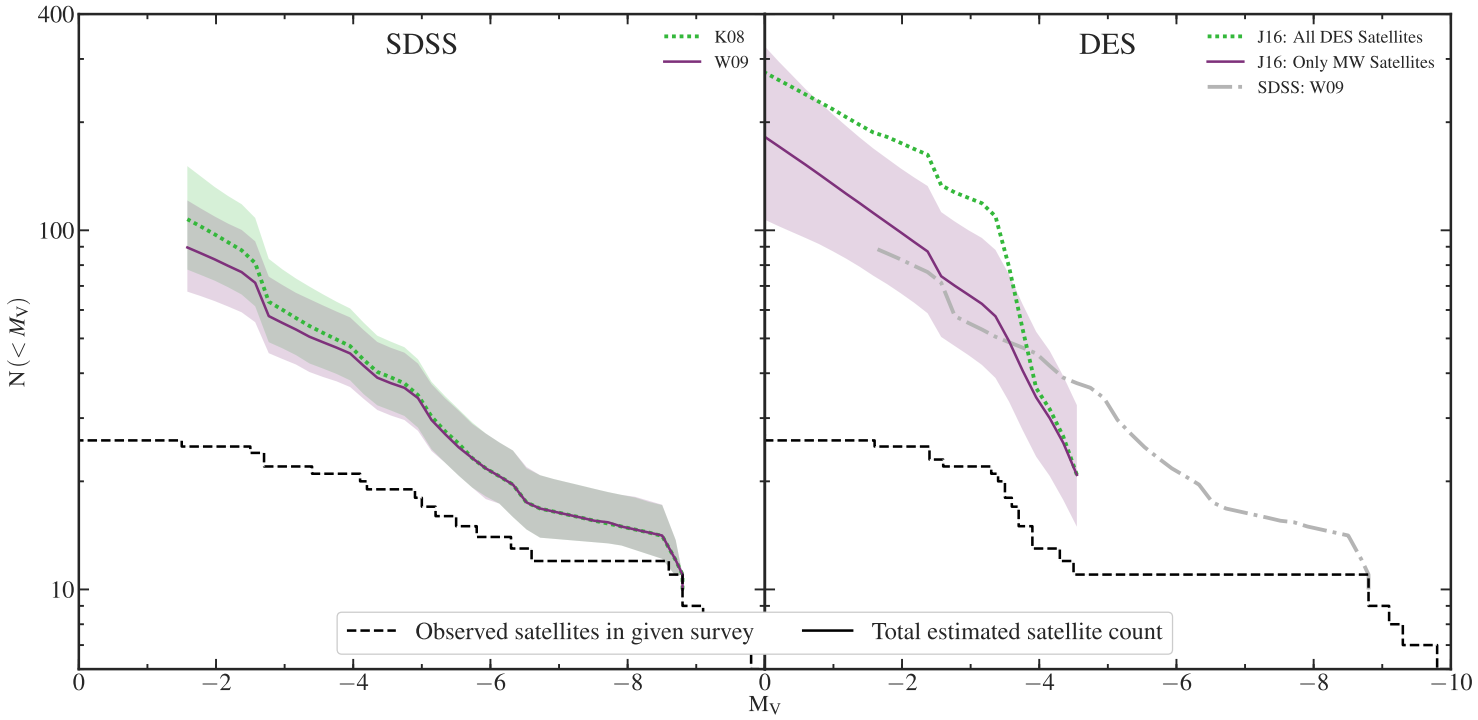


Figure 3.1

The total MW satellite galaxy luminosity functions inferred from the SDSS and DES surveys (left and right panels, respectively). The solid lines and corresponding shaded regions show the median estimates and associated 68 per cent uncertainties. The dashed lines indicate the number of observed satellites within 300 kpc in each of the two surveys; these are input into the Bayesian inference method. For the SDSS, we show estimates using the response functions of the two search algorithms devised by Koposov et al. (2008, K08) and Walsh et al. (2009, W09). Both algorithms detect the same number of satellites, but the latter probes down to fainter magnitudes. For DES, we use the Jethwa et al. (2016, J16) response function. This result is truncated at $M_V \leq -4.5$ as no satellites brighter than this have been observed in DES within 300 kpc. The DES estimate (solid line) accounts for the possibility that some objects observed by DES may be satellites of the LMC. For reference, we also plot a second estimate which assumes that all DES objects are associated with the MW (dotted line), as well as the SDSS W09 result (dot-dashed line).

unit volume has decreased, and the size of the satellite population inferred from this is lower as a result. If future surveys discover few new satellites, this estimate of the total population could prove to be overstated. This highlights the importance of combining multiple surveys which probe different parts of the sky to sufficient depth.

The result of applying our method to the DES is shown in the right-hand panel of Fig. 3.1; in this case we adopt the J16 response function. No satellites are detected in DES with magnitude in the range $-8.9 \lesssim M_V \lesssim -4.5$, so we interpolate between the values calculated at each end of the range. Including all the DES satellites in the inference method returns twice as many satellites with $M_V \lesssim -4$ than inferred from the SDSS satellites alone. This discrepancy is caused by the DES footprint being adjacent to the two Magellanic Clouds which, models suggest, are on their first infall (Kallivayalil et al., 2013; J16). If that were the case, then it is likely that the two Magellanic Clouds would have contributed their own complement of satellite galaxies. These are not distributed uniformly over the sky, but are still clustered around the Magellanic Clouds (Sales et al., 2011). As many as half of the satellites detected by DES could have come from the LMC (Sales et al., 2007; J16). Failing to account for these localized associations would lead to an overestimate of the total Galactic satellite population. We adopt the probabilities of association of each of the DES objects with the LMC inferred by J16 and include an additional step in our analysis: for each mock survey pointing, we generate a Monte Carlo realization in which the DES satellites are assigned either to the MW or to the LMC according to these probabilities. Only the DES satellites assigned to the MW are then included in the Bayesian inference.

The right-hand panel of Fig. 3.1 shows the satellite luminosity function accounting for the association of some DES satellites to the LMC. This estimate is in good agreement with the estimate from the SDSS for $M_V \lesssim -4$. The discrepancy at brighter magnitudes is due to the lack of detection in the DES survey of any satellites brighter than $M_V = -4.5$ within a distance of 300 kpc. While DES is deeper than SDSS, it covers a smaller area on the sky and thus, for $M_V \lesssim -5$ and $M_V \gtrsim -0.5$, DES samples a smaller effective volume than SDSS (see Fig. 2.1). Nonetheless, the luminosity function inferred from DES is generally consistent with that inferred from SDSS, given the large uncertainties in both estimates.

3.1.2 Combined estimate from SDSS+DES

The best estimate of the total satellite luminosity function is obtained by combining the SDSS and DES. We modify the analysis described in Section 2.3.2.1 by including a second conical region oriented relative to the first one such that it reproduces the approximate orientation of the real SDSS and DES. The SDSS vector is used to define the pointing 'direction' of this configuration; it uniformly samples the sky as before. The second vector—corresponding to the DES—is fixed at an angle of 120° relative to the SDSS vector but is allowed to rotate around it. For each SDSS pointing a configuration is generated and a combined SDSS+DES luminosity function is calculated. In practice, this analysis corresponds to that of a survey of effective volume, $V_{\mathrm{eff,\,SDSS}} + V_{\mathrm{eff,\,DES}}$, consisting of two disjoint regions. The analysis otherwise proceeds as before.

The predicted total satellite luminosity function from the combined SDSS+DES data is shown in Fig. 3.2. This estimate is consistent with those from the separate analyses of SDSS and DES data: except in a few bins, the medians of the individual estimates lie within the 68 per cent uncertainty range of the SDSS+DES estimate. When comparing with the combined result, we find that the SDSS-only estimate overpredicts the satellite count for $\mathrm{M}_V \leq -4$, which is to be expected given that DES did not find any satellites brighter than $\mathrm{M}_V = -4.5$ within our fiducial radius of 300 kpc. In contrast, for $\mathrm{M}_V > -4$, the SDSS-only estimate occasionally lies slightly below the total satellite count, reflecting the large number of satellites with $\mathrm{M}_V \geq -4.5$ observed by DES. The data associated with Fig. 3.2 are provided in Table 3.1.

We find that the total satellite luminosity function is well-fitted by the broken power law:

$$\log_{10} N(<\mathrm{M}_V) = \begin{cases} 0.095\mathrm{M}_V + 1.85 & \text{for } \mathrm{M}_V < -5.9 \\ 0.156\mathrm{M}_V + 2.21 & \text{for } \mathrm{M}_V \geq -5.9 \end{cases}, \tag{3.1.1}$$

that is, the faint end of the luminosity function is described by a significantly steeper power law than the bright end.

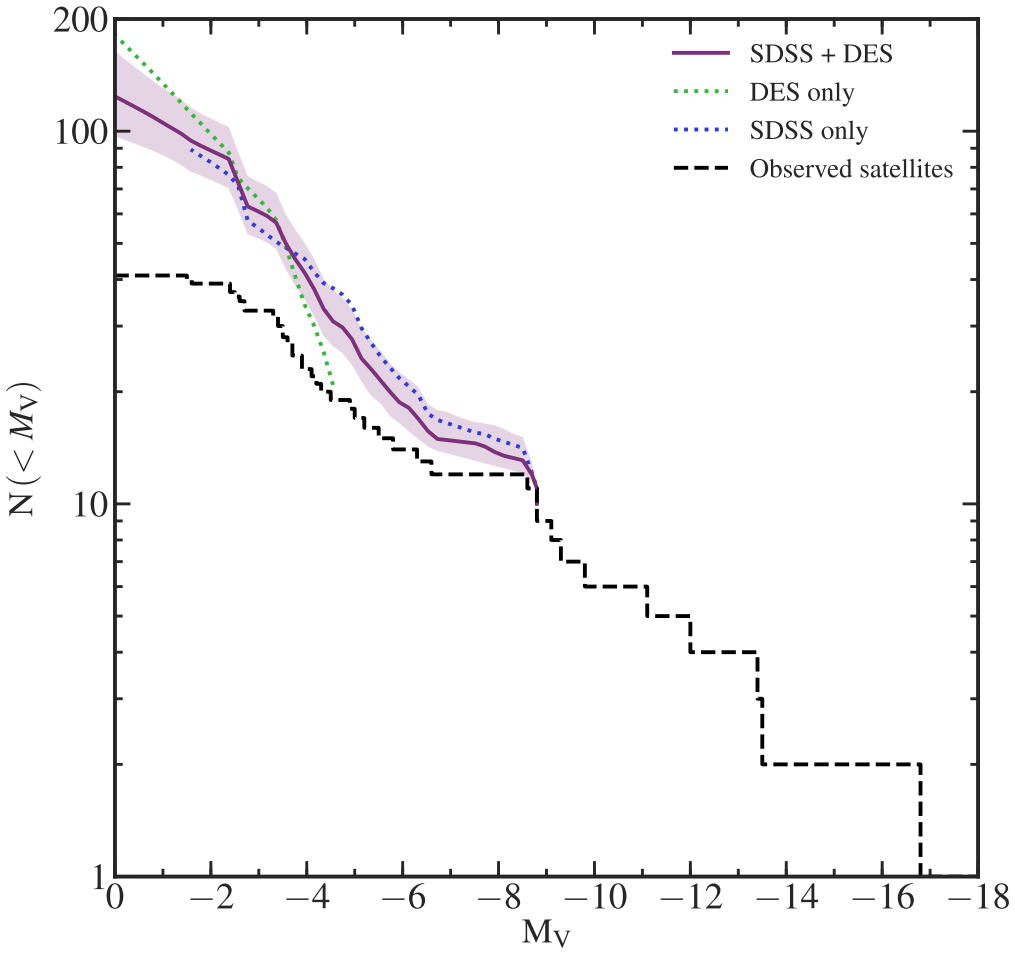


Figure 3.2

The total luminosity function of dwarf galaxies within a radius of 300 kpc from the Sun obtained from combining the SDSS and DES data. The solid line and the shaded region show the median estimate and its 68 per cent uncertainty, respectively. The two dotted lines show the median satellite luminosity functions using SDSS and DES data separately. The luminosity function of all observed satellites within the SDSS and DES footprints inside 300 kpc is indicated by the dashed line. The total satellite luminosity function is well-fitted (not shown here) by the broken power law given in equation (3.1.1).

Table 3.1

Cumulative number of satellites as a function of absolute magnitude within a heliocentric distance of 300 kpc for a 1.0×10^{12} $M_\odot$ MW halo, inferred from a Bayesian analysis of the SDSS DR9 + DES observed satellites. The cumulative number of these observed satellites is provided for reference. The quoted confidence limits are for statistical errors only.

M_V	$N(< M_V)$		Confidence limits: lower – upper		
	Observed	Predicted	68%	95%	98%
−8.8	11	11	...	...	...
−8.5	12	13	12 – 15	12 – 19	12 – 21
−8.0	12	14	13 – 16	12 – 20	12 – 21
−7.5	12	15	13 – 17	13 – 21	13 – 22
−7.0	12	15	14 – 17	13 – 21	13 – 23
−6.5	13	16	14 – 19	13 – 23	13 – 25
−6.0	14	19	16 – 22	15 – 27	15 – 30
−5.5	16	22	19 – 26	17 – 32	16 – 34
−5.0	18	27	23 – 32	20 – 39	20 – 43
−4.5	20	31	27 – 38	23 – 47	22 – 50
−4.0	23	41	35 – 49	30 – 60	29 – 64
−3.5	30	52	44 – 62	39 – 76	37 – 82
−3.0	33	61	51 – 73	44 – 89	43 – 95
−2.5	37	77	64 – 93	55 – 114	52 – 123
−2.0	39	89	74 – 108	63 – 133	60 – 142
−1.5	41	96	79 – 118	67 – 147	63 – 158
−1.0	41	105	86 – 131	72 – 163	68 – 175
−0.5	41	115	92 – 146	75 – 186	71 – 203
0.0	42	124	97 – 164	78 – 225	73 – 249

3.2 Factors influencing the luminosity function

3.2.1 Dependence on the tracer population

In Section 2.3.1 we argued that in order to make accurate predictions, it is necessary to incorporate two effects into the analysis: the inclusion of unresolved subhaloes, i.e. 'orphan galaxies', and the depletion of subhaloes due to tidal disruption by the central galaxy disc (i.e. baryonic effects). These changes primarily involve the inner ∼50 kpc of the halo, the region to which the faint end of the luminosity function is most sensitive. Although these two effects have opposite sign, they do not cancel out completely. In Fig. 3.3 we show the effect of each of the two corrections, which are only important for the faintest satellites ($M_V > -2$). Prior to any correction, the M_V=0 satellite count is 141^{+54}_{-35}; the addition of unresolved subhaloes reduces this to 113^{+34}_{-24}. This is because the unresolved subhalo population is very centrally concentrated; on average some ∼85 per cent of them lie within 50 kpc. Accounting for subhalo depletion due to baryonic effects produces a small upward shift in the median to 124^{+40}_{-27}; a decrease of ∼12 per cent relative to the uncorrected luminosity function inferred using the L2 subhalo distribution of Aquarius haloes.

3.2.2 Dependence on the mass of the MW halo

As we discussed in Section 2.3.1.2, the MW halo mass is poorly constrained, with recent estimates varying within a factor of 2 from our fiducial choice of $M_{\rm MW} = 1.0 \times 10^{12}$ $M_\odot$ (see the compilation of Wang et al., 2015). To investigate the sensitivity of the inferred total satellite luminosity function to the MW halo mass, we repeated our analysis for two extreme mass values, 0.5×10^{12} $M_\odot$ and 2.0×10^{12} $M_\odot$, corresponding roughly to lower and upper bounds for the MW halo mass (e.g. Wang et al., 2015). To obtain estimates for these halo masses, we rescaled the fiducial radial distribution of subhaloes using the procedure described in Section 2.3.1.2. The inferred dwarf galaxy luminosity functions

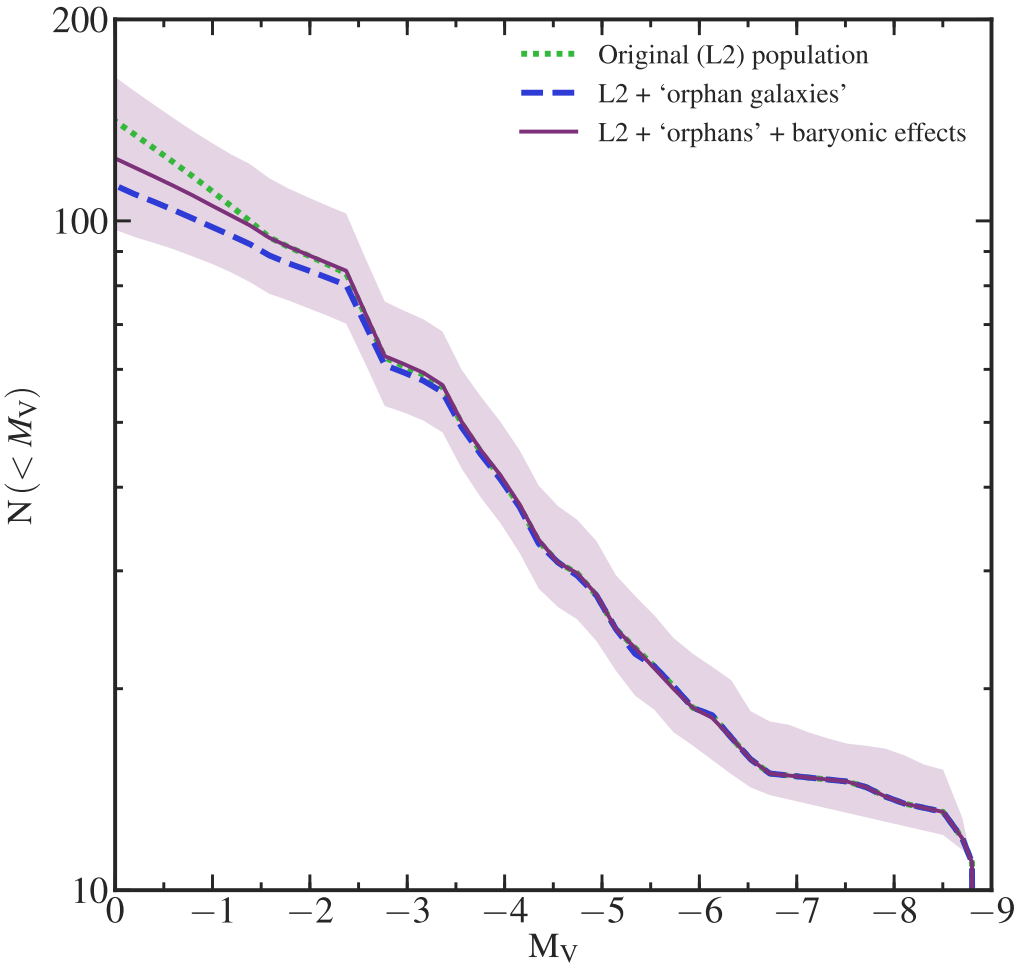


Figure 3.3

The sensitivity of the inferred satellite luminosity function to the two corrections applied to the subhalo population. The dotted line shows the inferred satellite count using the original subhalo distribution of Aquarius. The dashed line shows the effect of adding subhaloes missing due to resolution effects, the so-called 'orphan galaxies'. The solid line shows the results from our analysis, in which we also account for subhalo depletion due to baryonic effects. The shaded region indicates the 68 per cent uncertainty region of our final result.

are displayed in Fig. 3.4, which shows that despite the factor of 4 difference between the lowest and highest halo masses considered, no large discrepancies begin to emerge until $M_V \geq -2.5$. Even at fainter magnitudes, the differences are well within the 68 per cent uncertainty range for a given MW halo mass.

The number of subhaloes in a DM halo scales strongly with halo mass (e.g. Wang et al., 2012; Cautun et al., 2014a), so naively we might assume that the inferred satellite count follows the same relation. As Fig. 3.4 demonstrates, that is not the case; we see only a weak variation of $N_{\rm tot}$ with $M_{\rm halo}$. The inferred satellite count depends only on the shape of the normalized radial profile of subhaloes, and not on the *total* number of subhaloes. When expressed in terms of r / R_{200}, i.e. radial distance in units of the virial radius of the halo, the radial profile is largely independent of host mass (Springel et al., 2008; Han et al., 2016; Hellwing et al., 2016). Different host masses correspond to different values of R_{200}, and thus any features in the radial profile are mapped on to different physical radial distances. If the radial distribution of subhaloes were a power law, then the inferred satellite count would be independent of halo mass: for fixed r, changing R_{200} would only lead to a shift in the normalization of the radial profile, which is unimportant for our analysis.

3.2.3 Dependence on the outer radius cut-off

Fig. 3.5 illustrates the dependence of the total satellite count within a given radius, r, as a function of r. These estimates follow from the observation that the radial number density of subhaloes selected above a $v_{\rm peak}$ threshold is independent of the value of the threshold (see Fig. 2.3), which suggests that the radial distribution of satellites should also be independent of satellite luminosity.

The fiducial radial distribution of subhaloes is well described by an Einasto profile: the number of satellites within $\chi = r$ / R_{200} is given by:

$$N(<\chi) = 4\pi \int_0^{\chi} n(\chi')\, \chi'^2 \, \mathrm{d}\chi' \,, \tag{3.2.1}$$

with $n(\chi')$ the Einasto profile given by equation (2.3.3). Performing the integration and

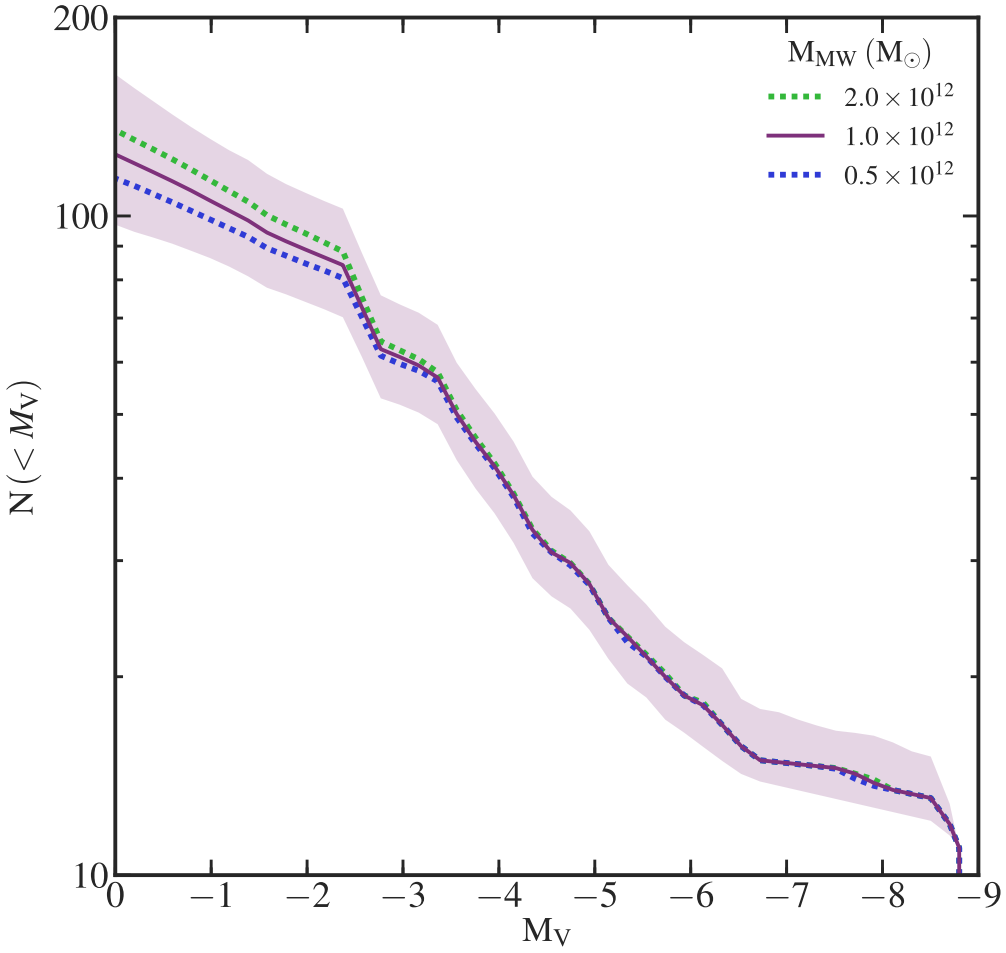


Figure 3.4

The dependence of the inferred total dwarf galaxy luminosity function within 300 kpc on the assumed mass of the MW halo. The lines show estimates for our fiducial MW halo mass of 1.0×10^{12} $M_\odot$ (used in previous plots) and for lighter and heavier MW haloes, as indicated in the legend. For the fiducial case, we show the median estimate (solid line) and the 68 per cent uncertainty (shaded region). For the other two cases we show only the median estimates (dotted lines).

substituting for χ gives:

$$N(<r) = N(<300\ \mathrm{kpc})\ \frac{\gamma\left(\frac{3}{\alpha}, \frac{2}{\alpha}\left[c\chi\right]^{\alpha}\right)}{\gamma\left(\frac{3}{\alpha}, \frac{2}{\alpha}\left[c\frac{300\ \mathrm{kpc}}{R_{200}}\right]^{\alpha}\right)}, \tag{3.2.2}$$

where the function γ is given by equation (2.3.4). The radial dependence of $N(<r)$ is affected by the assumed value for the MW halo mass through the dependence of R_{200} on halo mass. Fig. 3.5 shows the radial dependence of $N(<r)$ for the three MW halo masses assumed in Fig. 3.4; we find only a mild variation with MW halo mass. Extending to distances farther than 300 kpc leads only to modest increases in the satellite count, with an ~20 per cent increase at 400 kpc, which is roughly half way between the MW and M31. Of all the satellites within 300 kpc, ~80 per cent of them lie within 200 kpc, the R_{200} value for a 1.0×10^{12} $\mathrm{M}_{\odot}$ halo mass. At even smaller radial distances, we find ~45 per cent of the satellites within 100 kpc.

3.3 Apparent magnitude luminosity function

In this section we examine the prospects for discovery of faint satellites in future surveys of the MW. For simplicity we assume that the only factor that determines the detectability of a satellite is its apparent luminosity, rather than its size or surface brightness. We can then calculate the number counts of satellites as a function of V-band magnitude. To estimate apparent magnitudes, we assign an absolute magnitude, M_V, to subhaloes by sampling the inferred luminosity function from Section 3.1.2, i.e. the combined SDSS+DES estimate. We then use the subhalo distance from the halo centre to compute the distance modulus and thus the apparent magnitude. This process is repeated for the luminosity functions generated from each pointing and observer location combination—6000 in all. The results presented in this section are for a MW halo mass of 1.0×10^{12} $\mathrm{M}_{\odot}$ and for a 300 kpc outer radius.

Dwarf galaxy counts as a function of apparent magnitude are shown in Fig. 3.6, where we split the population into two classes: ultrafaint and hyperfaint dwarf galaxies, which we

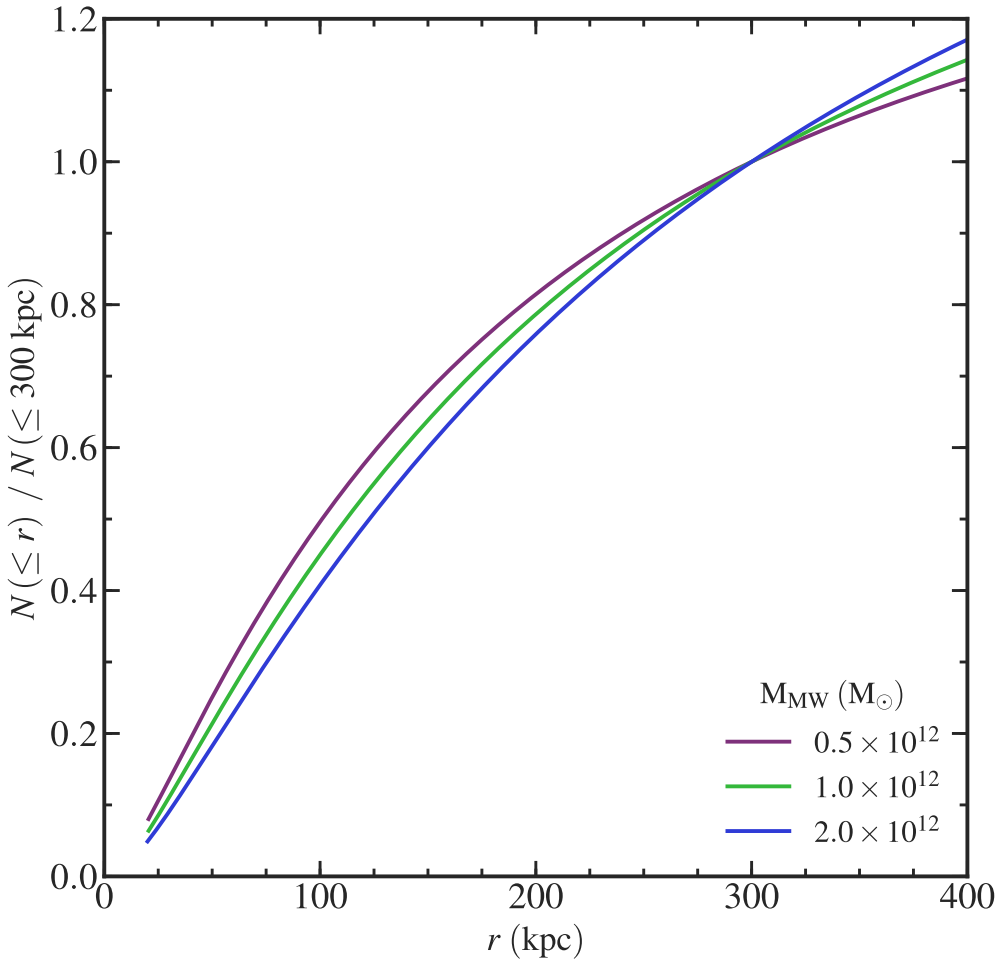


Figure 3.5

The radial dependence of the total number of satellites enclosed within radius r. The Y-axis gives the ratio of this number relative to the satellite count within 300 kpc, the fiducial radius used in this analysis. The result is independent of absolute magnitude, M_V, since subhaloes with different v_{peak} cuts have the same radial profile. There is little dependence on the mass of the MW halo.

define as objects in the absolute magnitude ranges: $-8 < M_V \leq -3$ and $-3 < M_V \leq 0$ respectively. Within 300 kpc from the MW, we expect to find 46^{+12}_{-8} (68 per cent CL, statistical error) ultrafaint and 61^{+37}_{-23} (68 per cent CL, statistical) hyperfaint dwarfs. The first number can be compared to the slightly higher estimate of 66^{+9}_{-7} (68 per cent CL) ultrafaints provided by Hargis et al. (2014), based solely on data from SDSS DR8. We showed in Fig. 3.2 that this population is usually overestimated in predictions based only on SDSS because of a higher abundance of ultrafaint satellites in the SDSS field than would be expected from the total observed population. As discussed in Section 2.3.3, their uncertainties are also 28 per cent too small as stochastic effects were not accounted for in their estimate. Most ultrafaints have apparent magnitudes brighter than 18, so surveys just 0.5 magnitudes deeper than DES—which can detect satellites down to $m_V = 17.5$—should be deep enough to observe most ultrafaint dwarfs in the MW. The luminosity function of hyperfaint dwarfs extends much fainter, with most satellites having $m_V < 21.5$. Discovering these would require a survey 4 mag deeper than DES; the Large Synoptic Survey Telescope (LSST) is one such future survey. An all-sky DES-like survey would only lead to the detection of ∼30 hyperfaint dwarfs, a factor of 4 more than the currently known population.

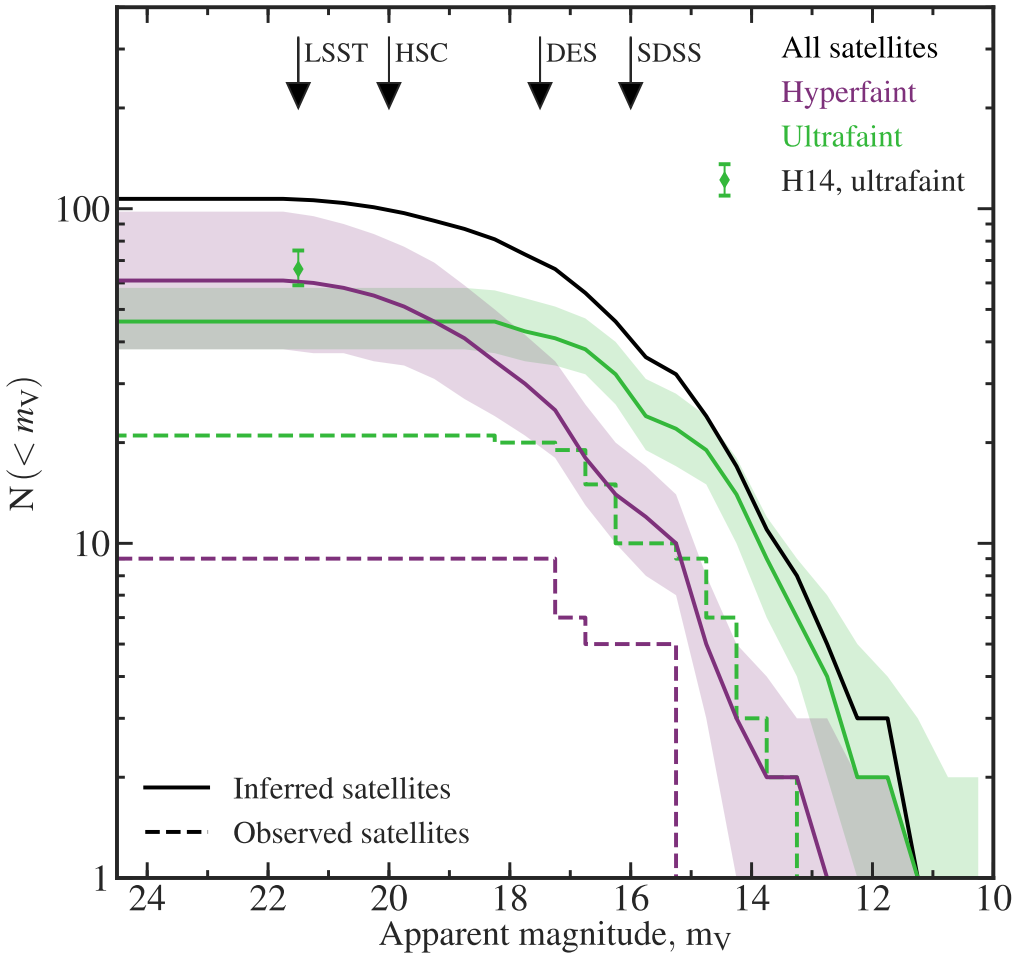


Figure 3.6

The inferred Galactic satellite number counts within 300 kpc as a function of apparent V-band magnitude, m_V. The satellites are split into ultra- and hyperfaint dwarf galaxies, which correspond to objects with absolute magnitude in the range $-8 < M_V \leq -3$ and $-3 < M_V \leq 0$, respectively. The solid lines display the median prediction, with the corresponding shaded regions indicating the 68 per cent uncertainties. For reference the sum of the median predictions of both populations is also provided (black line). The diamond and associated error bars represent the Hargis et al. (2014, H14) prediction and 68 per cent uncertainty region for the total expected number of ultrafaint satellites. As before, the dashed lines display number counts of observed ultra- and hyperfaint dwarf galaxies within the SDSS and DES. The vertical arrows indicate the faintest satellites that can be detected in past and future surveys: SDSS ($m_V = 16.0$), DES ($m_V = 17.5$), HSC ($m_V = 20.0$) and LSST ($m_V = 21.5$).

3.4 Discussion

We have made new predictions for the total MW satellite luminosity function by extrapolating the numbers of satellites currently known using a new Bayesian inference method. As input data we use a combination of the recently discovered satellites in the DES and the population previously known from SDSS DR9. As a prior for the radial distribution of the MW satellites, which is needed for the extrapolation, we use the radial distribution of subhaloes in the Aquarius simulations of galactic haloes having peak maximum circular velocity, $v_{\rm peak}$, above a given threshold. We correct the subhalo distribution for unresolved subhaloes and account for subhalo depletion due to tidal disruption by the central disc. We showed in Fig. 2.5 that the radial distribution of $v_{\rm peak}$-selected subhaloes provides a good match to that of the observed MW satellites. We improve upon previous studies by introducing a new Bayesian inference method, which overcomes the limitations of earlier approaches. We also explore the effect of uncertainties in the MW halo mass and derive a relation for rescaling our estimates to different radii.

We find that, for a 1.0×10^{12} $M_\odot$ MW halo, there are 124^{+40}_{-27} (68 per cent CL, statistical error) satellites brighter than M_V=0 within 300 kpc of the Sun, which is slightly inconsistent with the result from Hargis et al. (2014). Our estimate is consistent with that of J16 when adjusted for differing outer radii; their estimate lies at the upper end of our 68 per cent uncertainty range. Our lower estimate is due to the inclusion of orphan galaxies and baryonic effects, which decrease the inferred count of MW satellites (see Fig. 3.3). Compared with the T08 estimate of 322^{+144}_{-76} (98 per cent CL) satellites brighter than $M_V = -2.7$ within 300 kpc, our estimate of 66^{+39}_{-20} (98 per cent CL, statistical) is a factor of ~5 lower. The origin of this discrepancy is primarily the use by Tollerud et al. of the shallower K08 response function as opposed to the W09 function that we use here. Furthermore, since their work the SDSS survey footprint has increased in size by ~80 per cent, while the number of discovered satellites inside this footprint has increased by very little. We also note that previous studies have underestimated their uncertainty ranges because they have not properly accounted for stochastic effects, which are broadly

independent of satellite brightness (see Section 2.3.3 for a more in-depth discussion).

The future detection of dwarfs depends on their apparent magnitude and we can estimate the luminosity thresholds that future surveys will need to exceed in order to detect the satellite population inferred in this study. In our total inferred population there are 46^{+12}_{-8} (68 per cent CL, statistical) ultrafaint dwarf galaxies (with magnitudes in the range $-8 < M_V \leq -3$), of which ∼20 have been observed so far. We find that the majority of these have apparent magnitudes brighter than m_V=18; these would be discoverable with surveys just 0.5 magnitudes deeper than DES. There are ∼30 such dwarfs still to be discovered in the MW, of which ∼7 should lie inside the SDSS DR9 footprint but beyond its detection limit. Our 61^{+37}_{-23} (68 per cent CL, statistical) hyperfaint dwarfs (with magnitudes $M_V \geq -3$) make up some 62 per cent of our total population and have apparent magnitudes brighter than m_V=21; discovering these would require a survey 4 mag deeper than DES. The planned LSST survey should cover approximately half of the sky and will therefore be able to find half of the inferred count of 61^{+37}_{-23} hyperfaint dwarfs. The sizes of both populations are slightly inconsistent with the lower end of estimates by Hargis et al. (2014).

Our inferred satellite galaxy luminosity function likely represents a lower limit to the true population. Our method takes the observed satellites, which are found in surveys with various detectability limits, as a sample of the global population. In particular, the observed surface brightness cut-off suggests that there could be a population of faint, spatially extended dwarfs that are inaccessible to current surveys (e.g. see Torrealba et al., 2016a). To account for this in our method would require deeper observations than are currently available.

A further complication arises from the presence of the LMC, which, given its large mass, is likely to have brought its own complement of satellites. The LMC may be on its first infall (Sales et al., 2011; Kallivayalil et al., 2013; J16) and the spatial distribution of the satellites it brought with it could be very anisotropic (J16). While we accounted for the probability that a large fraction of DES detections may be associated with the LMC, our analysis does not account for the presence of LMC satellites outside the DES footprint.

To do so would require a prior on the present-day spatial distribution of LMC satellites. Before infall, the LMC could have had perhaps as much as a third of the MW satellite count (J16), though this estimate is very uncertain due to poor constraints on the MW and especially the LMC halo mass. At face value, this could add at most ∼50 satellites to the total count.

Inherent to all analyses that estimate the satellite luminosity function are several systematics which, with a few exceptions, mainly affect the faint end of the luminosity function. The most important of these is the assumed radial distribution of subhaloes, which needs to be determined from cosmological simulations. We showed that the distribution of $v_{\rm peak}$-selected subhaloes matches both the luminosity-independent radial distribution of observed MW satellites and that of state-of-the-art hydrodynamic simulations such as APOSTLE (see Figs 2.3 and 2.5); consequently, we think that any systematic effect on the inferred satellite count arising from our choice of fiducial tracer population is likely to be small. To obtain our fiducial subhalo sample, we needed to correct for two effects that are not well understood. Even the highest resolution simulations, such as those of the Aquarius Project, can suffer from resolution effects, particularly near the centre of the host halo. This issue is common to *all* cosmological simulations, and we addressed it by including 'orphan galaxies' (i.e. galaxies whose haloes have been disrupted) identified by applying the Durham semi-analytic model of galaxy formation, GALFORM, to the Aquarius simulations. This effect is only significant for the faint end of the satellite luminosity function ($M_V \gtrsim -3$) since ∼85 per cent of the orphan population lies within 50 kpc of the centre, the region to which the faint end is most sensitive. We also accounted for baryonic effects on the subhalo mass function by lowering its amplitude in accordance with the prescription in Section 2.3.1.1, using depletion factors based on the APOSTLE Project (Sawala et al., 2017). Garrison-Kimmel et al. (2017) argued for a larger depletion in the inner ∼30 kpc than Sawala et al., while Errani et al. (2017) claim that, due to their limited resolution, most simulations overpredict the subhalo depletion factor. As discussed in Section 3.2.1, although this correction introduces noticeable changes in the predicted satellite luminosity function, these lie within our error bounds, and are smaller

in magnitude than those introduced by the addition of orphan galaxies. These changes primarily affect the faint end of the satellite luminosity function above $M_V \geq -2$, which is also the most theoretically and observationally uncertain part of the luminosity function independently of these effects.

A second important systematic is the choice of observed satellite population. In this work we used satellites discovered in the SDSS and DES. Although all satellites in the former have been spectroscopically confirmed as DM-dominated dwarf galaxies, over three-quarters of the DES satellites have not (yet). We choose to use *all* DES satellites in our analysis. This is motivated by considering the size-magnitude plane (e.g. Drlica-Wagner et al., 2015, fig. 4) that shows that most DES satellites are more consistent with the properties of Local Group galaxies than with the population of known globular clusters. Reclassifying some of the DES detections as globular clusters would lower the inferred total satellite count at the faint end of the luminosity function ($M_V \geq -4$), but would not affect the bright end. Given the good agreement between the SDSS-only and DES-only estimates of the total satellite count, we predict that most DES detections are dwarf galaxies.

The mass of the MW halo is poorly constrained. However, the inferred satellite luminosity function is largely independent of the host halo mass, except at magnitudes fainter than $M_V = -3$ where it shows a very weak mass dependence (see Fig. 3.4). Instead of marginalizing over the MW halo mass distribution, we provide a means of converting between halo masses at the extremes of the range of constraints.

The MW is the smaller partner of a paired system, which could introduce anisotropies into the MW's substructure due to interactions with M31; these would be manifest in the form of more correlated structure. Our choice of 300 kpc for our fiducial radius is less than the midpoint of the MW-M31 distance, minimizing any effects from interactions with M31 and allowing us to model the MW approximately as an isolated halo. In addition, this value is often used in the literature (e.g. Hargis et al., 2014; J16) and is close to the expected virial radius of the MW halo. Our choice of fiducial radius should not be interpreted as precluding the eventual discovery of other satellites further out than this.

The dependence of the total satellite count on MW halo mass is not determined by the

number of subhaloes at fixed mass, but by the shape of the normalized subhalo radial number density profile. A weak halo mass dependence arises from the non-power law nature of the subhalo radial profile: features in this profile are remapped to different physical distances for different halo masses, resulting in a variation in the predicted luminosity function. As a direct consequence, this implies that changes in the assumed MW halo mass, which determines the number of DM substructures, alter the abundance matching relation for Galactic dwarfs; in this regime not all subhaloes of a given mass host a visible galaxy (Sawala et al., 2015). We find that doubling the halo mass roughly doubles the number of subhaloes (Wang et al., 2012; Cautun et al., 2014a), so that there are more of them at fixed $v_{\rm peak}$. A more massive MW halo would then require the same dwarfs to be placed in subhaloes with higher $v_{\rm peak}$ than they would for a lower MW-mass halo.

The spatial distribution of subhaloes—upon which our predictions rely—is partly determined by cosmology but is also affected by the internal dynamics of haloes. In turn, these are influenced by the mass function of subhaloes and their accretion rate, both of which are fairly universal in both ΛCDM and WDM models (Springel et al., 2008; Ludlow et al., 2016). Recent work by Bose et al. (2017) has shown that the radial distribution of subhaloes is broadly independent of the nature of the DM. Our predictions are therefore applicable to other DM models and can, in fact, be used to constrain the masses of WDM particles.

3.5 Conclusions

An estimate of the MW's complement of satellite galaxies is required until deeper, more complete surveys that could discover more faint galaxies are undertaken in the next few years. These predictions can be used to address numerous outstanding astrophysical questions, from understanding the effects of reionization on low mass haloes, to constraining the properties of dark matter particles.

In this work we have, for the first time, combined data from SDSS and DES—which together cover nearly half of the sky—to infer the MW's full complement of satellite

galaxies. Our method requires a prior for the radial distribution of satellites, which we obtain from the subhalo populations of the Aquarius suite of high-resolution DM-only simulations in which we account for the competing effects of resolution and subhalo depletion due to interaction with the central baryonic disc (see Section 3.4). We have shown that selecting subhaloes by their peak maximum circular velocity provides a good match to the radial distribution of observed MW satellites (see Fig. 2.5).

The Bayesian method we have introduced to make these estimates overcomes some of the limitations of previous analyses (see Fig. 2.7), and properly accounts for stochastic effects. For each observed dwarf galaxy, the method estimates how many objects are needed to find one such satellite in the survey volume. These results are averaged over multiple DM haloes to characterize uncertainties arising from halo-to-halo variation.

Within 300 kpc of the Sun—and assuming a MW halo mass of 1.0×10^{12} $M_{\odot}$—we predict that the MW has 124^{+40}_{-27} (68 per cent CL, statistical error) satellites brighter than M_V=0 (see Fig. 3.2). Of these, we expect to find 46^{+12}_{-8} (68 per cent CL, statistical) ultrafaint dwarf galaxies ($-8 < M_V \leq -3$), a result that is marginally inconsistent with the lower end of the Hargis et al. (2014) estimate, but nearly a factor of 5 smaller than the T08 estimate. All the Galactic ultrafaints could be detected by a survey just 0.5 magnitudes deeper than DES. We also expect to find a population of 61^{+37}_{-23} (68 per cent CL, statistical) hyperfaint dwarfs ($-3 < M_V \leq 0$), and to obtain a full census of this population would need a survey 4 mag deeper than DES. The LSST survey should be able to see at least half of this faint population of dwarf galaxies in the next decade.

In all methods seeking to estimate the total luminosity function certain assumptions must be made. In particular, an important assumption is the radial distribution of the true satellite population, which is best inferred from a cosmological simulation. Here, we have used a set of the highest resolution DM-only simulations available and, most importantly, a method for selecting the subhaloes that are expected to host satellites that has been shown to give consistent results for a number of observed properties of the MW satellite population, such as the radial distribution of and counts of bright observed MW satellites. This does not guarantee that the extrapolation is free of systematic effects but as Fig. 2.5

shows, in the regime where we can check with available data, any such systematics are small.

The estimates above represent only lower limits to the total number of Galactic satellites (see Section 3.4) because they do not take into account very low surface brightness objects that may have been missed in current observations. In addition, the estimate does not account for some of the satellites brought in by the LMC which today lie outside the DES footprint. Assuming that 12 of the DES satellites originate from the LMC (i.e. that have $p_{LMC} > 0.5$ in Table 2.1), and that the LMC contributes up to ~50 satellites to the MW (see J16), LMC satellites outside the DES footprint could increase the total count by at most 30 per cent.

While our key results assume a MW halo mass of 1.0×10^{12} $M_{\odot}$, our analysis shows that the predicted dwarf galaxy luminosity function is independent of host halo mass for objects brighter than $M_V = -3$ (see Fig. 3.4). For fainter satellites we find a weak dependence on halo mass, with a more massive MW halo playing host to more satellites. Our tests assuming extreme MW halo mass values ($[0.5, 2.0] \times 10^{12}$ $M_{\odot}$) reveal that the resulting luminosity functions lie well within the 68 per cent uncertainty range calculated for our fiducial MW halo mass. Of the dwarfs within our fiducial distance of 300 kpc, ~45 per cent and ~80 per cent are found within 100 and 200 kpc, respectively.

The results of this study provide a useful reference point for comparing theoretical predictions with the measured abundance of satellite galaxies in the MW. However, it must be borne in mind that the MW is only one system and that the abundance of satellites around similar galaxies exhibits considerable scatter (Guo et al., 2012; Wang & White, 2012).

The code that implements our method to estimate the total population of MW satellite galaxies is available online (Newton & Cautun, 2018). In addition, we also make available all data that are required to reproduce our results (e.g. Fig. 3.2).

In the next two chapters, we address the second strand of this thesis: considering alternative models to the standard cosmological paradigm, and testing these using visible tracers of DM structure. The satellite galaxy population of the MW is a sensitive probe of this, and

we introduce and improve upon previous methods that use estimates of the satellite galaxy complement to constrain the parameter space of viable DM models.

CHAPTER 4

Using estimates of MW DM substructure to constrain the properties of WDM models

---

4.1 Introduction

Recent astrophysical observations have provided indirect evidence for a candidate DM particle with mass in the keV range (e.g. Boyarsky et al., 2014b; Bulbul et al., 2014b). Such a particle would be much lighter than is proposed by CDM models and would have very different clustering properties on small scales (Boyarsky et al., 2015; Cappelluti et al., 2018). This, together with a lack of any experimental detection of a CDM particle despite considerable advances in particle detector technology (e.g. Liu et al., 2017; XENON Collaboration et al., 2017), has motivated a renewed interest in possible alternatives to the CDM model (e.g. Bœhm et al., 2014; Marsh, 2016; Escudero et al., 2018). These seek to replicate the success of CDM on large scales and to explain the small-scale problems with less reliance on uncertain 'baryonic processes'. One family of these alternative DM models posit a less massive WDM particle that would have a much higher thermal velocity than its CDM counterpart at early times in the evolution of the Universe. These 'thermal relics' are formed in equilibrium with the primordial plasma with masses such that they

are relativistic at decoupling but non-relativistic by matter-radiation equality (Avila-Reese et al., 2001; Bode et al., 2001). Such particles would free-stream out of small-scale primordial density perturbations, preventing their condensation into small haloes and producing a cut-off in the linear matter power spectrum on astrophysically relevant scales. Detecting this suppression of structure relative to CDM predictions would provide a means of discriminating between the prevailing cosmological paradigm and viable WDM models. The goal of this chapter is to use visible tracers of the DM substructure to rule out thermal relic WDM models that do not produce enough subhaloes to host the observed number of low-mass galaxies.

Low mass DM-dominated galaxies provide an excellent probe of the 'small-scale' DM structure (Shen et al., 2014; Sawala et al., 2015, 2016a; Wheeler et al., 2015). The smallest and faintest of these can be observed best in the environs of the MW; however, the current census of ∼50 satellite galaxies is highly incomplete as extant surveys do not cover the entire sky to sufficient depth, and large parts of it are partially or totally obscured by the MW itself (Koposov et al., 2008; Walsh et al., 2009; Hargis et al., 2014). Simple volume corrections to the observed complement of satellite galaxies have been used already to constrain the viable parameter space of thermal relic WDM models by comparing the number of DM substructures in MW-mass haloes with the number of observed satellites (Kennedy et al., 2014; Lovell et al., 2014). Such approaches make several assumptions about the expected spatial distribution of the underlying DM substructure and about the completeness of the surveys, which could lead to a misestimation of the real satellite population. More recent estimates of the satellite galaxy luminosity function that account for the stochasticity of observational data and uncertainties arising from the variability of host haloes at fixed halo mass suggest that the size of the total complement of MW satellites could be several times larger than previously assumed (see chapters 2 and 3).

This chapter improves on previous work to constrain the properties of candidate WDM particles in several important ways, which we demonstrate using the thermal relic class of WDM models. First, we use our estimate of the total satellite population of the MW from Chapter 3, which takes advantage of new observational data to infer a population

of 124^{+40}_{-27} satellites brighter than $M_V = 0$ within 300 kpc of the Sun (this was published in Newton et al., 2018, which forms the substantive content of chapters 2 and 3). This properly accounts for the incompleteness of current surveys and the method used to obtain this estimate has been tested robustly using mock observations. Secondly, our results account for resolution effects in *N*–body simulations that prevent the identification of DM subhaloes that survive to the present day but fall below the resolution limit of subhalo finders or are destroyed by numerical effects that enhance tidal stripping (e.g. Springel et al., 2008; Onions et al., 2012; van den Bosch & Ogiya, 2018). This significant effect has been overlooked in previous studies and produces constraints on the viable parameter space of WDM models that are too restrictive. Finally, we incorporate the uncertainty in the total number of satellite galaxies, which has not been included in previous analyses.

We organize this chapter as follows. In Section 4.2, we describe a new method to constrain the properties of WDM models by comparing their predictions of the size of the MW satellite galaxy population with estimates of the total number of MW satellite galaxies from observations. We apply this methodology to thermal relic WDM and present our main results in Section 4.3. We investigate further the effect of galaxy formation processes on the constraints that we can obtain in Section 4.3.1. Section 4.4 discusses the implications of our results and considers some of the limitations of our method. We present concluding remarks in Section 4.5.

4.2 Methods

Our goal is to use the satellite luminosity function of our Galaxy to constrain the properties of WDM models using a minimal set of assumptions. In DM cosmologies, galaxies of all masses form almost exclusively¶ within DM haloes. The abundance of these can be probed readily with numerical simulations which provide a useful tool to investigate the predictions of different models; we introduce these in Section 4.2.1. A DM model is viable only if it forms enough subhaloes to host each MW satellite galaxy. To test for this condition we need two ingredients. First, we need an accurate estimate of the MW satellite luminosity function, which we discuss in Section 4.2.2. Secondly, we need a model to predict the number of substructures given the properties of the WDM particle and the mass of the host DM halo, which we describe in Section 4.2.3.

4.2.1 *N*–body simulations

We calibrate our method using high-resolution DM-only *N*–body simulations of cosmological volumes. The Copernicus Complexio (COCO) suite consists of two zoom-in simulations: one of ΛCDM that we refer to as COCO-COLD (Hellwing et al., 2016), and the other of 3.3 keV thermal relic WDM, hereafter COCO-WARM (Bose et al., 2016). These two versions differ only in the matter power spectra used to perturb the simulation particles in the initial conditions (see discussion in Section 1.4). Details of the techniques needed to generate the perturbations from the matter power spectra can be found in Efstathiou et al. (1985). Both COCO-COLD and COCO-WARM are simulated in periodic boxes of side-length $70.4\,h^{-1}$Mpc using the GADGET3 code that was developed for the Aquarius Project (Springel et al., 2008). The high-resolution regions correspond approximately to spherical volumes with radii $\sim 18\,h^{-1}$Mpc that each contains $\sim 1.3 \times 10^{10}$

¶Dwarf galaxies can also form during the collision of gas-rich massive galaxies and are known as the so-called 'tidal dwarf galaxies' (e.g. Kaviraj et al., 2012; Lisenfeld et al., 2016; Ploeckinger et al., 2018; Haslbauer et al., 2019). These are low-mass and possess negligible DM content; consequently, they are thought to be short-lived. As our Galaxy has not experienced any recent major mergers, the MW is unlikely to contain a significant population of tidal dwarf galaxies.

DM particles with mass, m_p=1.135 × 10^5 $h^{-1}\mathrm{M}_\odot$. Haloes at the edges of these regions can become contaminated with high-mass simulation particles which disrupt their evolution. We identify these contaminated haloes as having a low-resolution DM particle within $3R_{200}$ of the halo centre. The cleaned catalogues provide large samples of haloes in both cosmological models and both simulations resolve the subhalo mass functions of DM haloes down to masses $\sim 10^7$ $\mathrm{M}_\odot$. The cosmological parameters assumed for this suite of simulations are derived from the WMAP seventh-year data release (Komatsu et al., 2011): $H_0 = 70.4\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_\mathrm{M} = 0.272$, $\Omega_\Lambda = 0.728$, $n_s = 0.967$, $\sigma_8 = 0.81$.

In N–body cosmological simulations the discreteness of the simulation particles can give rise to gravitational instabilities that influence the formation of structure artificially. Models such as WDM that impose a cut-off in the primordial matter power spectrum are especially susceptible to these effects (Wang & White, 2007; Angulo et al., 2013; Lovell et al., 2014). The instabilities are resolution-dependent and lead to the artificial fragmentation of filaments, giving rise to small 'spurious' haloes that create an upturn at the low mass end of the WDM halo mass function. Lovell et al. (2014) developed a method to identify and remove these objects from halo catalogues using their mass and particle content. The onset of numerical gravitational instabilities translates into a resolution-dependent mass threshold. Haloes that are unable to surpass this during their formation and subsequent evolution are likely to be spurious. This coarse requirement is refined further by a second criterion on the particles that compose the halo when its mass is half that of its maximum, $M_\mathrm{max}\,/\,2$. In the initial conditions of the simulation, the Lagrangian regions formed by the particles in spurious haloes are highly aspherical. This is parametrized by $s_\mathrm{half-max}$=$c\,/\,a$, where a and c are the major and minor axes of the diagonalized moment of inertia tensor of the DM particles in Lagrangian coordinates. These criteria were applied to the COCO-WARM simulation by Bose et al. (2016) who find that over 91 per cent of all haloes satisfy the criteria: M_max<3.1 × 10^7 $h^{-1}\mathrm{M}_\odot$ and $s_\mathrm{half-max}$<0.165, indicating that they are spurious. The details of the calculation of these threshold values can be found in Bose et al. (2016, section 2.3). We follow the same prescription to 'clean' the COCO-WARM catalogues of spurious haloes for use throughout

the rest of this chapter.

The resolution of a simulation also affects the identification of subhaloes in the inner regions of simulated haloes (e.g. Springel et al., 2008; Onions et al., 2012). Subhaloes that fall below the resolution limit at any time are discarded by substructure finders, and some others are disrupted artificially by numerical effects. Consequently, these objects do not appear in the subhalo catalogue, even though they may still exist at the present day. We correct for this by identifying such subhaloes in COCO-COLD and COCO-WARM, tracking them to z=0, and restoring them to the subhalo catalogues that we use to calibrate our methodology. Full details of this procedure are supplied in Section 2.3.1.

4.2.2 Model-independent radial density profile

To obtain the best constraints on the WDM particle mass we need a complete census of the Galactic satellites. The satellite population is dominated by ultra- and hyperfaint galaxies with absolute magnitudes fainter than $M_V = -8$ (see chapters 2 and 3, and also Tollerud et al., 2008; Hargis et al., 2014), which can be detected only in deep surveys. This means that large areas of the sky remain unexplored and that currently, we have only a partial census of the MW satellites. However, there are several methods that use the current observations to infer the total satellite count of our Galaxy (see chapters 2 and 3, and also Tollerud et al., 2008). Here, we use the estimates from Chapter 3 that are based on a Bayesian formalism that has been robustly tested using mock observations. These results were obtained by combining the observations of the Sloan Digital Sky Survey (SDSS; Alam et al., 2015) and the Dark Energy Survey (DES; Bechtol et al., 2015; Drlica-Wagner et al., 2015), which together cover nearly half the sky area, and estimating the MW satellite luminosity function down to a magnitude, $M_V = 0$ (this roughly corresponds to galaxies with stellar mass higher than 10^2 $M_\odot$, e.g. Bose et al., 2019).

The method described in Chapter 2 takes two input components. First, it uses the sky coverage of a given survey and the detection distance from the Sun out to which a satellite galaxy of a given magnitude can be detected within it. This depends on the depth of

the survey and the satellite detection algorithm that is used. Secondly, the method needs the shape of the radial number density profile of satellites. Simulations of DM-only CDM haloes show that subhaloes selected by $v_{\rm peak}$, the highest maximum circular velocity achieved in their evolutionary histories, have the same radial number density profile as that of the observed satellites (see Chapter 2, and discussion therein).

To obtain a good estimate of the faint end of the luminosity function and the total satellite count the radial profile must be sampled well at small radii. However, in models such as WDM where the total number of small structures is suppressed relative to CDM the inner radial profile is sampled less well. Applying the method described in Chapter 2 to such haloes would create halo-to-halo scatter that would introduce artificial stochasticity into estimates of the luminosity function, particularly at the faint end. This can be addressed by 'stacking' together subhaloes from multiple haloes according to their distance from the halo centre rescaled by the R_{200} of the halo. Such an approach ameliorates the issue of small number statistics in the number of subhaloes close to the halo centre and samples well the inner radial profile; however, it reduces our ability to characterize uncertainties in the luminosity function that arise from genuine halo-to-halo variation, as this is averaged out by the stacking procedure.

In Fig. 4.1, we compare the normalized radial number density profiles of stacked populations of subhaloes obtained from the COCO-WARM and COCO-COLD simulations. The fiducial populations were obtained by selecting subhaloes with $v_{\rm peak} \geq 20$ km s^{-1} and identifying and including subhaloes that would exist at z=0 if they had not fallen below the resolution limit of the simulation (for details see Section 2.3.1). The size of the complement of MW satellite galaxies inferred using the method described in Chapter 2 depends on the *shape* of the radial number density profile and not on the total number of subhaloes. While the number of subhaloes scales strongly with halo mass, the radial number density profiles shown in Fig. 4.1 are almost completely independent of it. Instead, features in the radial density profile are mapped to different physical radial distances depending on the value of R_{200}, leading to very weak dependence on the host halo mass and a similarly weak dependence on the total number of subhaloes (see Fig. 3.5). As the radial profiles of the

CDM and WDM models are in excellent agreement at all radii, estimates of the number of MW satellites obtained using subhalo populations taken from WDM haloes should be identical to those obtained using CDM haloes. By using CDM haloes to make such predictions we can take advantage of the improved statistics that a better-sampled radial distribution provides and to incorporate into our estimates genuine halo-to-halo variation that would otherwise be averaged out by stacking.

In this chapter, we infer the satellite galaxy luminosity function of the MW within R_{200} for assumed host halo masses in the range, $\mathrm{M}_{200} = [0.5,\ 2.0] \times 10^{12}\ \mathrm{M}_{\odot}$ using the Bayesian methodology presented in Chapter 2. As we mentioned above this requires two components: (i) a tracer population of DM subhaloes with a radial profile that matches that of the observed satellites, and; (ii) a set of satellite galaxies detected in surveys for which the completeness is characterized well. For the former, we use the same v_{peak}-selected $\left(v_{\mathrm{peak}} \geq 10\ \mathrm{km\,s^{-1}}\right)$ fiducial CDM subhalo populations as used in Chapter 2. These are obtained from five high-resolution ΛCDM DM-only N–body simulations of isolated MW-like host haloes from the Aquarius suite of simulations (Springel et al., 2008). For the latter, we use the observations of nearby dwarf galaxies from the SDSS and DES supplied in Table 2.1 of Section 2.2.

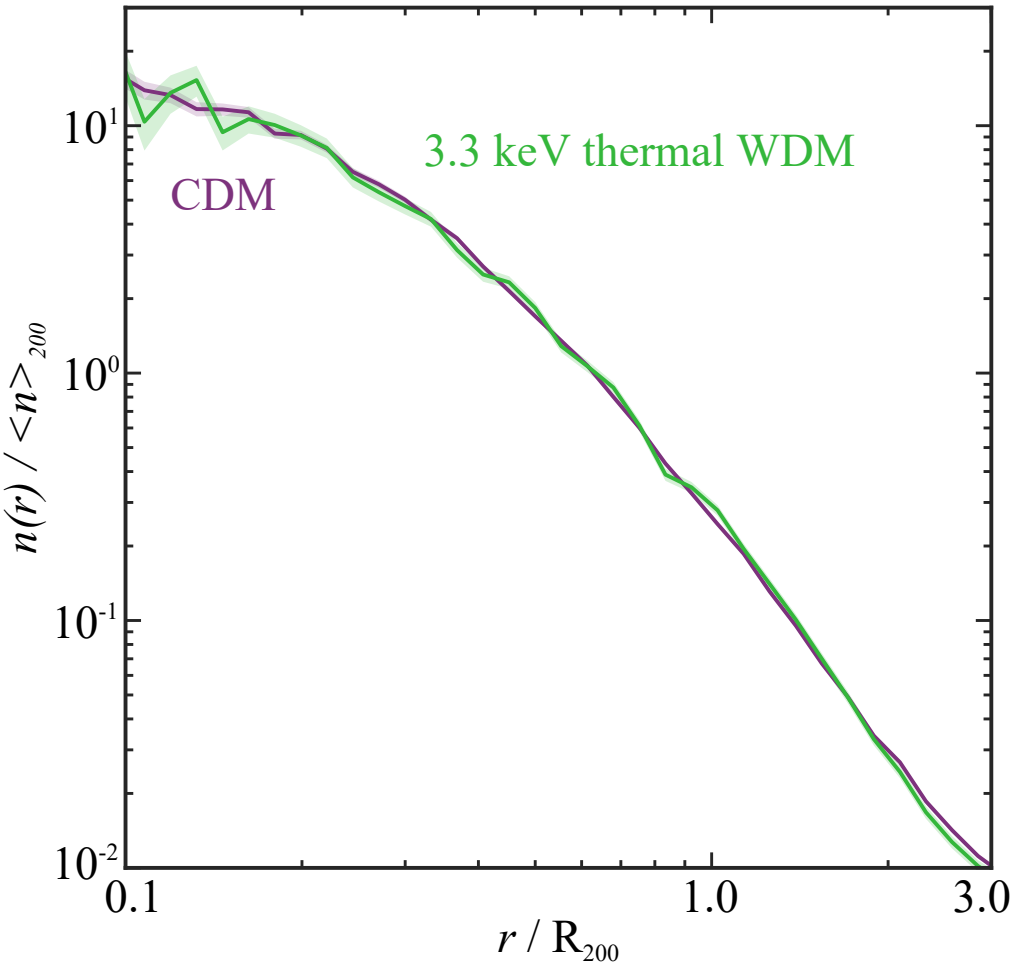


Figure 4.1

The radial number density of subhaloes with $v_{\rm peak} \geq 20$ km s^{-1} normalized to the mean density within R_{200}. The solid lines show the profiles averaged over 805 and 798 host haloes with masses $M_{200} \geq 10^{11}$ $M_\odot$ from COCO-COLD and COCO-WARM, respectively. The corresponding shaded region shows the 68 per cent scatter of the profiles over all haloes in the sample.

4.2.3 Estimating the amount of halo substructure

Estimates of the average number of subhaloes in MW-like DM haloes can be obtained using the Extended Press–Schechter (EPS) formalism (Press & Schechter, 1974; Bond et al., 1991; Bower, 1991; Lacey & Cole, 1993; Parkinson et al., 2008). In this approach, the linear matter density field is filtered with a window function to identify regions that are sufficiently dense enough to collapse to form virialized DM haloes at low redshift. In CDM models the filter employed for this task takes the form of a top-hat in real space. However, applying this to models such as WDM in which power is suppressed at small scales leads to an over-prediction of the number of low-mass haloes (Benson et al., 2013). This occurs because the variance of the smoothed density field at small radii becomes independent of the shape of the linear matter power spectrum if the latter decreases faster than k^{-3}. Consequently, the halo mass function continues to increase at small masses rather than turning over (Lovell et al., 2016; Leo et al., 2018, section 3.1), making the top-hat filter an inappropriate choice. Using a sharp-k space filter seemed to address this by accounting for the shape of damped power spectra at all radii (Benson et al., 2013; Schneider et al., 2013); however, subsequent work by Leo et al. (2018) demonstrates that this over-suppresses the production of small haloes. They find that using a smoothed version of the sharp k-space filter produces halo mass functions in best agreement with N–body simulations. Throughout this and the next chapter, we use the Leo et al. (2018) smooth k-space filter for the WDM models that we consider.

To obtain our estimates of the number of substructures, $N_{\rm sub}$, within R_{200} of MW-like haloes we follow the approach described by Giocoli et al. (2008) that was subsequently modified by Schneider (2015, section 4.4) for use with sharp k-space filters. Using the Leo et al. (2018) filter a conditional halo mass function, $N_{\rm SK}$, is generated from the primordial linear matter power spectrum. Bode et al. (2001) showed that WDM power spectra, $P_{\rm WDM}(k)$, are related to the CDM power spectrum, $P_{\rm CDM}(k)$, by $P_{\rm WDM}(k) = T^2(k)\, P_{\rm CDM}(k)$, where $T(k)$ is the transfer function given by

$$T(k) = \left[1 + (\alpha k)^{2\nu}\right]^{\frac{-5}{\nu}} . \tag{4.2.1}$$

Here, $\nu = 1.12$ and α is described by Viel et al. (2005) as being a function of the WDM particle mass, $m_{\rm th}$, given by

$$\alpha = 0.049 \left[\frac{m_{\rm th}}{\rm keV}\right]^{-1.11} \left[\frac{\Omega_{\rm WDM}}{0.25}\right]^{0.11} \left[\frac{h}{0.7}\right]^{1.22} h^{-1}{\rm Mpc}. \tag{4.2.2}$$

Schneider (2015) showed that integrating the conditional halo mass function over the redshift-dependent spherical collapse threshold of a given progenitor, $\delta_c(z)$, gives the subhalo mass function

$$\frac{{\rm d}N_{\rm sub}}{{\rm d}\ln M} = \frac{1}{N_{\rm norm}} \int_{\delta_c(0)}^{\infty} \frac{{\rm d}N_{\rm SK}}{{\rm d}\ln M}\, {\rm d}\delta_c\,, \tag{4.2.3}$$

where M is the filter mass and $N_{\rm norm}$ is a normalization constant. The latter term, which is a free parameter, corrects the total count for progenitor subhaloes that exist at multiple redshifts which are counted more than once. Using the Leo et al. (2018) filter introduces two other free parameters, $\hat{\beta}$ and $\hat{c}$, that control the 'smoothness' and the mass-radius relationship of the filter function.

We calibrate the free parameters of the EPS formalism by comparing its predictions of DM substructure with the fiducial subhalo populations of COCO haloes in the mass bin $\rm M_{200} = [0.95,\ 1.10]\times 10^{12}\ M_\odot$. Specifically, we determine the EPS free parameters by applying the following two criteria:

(i) the EPS estimate of the mean number of CDM subhaloes with mass $M_{\rm sub} \geq 10^9\ \rm M_\odot$ must equal the mean number of such objects in COCO-COLD haloes, and;

(ii) the EPS prediction of the mean number of WDM subhaloes with $M_{\rm sub} \geq 10^6\ \rm M_\odot$ (i.e. all subhaloes) must equal the mean number of such objects in COCO-WARM haloes.

We choose the latter mass as it is below the turnover in the WDM power spectrum. We obtain excellent agreement between the mean EPS estimates and the COCO simulation results by setting $N_{\rm norm} = 1.51$, $\hat{\beta} = 4.6$, and $\hat{c} = 3.9$. This is shown in Fig. 4.2, which is discussed below.

The EPS formalism described above predicts only the mean number of subhaloes in DM haloes of a given mass, and not the host-to-host scatter in the subhalo count. As we

will discuss later, including this scatter is very important to obtain unbiased results and thus needs to be accounted for. We do this using the results of cosmological N–body simulations that have shown that the scatter in the subhalo mass function is modelled well by a negative binomial distribution (Boylan-Kolchin et al., 2010). This takes the form

$$\mathrm{P}\,(N|\ r,\ p) = \frac{\Gamma(N+r)}{\Gamma(r)\,\Gamma(N+1)}\,p^r(1-p)^N\ , \tag{4.2.4}$$

where N is the number of subhaloes and $\Gamma(x) = (x-1)!$ is the Gamma function. The variable $p = \langle N\rangle\ /\,\sigma^2$, where $\langle N\rangle$ and σ^2 are, respectively, the mean and the dispersion of the distribution. This scatter in the subhalo count can be described best as the convolution of a Poisson distribution with a second distribution that describes the additional intrinsic variability of the subhalo count within haloes of fixed mass, such that $\sigma^2 = \sigma^2_{\rm Poisson} + \sigma^2_I$. The parameter r then describes the relative contribution of each of these two terms: $r = \sigma^2_{\rm Poisson}\,/\,\sigma^2_{\rm I}$. We find that the scatter in the subhalo count of haloes in the COCO suite is modelled well by σ_I=0.12 $\langle N\rangle$, as depicted in Fig. 4.2, which is discussed below. We use this approach to characterize the scatter associated with the EPS predictions throughout the remainder of this chapter.

In Fig. 4.2 we compare the EPS predictions for haloes in the mass range $[0.5,\ 2.0]\times 10^{12}\,\mathrm{M}_\odot$ to the number of subhaloes in individual COCO haloes of the same mass. We obtain excellent agreement with N–body results across the entire halo mass range of interest for this study. In particular, our approach reproduces very well both the mean number of subhaloes as well as its halo-to-halo scatter, which is shown in Fig. 4.2 by the grey shaded region and the vertical error bars.

4.2.4 Calculating model acceptance probability

We rule out sections of the viable thermal relic WDM parameter space by calculating the fraction, $f_{\rm acc}$, of WDM systems that have at least as many subhaloes as the total number of MW satellites. Let us denote with $p^{\rm EPS}$ the probability density function of the number of DM subhaloes predicted by the EPS formalism. Then, the fraction of haloes with $N^{\rm MW}_{\rm sat}$ or

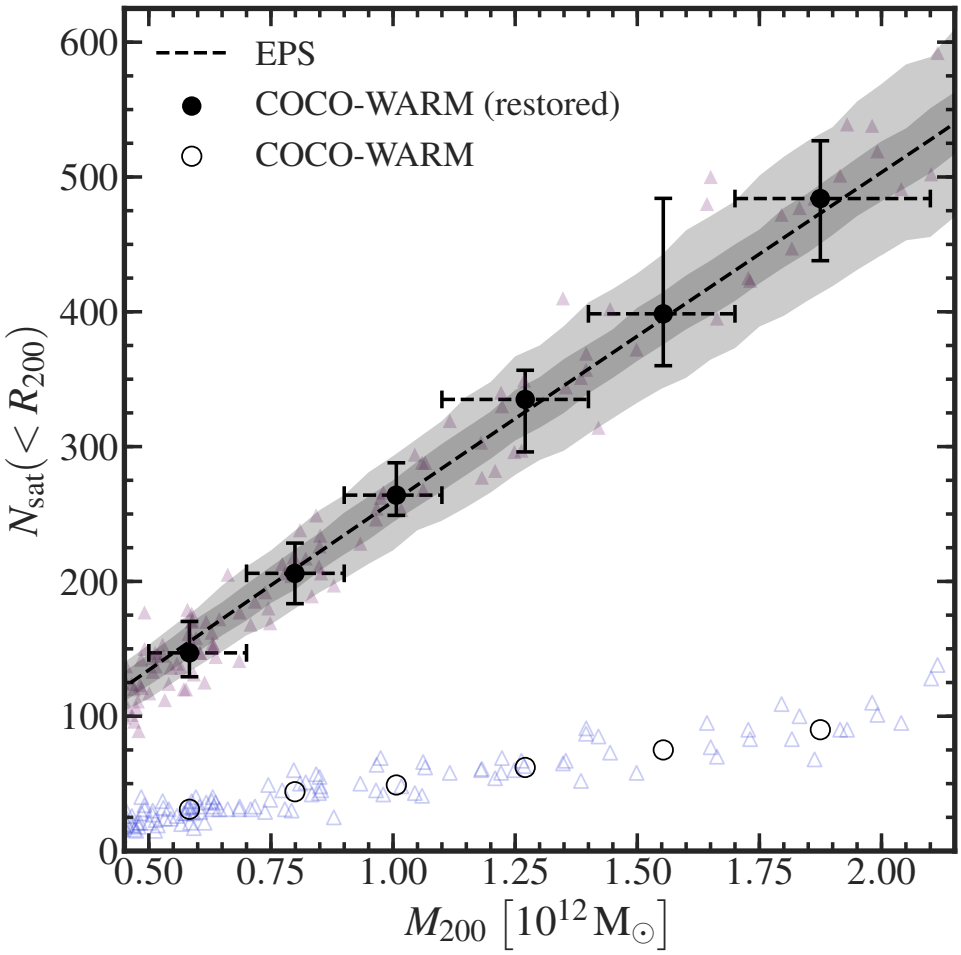


Figure 4.2

The total number of DM subhaloes within R_{200} as a function of DM halo mass, M_{200}. The dashed line shows the mean number of subhaloes predicted by the EPS formalism and the dark shaded region indicates the associated 68 per cent Poisson scatter. The light shaded region gives the 68 per cent scatter modelled using equation (4.2.4). The symbols represent haloes from the COCO-WARM simulations: unfilled symbols are from a subhalo catalogue where the 'missing' subhaloes have not been recovered, and filled symbols indicate the same haloes from a subhalo catalogue after restoration of the 'missing' subhaloes. Triangles represent individual haloes and circles represent the mean number of subhaloes in a halo mass bin with width indicated by the horizontal dashed error bars. The 68 per cent scatter in each halo mass bin obtained from the COCO-WARM simulation suite is represented by vertical error bars.

more subhaloes is given by:

$$f_{\rm acc}\left(N_{\rm sub} \geq N_{\rm sat}^{\rm MW}\right) = \int_{N_{\rm sat}^{\rm MW}}^{\infty} {\rm d}N_{\rm sub}\, p^{\rm EPS}(N_{\rm sub}) \ . \tag{4.2.5}$$

However, as we discussed in Section 4.2.2, the total number of MW satellite galaxies is affected by uncertainties. We can account for these by marginalizing over the distribution of MW satellite counts, $p^{\rm MW}\left(N_{\rm sat}^{\rm MW}\right)$. Combining everything we find that the fraction of WDM haloes with at least as many subhaloes as the MW satellite count is given by:

$$f_{\rm acc} = \int_0^{\infty} dN_{\rm sat}^{\rm MW} \left[p^{\rm MW}\left(N_{\rm sat}^{\rm MW}\right) \int_{N_{\rm sat}^{\rm MW}}^{\infty} {\rm d}N_{\rm sub}\, p^{\rm EPS}(N_{\rm sub}) \right] \ . \tag{4.2.6}$$

Although the number of MW satellites inferred in Chapter 3 has only a very weak dependence on the assumed mass of the DM halo (see Fig. 3.4), we choose to calculate the expected number of satellites within R_{200} for each MW halo mass. As the number of subhaloes varies rapidly with halo mass (e.g. see Fig. 4.2) the fraction of valid WDM haloes depends strongly on the assumed mass of the Galactic halo.

This approach to calculating the fraction of viable WDM systems for the first time incorporates the scatter in $N_{\rm sub}$ at fixed halo mass *and* the uncertainty in the inferred total MW satellite population. As we demonstrate in Fig. 4.3, excluding one, or both, of these sources of uncertainty produces constraints on $m_{\rm th}$ that are too severe. When using our methodology, these constraints would be ∼15 per cent more restrictive across the halo mass range considered than our reported values. Therefore, the results obtained by some previous analyses rule out too much of the parameter space.

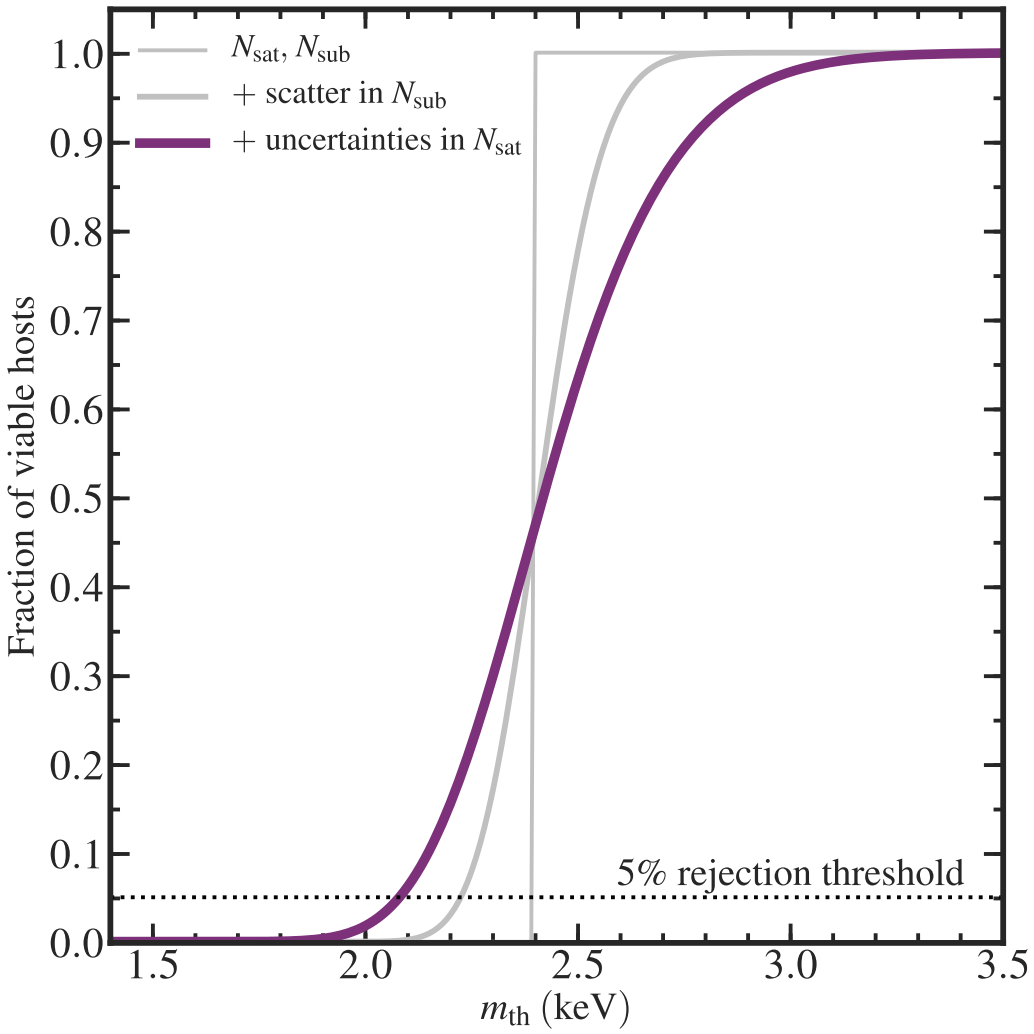


Figure 4.3

The fraction, $f_{\rm acc}$, of WDM systems with at least as many DM subhaloes, $N_{\rm sub}$, as the inferred total number of MW satellites, $N_{\rm sat}$, for a DM halo with $\rm M_{200} = 1 \times 10^{12}\ M_\odot$. Thermal relic masses with $f_{\rm acc} \leq 0.05$ are ruled out with 95 per cent confidence. Earlier works that do not account for the uncertainty in $N_{\rm sat}$ or the scatter in $N_{\rm sub}$ at fixed halo mass (thin lines) artificially exclude too many thermal relic particle mass values. In this work (thick line) we include both sources of uncertainty in our calculation. The horizontal dotted line indicates the 5 per cent rejection threshold that we use to rule out parts of the WDM parameter space.

4.3 Thermal relic particle mass constraints

Here we present the results of our analysis that were obtained using the EPS formalism calibrated to fiducial subhalo populations from the COCO-COLD and COCO-WARM simulations. Our most robust result assumes that *all* DM subhaloes that form host a galaxy, thereby making no assumptions about galaxy formation processes which cannot produce galaxies more efficiently than this.

We compute the model acceptance distributions of DM haloes in the mass range $\mathrm{M}_{200} = [0.5,\ 2.0] \times 10^{12}\,\mathrm{M}_{\odot}$ for several thermal relic WDM models. In addition to incorporating the scatter in N_{sub} and the uncertainty in N_{sat}, we account for resolution effects in the N–body simulations with which we calibrate the EPS formalism by including subhaloes that have been lost below the resolution limit or destroyed artificially by tracking the most bound particle of these objects to z=0 (see Section 2.3.1 for details). This problem persists even when using high-resolution simulations, and the effect is amplified as simulation resolution decreases. We rule out with 95 per cent confidence all combinations of M_{200} and m_{th} with $f_{\mathrm{acc}} \leq 0.05$.

The results that we obtain using this approach are displayed in Fig. 4.4. The shaded region represents the parameter combinations that we rule out with 95 per cent confidence. Within the range of halo mass values favoured by Callingham et al. (2019), who find a MW halo mass of $M_{200}=1.17^{+0.21}_{-0.15} \times 10^{12}\ \mathrm{M}_{\odot}$ (68 per cent CL), we rule out all models with $m_{\mathrm{th}} \leq 1.95$ keV. These constraints do not depend on uncertain galaxy formation physics and therefore are the most robust constraints to be placed on the thermal relic particle mass to date. A more realistic treatment of galaxy formation processes—the effect of which would be to render a large number of low-mass subhaloes invisible—would allow us to rule out more of this parameter space as fewer WDM models would produce a sufficient number of satellites to be consistent with the inferred total population. We consider this possibility in more detail in Section 4.3.1.

In Fig. 4.4, we include for comparison the constraints obtained by Polisensky & Ricotti

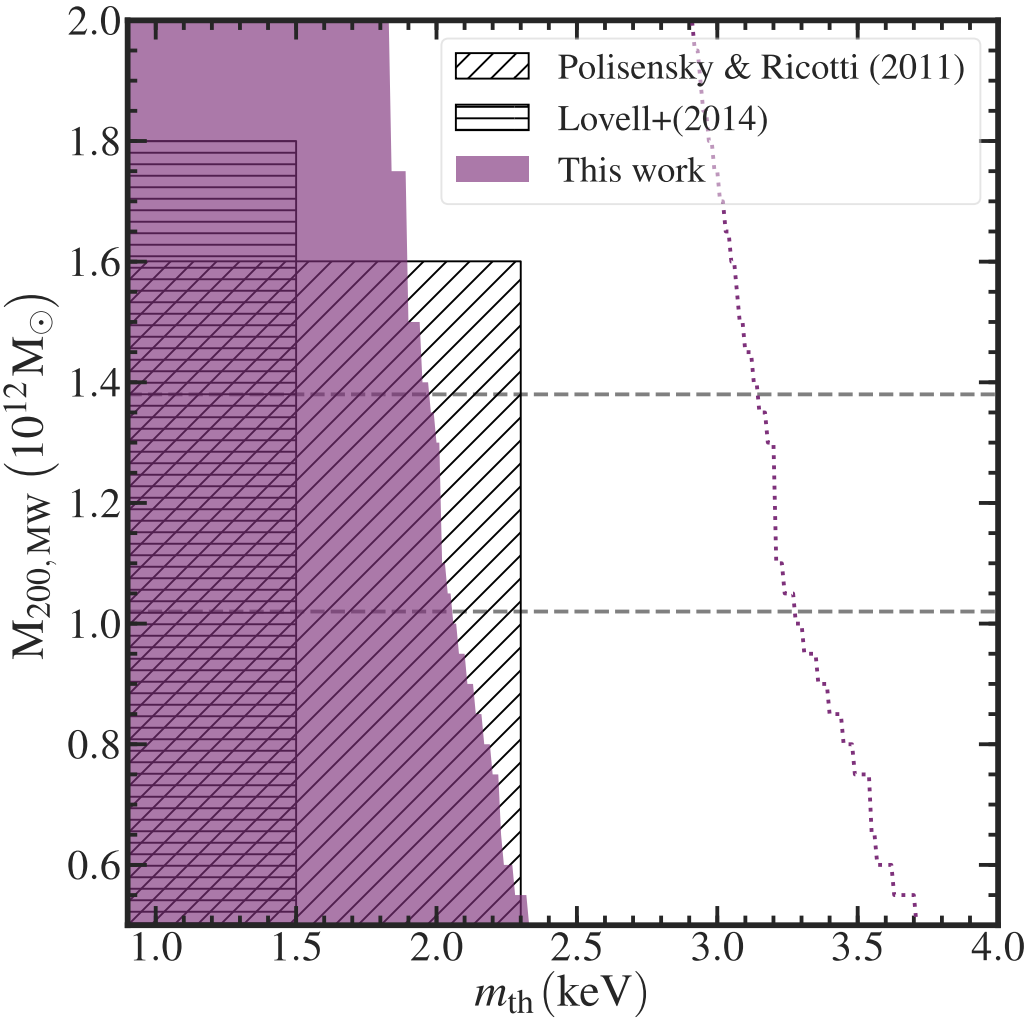


Figure 4.4

Constraints on the particle mass, m_{th}, of the thermal relic WDM model. These depend on the assumed mass of the MW halo, which is shown on the vertical axis. We exclude with 95 per cent confidence parameter combinations in the shaded region. The dotted line indicates the extent of the exclusion region if we do not include 'missing' subhaloes when calibrating the EPS formalism with the COCO simulations (see Section 4.2.3 for details). The constraints obtained by previous works, which do not consider some of the highest MW halo masses displayed here, are indicated by the hatched regions. These are overly restrictive as they do not account for some sources of uncertainty (see Section 4.2.4 for details). The two dashed horizontal lines show the 68 per cent confidence range for the mass of the MW halo (Callingham et al., 2019).

(2011) and Lovell et al. (2014) who use similar analysis techniques. These constraints suffer from resolution effects that suppress the identification of some substructures that survive to the present day. The dotted line demarcates the exclusion region that we would obtain in our analysis if we *did not* account for these missing subhaloes. Such issues are not revealed by numerical convergence tests that are typically used to assess the reliability of particular simulations. For example, even the 'level 2' simulations of Aquarius haloes, which are some of the highest resolution DM-only haloes available, are not fully converged (see Section 2.3.1).

4.3.1 Modelling galaxy formation processes

In the preceding sections, we described an approach to determine the viability of WDM models that we used to produce a highly robust, albeit conservative, lower limit on the allowed mass of the thermal relic particle. However, it does not account for the effect of galaxy formation processes on the satellite complement of the MW. These mechanisms play an important role in the evolution of the satellite galaxy luminosity function but still are not fully understood. Semi-analytic models of galaxy formation allow for the fast and efficient exploration of such processes and thus to understand how they affect the WDM constraints.

GALFORM (Cole et al., 1994, 2000) is one of the most advanced semi-analytic models that are currently available and is tuned to reproduce a selection of properties of the local galaxy population. A complete summary of the observational constraints used to calibrate the GALFORM model parameters is provided in Lacey et al. (2016, section 4.2; hereafter L16). Of particular interest to our study is the reionization of the Universe, which is the main process that affects the evolution of the faint end of the galaxy luminosity function. In particular, the UV radiation that permeates the Universe (and which is responsible for reionization) heats the intergalactic medium and prevents it from cooling into low-mass haloes, thus depleting the reservoir of cold gas from which stars form.

In GALFORM, the effect of reionization on haloes is modelled using two parameters: a

circular velocity cooling threshold, $V_{\rm cut}$, and the redshift of reionization, $z_{\rm reion}$. The intergalactic medium is taken to be fully ionized at a redshift, z=$z_{\rm reion}$, whereafter the cooling of gas into haloes with circular velocities, $v_{\rm vir} < V_{\rm cut}$, is prevented. This simple scheme has been verified against more sophisticated calculations of reionization, with which it has been shown to produce a good agreement (Benson et al., 2002a; Font et al., 2011). Recent studies by e.g. Bose et al. (2018) have characterized the sensitivity of the satellite galaxy luminosity function to changes in these parameters: a later epoch of reionization allows more faint satellites to form, and a smaller circular velocity cooling threshold permits those faint satellites to become brighter.

We use GALFORM to explore the effect of different parametrizations of reionization on the number of luminous structures around the MW. This parameter space has been constrained well already by theoretical calculations and the analysis of recent observational data, so we consider parametrizations in the ranges $6 \leq z_{\rm reion} < 7$ and $25\ {\rm km\,s^{-1}} \leq V_{\rm cut} \leq 35\ {\rm km\,s^{-1}}$ (Okamoto et al., 2008; Font et al., 2011; Robertson et al., 2015; Bañados et al., 2018; Davies et al., 2018; Mason et al., 2018; Planck Collaboration et al., 2018).

4.3.2 Constraints using GALFORM models

We explore the different parametrizations of reionization that are described in Section 4.3.1 by modifications to the L16 model applied to the COCO-COLD and COCO-WARM merger trees. We use these *N*–body results to calibrate the algorithm used in GALFORM to construct Monte Carlo merger trees of these models for the thermal relic WDM particle masses in the range of interest. The merger trees are generated within GALFORM using an implementation of the Parkinson, Cole & Helly (2008, hereafter PCH) merger tree algorithm which iteratively splits the present-day halo mass into different branches as it progresses to higher redshifts. The algorithm depends on three free parameters: G_0=0.57, a normalization constant; γ_1=0.38, which controls the mass distribution of the progenitor haloes; and γ_2= − 0.01, which controls the halo-splitting rate. PCH calibrated these parameters by comparing the Monte Carlo progenitor halo mass functions at several

redshifts with those from the Millennium simulation (Springel et al., 2005). This follows the evolution of 2160^3 particles with mass, $m_p = 8.6 \times 10^8\, h^{-1}\mathrm{M}_\odot$, resolving the halo mass function to $\sim 1.7 \times 10^{10}\, h^{-1}\mathrm{M}_\odot$, which is three orders of magnitude larger than the regime of interest for this study. The best-fitting free parameter values derived from the Millennium simulation predict a factor of two times more faint galaxies than is obtained by applying GALFORM to the COCO suite. John Helly performed the PCH calibration procedure using the COCO simulations and find best-fitting values of G_0=0.75, γ_1=0.1 and $\gamma_2 = -0.12$; however, this still overestimates the faint end galaxy luminosity function by a factor of 1.6 compared to COCO. Further improvements to the PCH algorithm require more thorough investigation and possibly the introduction of one or more additional free parameters; these are beyond the scope of this work.

To generate the model-dependent predictions in this section, we apply GALFORM to Monte Carlo merger trees calibrated as closely as possible to the COCO suite. By applying the modified L16 models to these merger trees we obtain the dwarf galaxy luminosity function for 1000 realizations of each MW halo mass, allowing us to compute the model acceptance distributions in the same manner as before (see Section 4.2.4). As the merger trees over-predict the faint-end galaxy luminosity function, our constraints on the thermal relic WDM particle mass are less restrictive than they would be if we were able to obtain better agreement between the Monte Carlo and N–body luminosity functions. In agreement with previous work, we find that the resulting constraints on m_{th} are stronger for larger values of z_{reion} and smaller values of V_{cut} and that the value chosen for V_{cut} has the largest effect on the number of substructures at fixed halo mass (e.g. Kennedy et al., 2014).

Several previous works that explored the effect of reionization on the satellite complement used the L16 model, which has $z_{\mathrm{reion}} = 10$ and $V_{\mathrm{cut}} = 30\ \mathrm{km\,s^{-1}}$. This combination of parameters is disfavoured by recent observations (e.g. Mason et al., 2018; Planck Collaboration et al., 2018). Additionally, others have noted that using this value of z_{reion} is not self-consistent and that a modified L16 model with $z_{\mathrm{reion}} = 6$ is a more appropriate choice (Bose et al., 2018). This modified L16 model is a viable parametrization that is consistent with observations, so we adopt this to obtain the constraints that we plot in

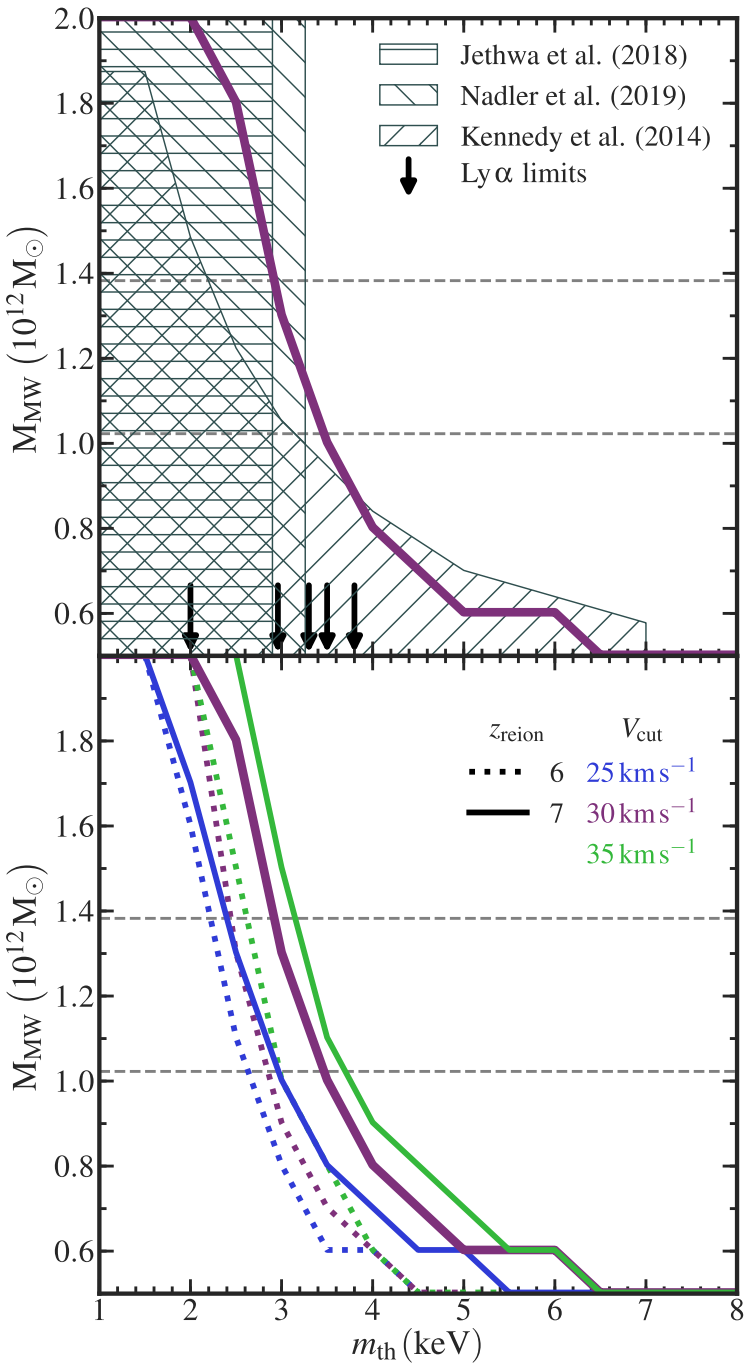


Figure 4.5

Upper panel: Constraints on the m_{th} mass obtained assuming our fiducial $z_{reion} = 7$, $V_{cut} = 30$ km s^{-1} model of reionization within the GALFORM galaxy formation model (thick solid line). Parameter combinations to the left of and beneath the envelope are ruled out with 95 per cent confidence. The constraints obtained by previous works that adopted similar approaches are displayed by the hatched regions (Jethwa et al., 2018; Nadler et al., 2019). Arrows indicate the 2 keV (Safarzadeh et al., 2018), 2.96 keV (Baur et al., 2016), 3.3 keV (Viel et al., 2013), 3.5 keV (Iršič et al., 2017), and 3.8 keV (Hsueh et al., 2019) envelopes of the most robust constraints on the thermal relic particle mass obtained from the Ly α forest. *Bottom panel:* Constraints obtained from other parametrizations of reionization. Different choices of $V_{cut} = [25, 30, 35]$ km s^{-1} are indicated by the colours purple, blue, or green. The dotted and solid lines indicate the envelopes of the constraint regions that assume $z_{reion} = 6$ or $z_{reion} = 7$, respectively.

the upper panel of Fig. 4.5. In this model, we rule out *all* thermal relic WDM particle masses with $m_{\rm th} \leq 2.0$ keV, and within the most likely MW halo mass range we rule out with 95 per cent confidence all models with $m_{\rm th} \leq 2.9$ keV. These constraints improve on our model-independent constraint and strengthen the constraints in different MW halo mass regimes when compared with work by Kennedy et al. (2014), Jethwa et al. (2018), and Nadler et al. (2019) which also incorporate the effects of galaxy formation models. In Fig. 4.5 we also include for comparison the most conservative constraints derived from the Ly α forest by Viel et al. (2013), Baur et al. (2016), Iršič et al. (2017), Safarzadeh et al. (2018), and Hsueh et al. (2019), which our results complement.

4.4 Discussion

We have placed new conservative and highly robust constraints on the mass of the thermal relic WDM particle by comparing EPS predictions of the DM subhalo content of WDM haloes with the total number of MW satellite galaxies inferred from observations. We obtain estimates of the total satellite complement using the approach described in Chapter 2 with recent observations of satellites from the SDSS and DES. To calibrate the EPS formalism we use DM haloes from the COCO simulation suite with masses in the likely MW halo mass range $\mathrm{M_{200}}=[0.5,\ 2.0] \times 10^{12}\ \mathrm{M_\odot}$. We improve upon previous constraints by incorporating for the first time the uncertainty in the size of the total MW satellite complement and by accounting for unresolved or numerically disrupted subhaloes in *N*–body simulations; these are not identified easily by convergence tests. We also explore the effect of various assumptions about galaxy formation processes on the constraints that we can place on the WDM particle mass.

We find that, for MW DM haloes with mass $\mathrm{M_{200}} \leq 1.38 \times 10^{12}\ \mathrm{M_\odot}$, thermal relic models with $m_{\rm th} \leq 1.95$ keV are ruled out with 95 per cent confidence (see Fig. 4.4). Our result is independent of assumptions about galaxy formation physics, as for our purposes we treat *all* DM subhaloes as hosts of visible galaxies. This ensures that the constraints provide a robust lower limit on the mass of the thermal relic WDM particle, improving

on the results reported in Lovell et al. (2014) across the entire MW halo mass range considered (see Fig. 4.4). Our results are competitive with but slightly less restrictive than the constraints obtained by Polisensky & Ricotti (2011) because we account for subhaloes that are 'missing' from the z=0 halo catalogues.

The resolution of a simulation can affect the population of haloes at z=0 in two major ways. First, haloes close to the resolution limit of simulations experience stronger tidal disruption due to numerical effects which can destroy the halo. Secondly, structure finders stop tracking haloes that fall below a mass threshold at *any* time during their evolution. Haloes composed of few simulation particles can occasionally fall below this, even if the main object survives to the present day; as a result, they are not included in the final catalogue. Excluding these objects significantly affects the constraints on the WDM parameter space, strengthening them artificially (see Fig. 4.4). This effect worsens as simulation resolution decreases, so constraints that are obtained using lower-resolution simulations and using methods that do not account for the 'missing' subhaloes will be a significant overestimate.

The processes underpinning the formation of galaxies are complex and are yet to be understood fully; nevertheless, they play an important role in shaping the luminosity function of the dwarf galaxies of the MW. Incorporating the effect of these mechanisms into our approach allows us to refine the constraints on the properties of the DM and rule out many more WDM models. In a modified version of the L16 GALFORM model with $z_{\rm reion} = 6$, and for MW DM haloes with mass $\mathrm{M_{200}} \leq 1.38 \times 10^{12}$ $\mathrm{M_\odot}$, we rule out with 95 per cent confidence thermal relic models with $m_{\rm th} \leq 2.9$ keV (see Fig. 4.5). Furthermore, we rule out all thermal relic WDM particle masses with $m_{\rm th} \leq 2.0$ keV. These improve on our model-independent results and are consistent with the constraints obtained in previous work by Kennedy et al. (2014), Jethwa et al. (2018), and Nadler et al. (2019), which adopted similar approaches. However, we note that the overestimation of the satellite galaxy luminosity function by the PCH tree algorithm suppresses our constraints compared to what should be achievable. This is noticeable in the $z_{\rm reion} = 6$ constraints, which are *less* restrictive on the thermal relic mass parameter space than our conservative estimates from Section 4.2.2. Future work to improve the algorithm will also improve

our constraints. Despite this caveat, this result compares favourably with complementary constraints derived from the Ly α forest by Baur et al. (2016) and Safarzadeh et al. (2018), and is consistent with other results by Viel et al. (2013), Iršič et al. (2017) and Hsueh et al. (2019).

Our key results assume a MW halo mass up to $M_{200} = 1.38 \times 10^{12}$ $M_\odot$, although we also consider halo masses in the range $M_{200} = [0.5, 2.0] \times 10^{12}\, M_\odot$. The constraints have a moderate dependence on host halo mass because the number of MW satellite galaxies inferred from observations scales much less strongly with halo mass than the number of subhaloes predicted by DM models (see Chapter 4). Better measurements of the mass of the MW halo will improve the number of thermal relic models that can be ruled out; in the most extreme case a MW halo with mass at the lowest end of the likely range would rule out thermal relic models with $m_{\rm th} \leq 2.3$ keV independently of galaxy formation physics. This estimate does not account for the effect of the central baryonic disc of the host halo, which destroys subhaloes, and would exclude more of the WDM parameter space. Accounting for reionization in our fiducial baryonic scenario, this rules out $m_{\rm th} \leq 4.5$ keV with 95 per cent confidence.

The size of the satellite population inferred by the method described in Chapter 2 is a lower limit to the true population as it cannot account for spatially-extended dwarf galaxies that fall below the surface brightness threshold of the surveys. Additionally, it does not encompass the contribution of the former satellites of the Large Magellanic Cloud that lie outside the DES footprint that could increase the size of the satellite complement still further. Taken together these caveats *strengthen* the robustness of our lower limits on the thermal relic particle mass as a larger inferred satellite complement would rule out an even larger WDM parameter space.

4.5 Conclusions

In the continued absence of the direct detection of a DM particle or the observation of an astrophysical phenomenon that unambiguously constrains the properties of the DM

particle, the debate about its exact nature—and the acceptability of the current cosmological paradigm—will continue. This has been spurred on by the apparent 'small-scale' challenges to ΛCDM: discrepancies between the observations of low-mass galaxies and predictions of DM substructure, which so far remain unexplained within this framework, leading to renewed interest in alternative DM models that are free of such issues (see Chapter 1, Section 1.3). One class of these, which are broadly termed WDM models, produces a cut-off in the linear matter power spectrum that, depending on the properties of the DM particle, can suppress the formation of DM haloes on the scale of (and smaller than) those that would usually host dwarf galaxies. One method to constrain the parameter space of these models is the use of sophisticated hydrodynamic simulations to simulate self-consistently the formation and evolution of dwarf galaxies in the Local Group, and around MW-like hosts in particular. However, the resolution that would be required to achieve this in a volume that is sufficiently large enough to attain reasonable statistical power is, at present, computationally prohibitive, and likely will be for some time to come. The development of other approaches to explore efficiently the viability of different cosmological models on these scales is, therefore, an imperative.

In this work, we develop an improved method to constrain the properties of WDM models by comparing EPS predictions of the amount of substructure within MW-mass WDM haloes with the most recent estimates of the size of the satellite population of the MW (see sections 4.2.3 and 4.2.4). This approach is complementary to previous work and for the first time accounts fully for limitations in the resolution of N–body cosmological simulations, incorporates the scatter in the number of substructures inside haloes at fixed DM halo mass, and includes the uncertainty associated with estimates of the number of satellite galaxies in the MW. The constraints that can be produced by this method rule out efficiently many WDM models independently of any particular choice of galaxy formation physics, making the results highly robust.

We demonstrate the utility of this approach by applying it to thermal relic WDM models to constrain the DM particle mass (see Section 4.3). Our most robust constraint rules out with 95 per cent confidence thermal relic WDM particles with masses $m_{\rm th} \leq 1.95$ keV. This

is competitive with existing limits that also use the abundance of MW satellite galaxies to constrain the WDM parameter space; however, our approach accounts for small subhaloes in N–body simulations that are not identified by substructure finders, despite some of these objects surviving to z=0. Excluding them from the subhalo catalogue reduces the number of subhaloes that are available to host dwarf galaxies, artificially strengthening restrictions on the viable thermal relic particle mass parameter space (see Fig. 4.4). This effect worsens as simulation resolution decreases, so constraints that are obtained using lower-resolution simulations, without accounting for the 'missing' subhaloes, will be a significant overestimate.

All methods that seek to constrain the properties of DM models using visible tracers of the underlying substructure must make assumptions about galaxy formation processes that affect the satellite complement of the MW. Here, to obtain our highly robust constraints on the allowed properties of candidate WDM particles independently of galaxy formation physics, we have assumed that a galaxy forms in *all* DM haloes. While the details of baryonic physics mechanisms are still not understood fully, it is clear that many small subhaloes are unlikely to host a luminous component. Accounting for these physical processes in models would reduce the effective size of the satellite complement and in our analysis would significantly improve the constraints on the WDM particle properties.

In particular, the reionization of hydrogen in the early Universe, and the size of DM haloes in which it suppresses galaxy formation is thought to be the dominant process that determines the luminosity function of MW satellite galaxies. We use the Durham semi-analytic model GALFORM to explore several possible descriptions of this process and examine how different parametrizations affect the constraints on thermal relic WDM (see Section 4.3.1). By assuming that reionization is complete by $z_{\rm reion} = 6$ and that galaxy formation is suppressed in DM haloes with circular velocity $v_{\rm vir}$<25 km s^{-1}, we rule out with 95 per cent confidence thermal relic particles with mass $m_{\rm th}\leq$2.9 keV in MW haloes with mass $\rm M_{200}\leq 1.38\times10^{12}$ $\rm M_\odot$ (see Fig. 4.5). Of all options considered this parametrization of reionization produced the largest population of luminous satellite galaxies in MW haloes of fixed mass, making this constraint a lower bound. This improves

on the Kennedy et al. (2014) result and is competitive with recent analyses using the Ly α forest (e.g. Iršič et al., 2017; Safarzadeh et al., 2018).

While a DM particle candidate remains undetected, thermal WDM models remain a feasible alternative to CDM. The satellite galaxy system of the MW provides a powerful means of probing structure formation on small scales and can help to discriminate between different cosmological models. However, the MW may not be typical of most DM haloes of similar mass. Hydrodynamic simulations that self-consistently model star formation and gas physics on the scale of dwarf galaxies will facilitate more robust astrophysical tests of this; however, achieving sufficient resolution at present is computationally prohibitive. A complementary means of testing the predictions of structure formation from different cosmological models is by considering their predictions of the evolution of structure across a range of mass scales and comparing these with observations. Currently, this is challenging to achieve as it is difficult to identify such faint objects at such vast distances against observational backgrounds. Future improvements in observational capability will offer the prospect to constrain further the parameter space of viable WDM models.

In the next chapter, we apply the technique introduced here to another model of WDM that shows promise as a candidate cosmological model: the Neutrino Minimal Standard Model (νMSM). In addition to the mass of the proposed DM candidate, this model introduces a dependence on the primordial lepton asymmetry of the Universe, granting it an extra degree of freedom that enables it to evade current astrophysical bounds on the mass of thermally-produced WDM. We constrain this parameter space using the estimates of the MW satellite galaxy population and discuss the implications of this result in the context of recent astrophysical observations.

CHAPTER 5

Constraining the particle properties of νMSM sterile neutrino DM

5.1 Introduction

As we discussed in Chapter 1, the emergence of discrepancies between CDM predictions of structure formation and observations of galaxy clustering motivated the consideration of alternative cosmological models (Efstathiou et al., 1990; Maddox et al., 1990). One class of these that gained attention was WDM, the DM particles of which have a warmer momentum distribution than CDM in the early Universe; however, at the time no mechanism to produce the DM particle had been found. Following the discovery of the accelerating expansion of the Universe (Riess et al., 1998; Schmidt et al., 1998; Perlmutter et al., 1999), the introduction of a positive cosmological constant solved many of the problems of CDM, and the combined ΛCDM model has achieved considerable success at reproducing observations of the large-scale clustering of galaxies and predicting numerous detailed properties of the CMB. However, it still suffers from several problems on smaller astrophysical scales (discussed in detail in Section 1.3), renewing interest in alternative cosmological models. One WDM model offers an appealing alternative to CDM in its own right: the νMSM.

The νMSM is a minimal extension of the Standard Model of particle physics (Asaka & Shaposhnikov, 2005; Asaka et al., 2005; Canetti et al., 2013a,b). It entails the addition of three massive right-chiral neutrinos to the family of three 'active' neutrinos that are already known. These 'sterile' neutrinos do not couple to the gauge bosons that mediate the fundamental forces of the Standard Model, which interact exclusively with left-chiral fermions and right-chiral anti-fermions. As a result, the only mechanism through which sterile neutrinos interact with almost all ordinary matter is via gravitational interactions, a key property of DM candidates. Additionally, the lightest sterile neutrino is expected to have $\mathcal{O}(\,\mathrm{keV})$ mass and could be long-lived on cosmological timescales (Adhikari et al., 2017), placing it in parameter space that is within reach of terrestrial and astrophysical experiments.

A definitive test of DM models is to identify and measure the properties of the particle. However, to date, no candidate DM particle has been detected in direct detection experiments (LUX Collaboration et al., 2014; XENON Collaboration et al., 2017, 2018) or particle collider searches (e.g. Khachatryan et al., 2016; Aaboud et al., 2018; Sirunyan et al., 2018). In the absence of such a detection, the most viable approach to infer *indirectly* the properties of the DM is with astrophysical observational techniques. In particular, one technique that has shown promise is the search for characteristic signatures of DM decay in the spectra of systems that are thought to host a significant DM component. Recent studies of such systems have led to the detection of an unexplained excess at 3.55 keV in the stacked X-ray spectrum of several galaxy clusters (Bulbul et al., 2014b) and the X-ray spectrum of the centre of M31 (Boyarsky et al., 2014b). This elicited considerable interest as there is no known astrophysical origin for a spectral line at this energy. Several mechanisms have been proposed to describe this line, such as Sulphur charge exchange (Gu et al., 2015; Shah et al., 2016) and via the modelling of observational backgrounds (Boyarsky et al., 2014a; Bulbul et al., 2014a; Jeltema & Profumo, 2015; Cappelluti et al., 2018); however, these are yet to be confirmed. If this line is interpreted as the decay product of a DM particle the parent particle mass, 7.1 keV, would sit comfortably in the range expected of a sterile neutrino candidate responsible for all of the DM component of

the matter density. This possibility spurred a considerable number of follow-up attempts to detect it in other systems at higher significance. These include the centre of the MW (Boyarsky et al., 2015; Perez et al., 2017), the Perseus cluster (Urban et al., 2015; Franse et al., 2016, but see also Aharonian et al., 2017), and in other galaxy clusters (Bulbul et al., 2016, but see also Hofmann et al., 2016). While this has met with some success, other studies of stacked galaxy spectra (Anderson et al., 2015) and the Draco dwarf galaxy (Ruchayskiy et al., 2016; Sonbas et al., 2016) either appear to rule out the existence of a 3.55 keV signal or are unable to determine its origin. It is evident that additional, higher resolution measurements of the soft X-ray spectrum in such objects, as will be achievable with XRISM (Tashiro et al., 2018; Lovell et al., 2019a,b), will be needed to determine unambiguously if the line truly exists or is merely some instrumental systematic.

Another approach to constrain the properties of the DM is to characterize its macroscopic effects on astrophysical structure. On large scales the clustering of galaxies in WDM models behaves identically to CDM (Bond et al., 1982; Olive & Turner, 1982; Pagels & Primack, 1982; Peebles, 1982b); however, on smaller scales, the two models exhibit different behaviour. In WDM models the relativistic velocities of the DM particles suppress the formation of structure relative to CDM below some characteristic length scale, leaving a distinctive and potentially observable fingerprint on the abundance of small DM haloes (see Section 1.4). We are best able to probe these using the population of nearby dwarf galaxies, which act as tracers of the underlying DM substructure of the MW halo.

The current incompleteness of satellite galaxy surveys—due variously to difficulties with the zone of avoidance and a lack of sky coverage at sufficient depth—has motivated several attempts to estimate the total size of the MW satellite galaxy population (see chapters 2 and 3, and also Koposov et al., 2008; Tollerud et al., 2008; Hargis et al., 2014). As we showed in Chapter 3, the most recent estimate of the size of this population indicates that we expect to find 124^{+40}_{-27} such objects with magnitudes brighter than $M_V = 0$ within 300 kpc of the MW. This provides a means of assessing the viability of different WDM models: those that are unable to produce a sufficient number of DM subhaloes to host this population can be ruled out. We introduced the methodology behind this approach in

Chapter 4 and tested it on the generic class of thermal relic WDM models. Here, we adopt the same approach to constrain the properties of the νMSM.

5.1.1 The neutrino Minimal Standard Model

In the νMSM, the sterile neutrinos are produced via the propagation of left-chiral neutrinos into the right-chiral states. This is characterized by the 'mixing angle', θ_M, which is extremely small and limits the oscillation rates of the right-chiral neutrino species into left-chiral, and vice-versa. They obtain their masses via a mechanism known as the 'Type–I seesaw' (Minkowski, 1977; Gell-Mann et al., 1979; Barbieri et al., 1980; Mohapatra & Senjanović, 1980; Yanagida, 1980). The mixing of the active and sterile neutrino states produces two distinct sets of mass eigenstates: as the masses in one set increase, the other set becomes lighter; hence, the 'seesaw' nomenclature. Such a mechanism offers a plausible explanation for the extremely small masses of the active neutrinos of the Standard Model. The high mass eigenstates are ascribed to the sterile neutrinos, producing one with mass $\mathcal{O}(\,\mathrm{keV})$—the DM candidate—and two others with masses $\mathcal{O}(100\ \mathrm{GeV})$ (for a detailed review see Abazajian et al., 2012). Such high mass states are unstable and extremely short-lived (Adhikari et al., 2017). Their decay into leptons soon after the Big Bang introduces an asymmetry in the abundance of leptons relative to anti-leptons in a process known as leptogenesis (Fukugita & Yanagida, 1986). Shortly thereafter this lepton asymmetry is partially reprocessed into a baryon asymmetry during baryogenesis, providing a natural explanation for the observed asymmetry of matter to anti-matter in the Universe (Shaposhnikov, 2008; Canetti et al., 2012, 2013b).

The lepton asymmetry in the early Universe also has a direct effect on the production of the DM sterile neutrino species, as active neutrinos propagating through matter have different oscillation parameters compared to those propagating in vacuum. This phenomenon is described by the Mikheyev–Smirnov–Wolfenstein (MSW) effect (Wolfenstein, 1978; Mikheyev & Smirnov, 1985). Active neutrinos propagating in a medium weakly interact with electrons present in the material, increasing their effective mass of the neutrinos. The

oscillation parameters are proportional to the square of the difference in neutrino mass of the flavour states, so neutrino oscillations in matter may be different from those in a vacuum. This can be seen readily in the flavour oscillation of active neutrinos passing through the Sun and the Earth which creates a deficit of ν_e compared to theoretical predictions of models of the fusion processes at the centre of the Sun. This discrepancy was originally known as the solar neutrino problem (Bahcall, 1964; Davis, 1964). In the early Universe, the lepton asymmetry affects the mixing parameters between active and sterile states in much the same way: the effective mass of the active neutrinos in the primordial plasma is increased, resonantly enhancing the production of sterile neutrinos (Laine & Shaposhnikov, 2008).

The lepton asymmetry of the early Universe can be parametrized as

$$L_6 = 10^6 \frac{n_{\nu_e} - n_{\bar{\nu}_e}}{s}, \tag{5.1.1}$$

the ratio of the difference between the number densities of electron neutrinos and electron anti-neutrinos ($\bar{\nu}_e$) to the entropy density of the Universe, s. Constraints on the cosmic DM density restrict the parameter space of the DM properties such that there is a monotonic relationship between L_6 and θ_M for a sterile neutrino of given particle mass, m_s (Lovell et al., 2016). Sterile WDM models can, therefore, be parametrized entirely by m_s and L_6. While observational constraints on L_6 are very weak and permit a large number of possible values (Oldengott & Schwarz, 2017), the mechanisms that generate lepton asymmetry in the νMSM restrict this range to $0 < L_6 \leq 700$ (Boyarsky et al., 2009, but see also Canetti et al., 2013a,b). We explore this parameter space for a range of possible values of m_s in the sections that follow.

The size of the lepton asymmetry can significantly affect the momentum distribution of the DM particles at early times. The behaviour is non-monotonic: at small L_6 values, resonances appear at low momenta. As L_6 increases the amplitude and location of these resonances shift, initially producing *colder* DM distributions. Only at high L_6 values do the resonances advance to high enough momentum states that the production of DM at all momenta is enhanced (Laine & Shaposhnikov, 2008; Lovell et al., 2016). The

'temperature' of the νMSM DM, therefore, depends non-monotonically on the lepton asymmetry in the early Universe, allowing these models to circumvent cosmological constraints on the particle mass, including Ly α bounds for thermal neutrinos. A more comprehensive discourse concerning the effects of L_6 on the momentum distribution of resonantly-produced sterile neutrino WDM can be found in Lovell et al. (2016).

In this chapter, we constrain the parameter space of resonantly produced sterile neutrino WDM models by adopting the approach introduced in Chapter 4. We summarize this approach as applied to νMSM WDM in Section 5.2 and calibrate the method using DM-only N–body simulations of these models. We present the constraints on the viable $m_{\rm s}$–L_6–MW halo mass parameter space in Section 5.3 and summarize our findings and present concluding remarks in Section 5.4.

5.2 Method

In Chapter 4, we described an approach to constrain the properties of WDM models by comparing their predictions of the abundance of DM substructure in MW-mass haloes with the total population of MW satellite galaxies inferred from observations. We tested this using thermal relic WDM models in Section 4.3. Here, we perform the same procedure to constrain the properties of the νMSM, using the EPS formalism to generate subhalo mass functions for combinations of $m_{\rm s}$ and L_6.

5.2.1 Calibrating the EPS formalism with numerical simulations

In the EPS formalism, the density field is filtered using a window function to identify regions that are dense enough to collapse into virialized haloes by the present day. From this, the abundance of structure at different mass scales can be calculated (see Section 4.2.3 for full details). The form of the filter function can affect predictions of the mean number of subhaloes in a halo of a given mass. In CDM models a top-hat filter produces a good agreement with the results of numerical simulations because the power spectrum changes

slowly on small scales. However, using this filter with DM models that have a cut-off in the power spectrum over-predicts the number of low mass objects (Benson et al., 2013). A smooth filter in k-space produces results in better agreement with simulations, which we verified in the previous chapter using the COCO suite. The smooth filter introduces two free parameters: $\hat{\beta}$, and $\hat{c}$ (Leo et al., 2018); and the subhalo mass function also depends on a normalization constant, $N_{\rm norm}$.

We calibrate the EPS parameters using DM-only simulation counterparts of the 'V2' and 'V5' sets of Local Group-analogue volumes introduced by Lovell et al. (2017). These were simulated using the same initial conditions (except for differences in the properties of the DM) as the first six simulation volumes of the APOSTLE Project (Fattahi et al., 2016; Sawala et al., 2016b) and are each composed of a CDM volume and two 7 keV νMSM volumes with $L_6 = [10,\ 120]$. All simulations adopt cosmological parameters from the WMAP seventh-year data release (Komatsu et al., 2011): $H_0 = 70.4\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_{\rm M} = 0.272$, $\Omega_\Lambda = 0.728$, $n_s = 0.967$, $\sigma_8 = 0.81$.

DM models with a power spectrum cut-off form 'spurious' haloes produced by the artificial fragmentation of filaments (Lovell et al., 2014). This is caused by resolution-dependent gravitational instabilities that are generated by the discreteness of the simulation particles. We identify and remove these objects from the subhalo catalogues by adopting the same procedure that we introduced in Chapter 4 to clean the halo catalogues of thermal relic WDM models (Lovell et al., 2014). We also correct for other resolution-dependent effects: the failure of structure finders to identify some small haloes that survive to the present day, and the tidal disruption of small haloes due to numerical effects. We restore these to the z=0 subhalo catalogue by tracking the most-bound particle in the structure to the present day using the procedure described in Section 2.3.1.

We obtain excellent agreement between the EPS predictions and the νMSM APOSTLE N–body results by setting $N_{\rm norm} = 1.4$, $\hat{\beta} = 4.2$, and $\hat{c} = 3.9$. These values are very similar to the values we obtained in Chapter 4 for thermal relic WDM. In Fig. 5.1 we plot the mean number of subhaloes predicted by the EPS formalism within R_{200} of the centre of MW-mass DM haloes in the $L_6 = 10$. The shaded regions represent the 68 per cent scatter

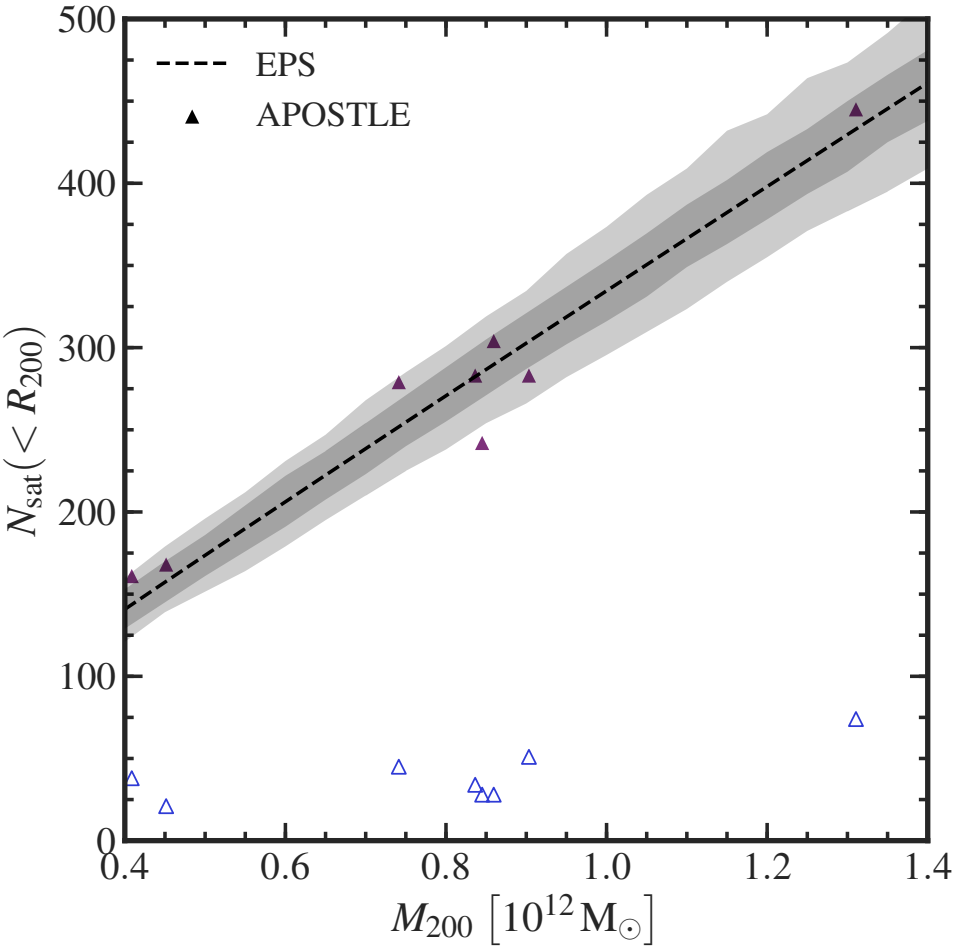


Figure 5.1

The total number of DM subhaloes within R_{200} as a function of DM halo mass, M_{200}. The dashed line shows the mean number of subhaloes predicted by the EPS formalism and the dark shaded region indicates the associated 68 per cent Poisson scatter. The light shaded region gives the 68 per cent scatter modelled using a negative binomial distribution (see equation 4.2.4). The symbols represent haloes from the 7.1 keV νMSM APOSTLE simulations for $L_6 = 10$: unfilled symbols are from a subhalo catalogue where the 'missing' subhaloes have not been recovered, and filled symbols indicate the same haloes after restoration of the 'missing' subhaloes by following the procedure detailed in Section 2.3.1.

in $N_{\rm sub}$ at fixed halo mass obtained analytically from a negative binomial distribution that models well the scatter seen in numerical simulations (Boylan-Kolchin et al., 2010, see also Section 4.2.3). When compared with the calibration plot of 3.3 keV thermal relic WDM (see Fig. 4.2), the 7.1 keV, $L_6 = 10$ νMSM model produces more DM subhaloes at fixed halo mass. Therefore, larger fractions of these systems host at least as many satellites as the inferred total MW satellite population. As no constraints could be placed on the 3.3 keV thermal relic model in Chapter 4, this approach will also be unable to constrain this νMSM model.

5.2.2 Computing the constraints

The constraints on the νMSM depend on the assumed mass of the MW, $m_{\rm s}$, and L_6. We consider MW halo masses in the likely range $(0.5–2.0) \times 10^{12}$ $\mathrm{M}_{\odot}$, and lepton asymmetry parameters, $L_6 = [0,\ 700]$; in the same range as used by Lovell et al. (2016).

For each combination of parameters, we compute the fraction of νMSM WDM systems that have at least as many DM subhaloes as the number of MW satellite galaxies inferred from observations using the methodology presented in Chapter 2. The model acceptance probability is given by equation (4.2.6) and accounts for the scatter in $N_{\rm sub}$ at fixed halo mass and the uncertainty in the number of MW satellite galaxies. We rule out combinations of parameters that have acceptance probability, $f_{\rm acc} \leq 0.05$, with 95 per cent confidence.

5.3 Constraints on the properties of sterile neutrinos in the νMSM

Here, we present the results of our analysis that were obtained using the EPS formalism calibrated to fiducial subhalo populations from DM-only νMSM versions of the APOSTLE simulations. Our analysis makes no assumptions about the specifics of galaxy formation processes. Instead, we assume that all DM subhaloes that form host a galaxy, which maximizes the model acceptance probability and produces extremely robust lower limits on the viable parameter space.

In Fig. 5.2 we plot the exclusion envelopes of the νMSM for several values of L_6. Parameter combinations to the left of, and below, the envelopes are ruled out with 95 per cent confidence. The coldest model, which predicts the highest number of subhaloes at fixed halo mass and corresponds to $L_6 = 12$, rules out all νMSM models with $m_s \leq 1.65$ keV when assuming a MW halo mass of $M_{200} = 1.38 \times 10^{12}$ $M_\odot$. This corresponds to the upper 84 per cent limit on the MW halo mass obtained by Callingham et al. (2019). The warmest model tested, with $L_6 = 700$, rules out $m_s \leq 7.85$ keV, although such a large value of L_6 is disfavoured by observational constraints if a 7.1 keV sterile neutrino composes all of the DM (Boyarsky et al., 2014b, 2015; Bulbul et al., 2014b; Iakubovskyi et al., 2015; Ruchayskiy et al., 2016).

The constraints on the νMSM WDM parameter space become less restrictive if the mass of the MW DM halo is at the upper end of the allowed range. In Fig. 5.3, we assume a fiducial MW halo mass of $M_{200} = 1.4 \times 10^{12}$ $M_\odot$, which is consistent with the upper 84 per cent bound on the halo mass obtained by Callingham et al. (2019). The parameter combinations in the shaded region are ruled out with 95 per cent confidence. The convoluted shape of the exclusion region arises from the response of the momentum distribution of the DM to different values of the lepton asymmetry. Values of L_6 in the middle of the range we consider, i.e. those values between $L_6 = 7$ and $L_6 = 30$, produce the coldest DM models. As a result, there is little suppression of small-scale power in these

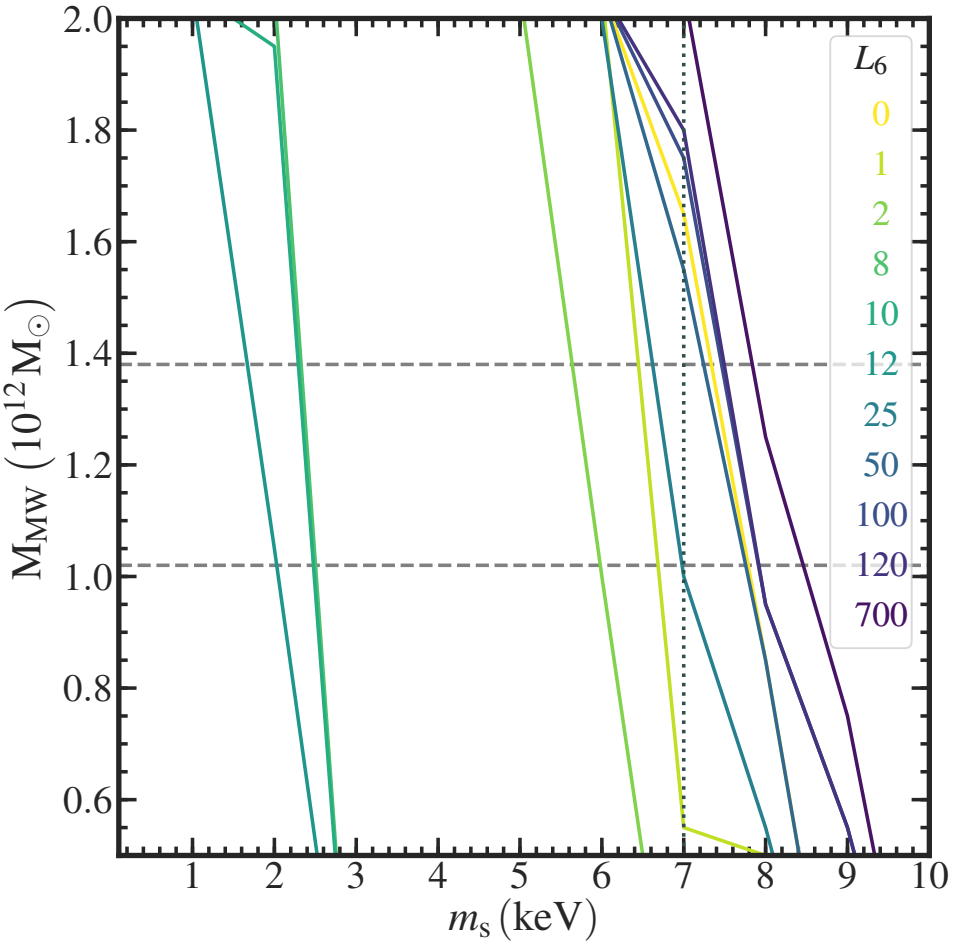


Figure 5.2

Constraints on viable parametrizations of the νMSM. These depend on the assumed mass of the MW halo, shown on the vertical axis, the mass of the sterile neutrino DM particle, shown on the horizontal axis, and the lepton asymmetry parameter, L_6. We exclude with 95 per cent confidence parameter combinations to the left of, and below, the envelopes represented by the solid lines. The two dashed horizontal lines show the 68 per cent confidence range for the mass of the MW halo obtained by Callingham et al. (2019).

models across almost all sterile neutrino rest mass values we consider. We also include in Fig. 5.3 the observational constraints from X-ray line non-detections (Watson et al., 2012; Horiuchi et al., 2014) and gravitational lensing measurements (Vegetti et al., 2018), which restrict further the allowed parameter space. As the DM neutrino mass decreases below $m_s = 6$ keV, the viable parameter space reduces significantly. At our fiducial halo mass we rule out all parametrizations of the νMSM with $m_s \leq 1$ keV; this constraint also holds at $M_{200} = 2 \times 10^{12}$ $M_\odot$, so this neutrino mass is ruled out entirely.

If the 3.55 keV line originates from the decay of a 7.1 keV sterile neutrino, and if such particles compose all of the DM content of the Universe, strong constraints can be placed on θ_M using X-ray flux observations (Boyarsky et al., 2014b, 2015; Bulbul et al., 2014b; Iakubovskyi et al., 2015; Ruchayskiy et al., 2016). Expressed in terms of the primordial lepton asymmetry, these observations favour $9 < L_6 \leq 11.2$ (Lovell et al., 2017), and we indicate this range with an error bar in Fig. 5.3. We find that this region of parameter space is unconstrained for a 7 keV νMSM neutrino; however, at values of $L_6 \geq 50$ the 7 keV model is ruled out with 95 per cent confidence.

In Fig. 5.4, we explore the effect of different assumed values of the MW halo mass on the constraints that we place on L_6 in the $m_s = 7$ keV νMSM parametrization. As before, combinations of parameters in the blue-shaded and hatched regions are ruled out with 95 per cent confidence, from our analysis and X-ray non-detections, respectively. We find that $L_6 = 700$ is ruled out for all likely values of the MW halo mass and if we assume our fiducial MW halo mass, values of $L_6 = 0$ and $L_6 > 40$ are also ruled out. The purple-shaded region indicates the range of L_6 values that are favoured by X-ray flux observations if the DM comprises 7.1 keV νMSM neutrinos. All L_6 values in this range and up to $L_6 = 15$ are not constrained by our analysis or by observational limits. We highlight the boundaries of the unconstrained region of parameter space between the favoured MW halo mass bounds with a thick solid line.

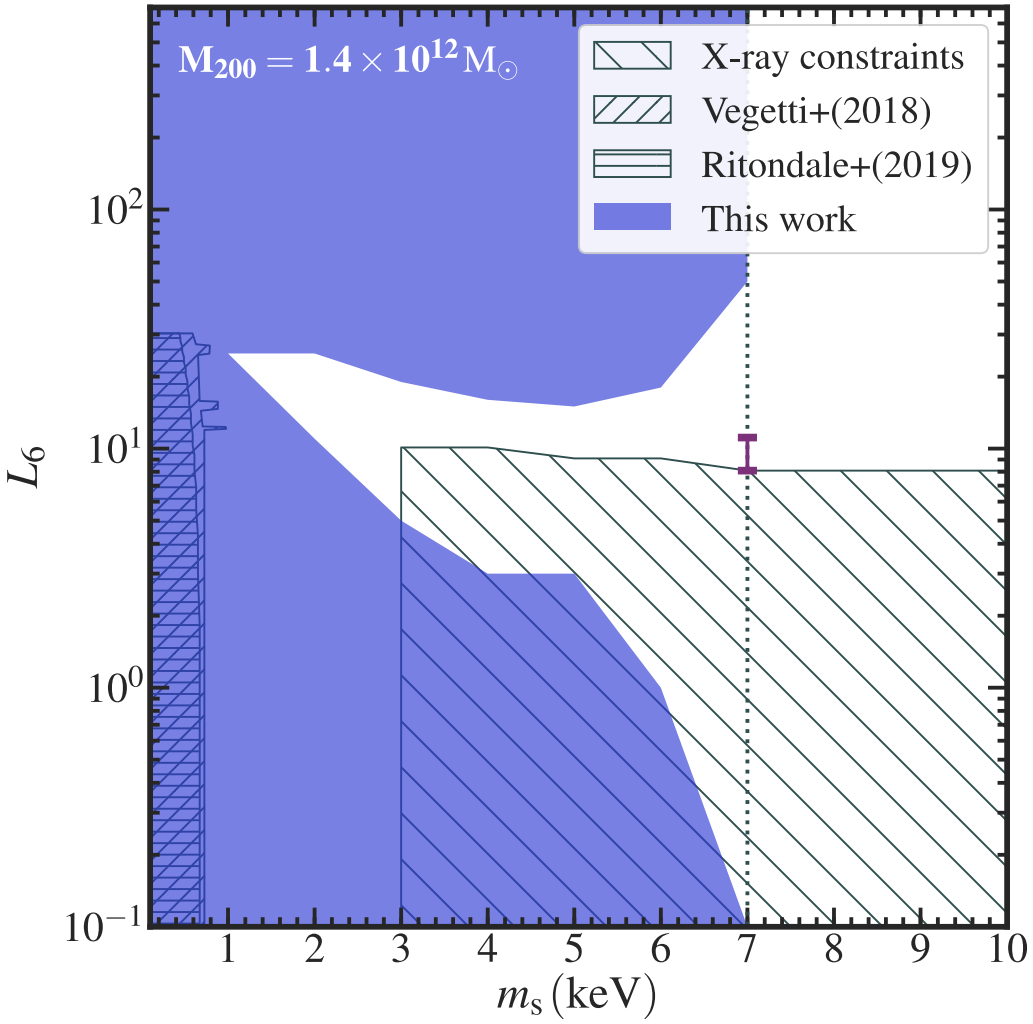


Figure 5.3

Constraints on the m_s–L_6 parameters of the νMSM assuming a MW halo mass of $M_{200} = 1.4 \times 10^{12}\ M_\odot$. As before, we exclude with 95 per cent confidence parameter combinations in the shaded region. The hashed regions provide constraints from X-ray non-detections (Watson et al., 2012; Horiuchi et al., 2014) and gravitational lensing (Vegetti et al., 2018; Ritondale et al., 2019), and the error bar indicates the $9 < L_6 \leq 11.2$ range favoured by X-ray detections if the DM comprises 7.1 keV νMSM neutrinos.

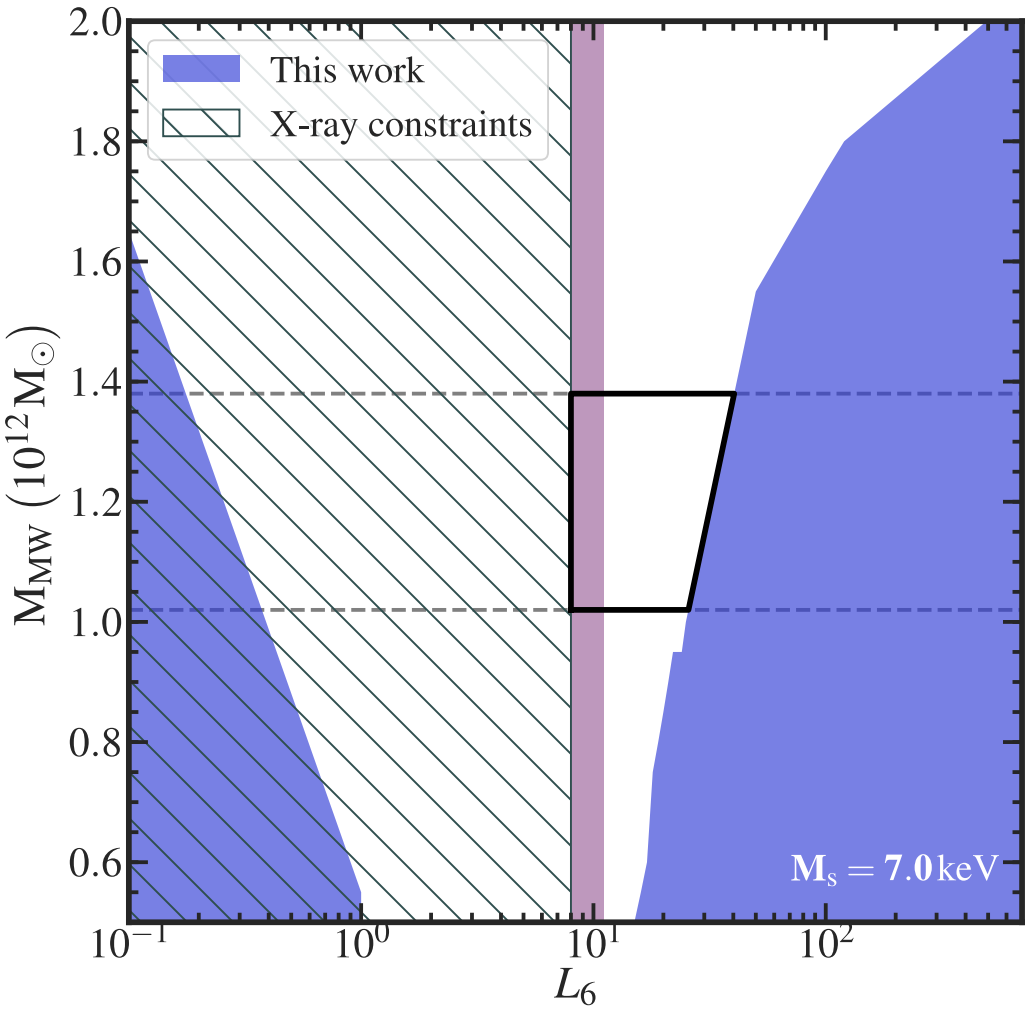


Figure 5.4

Constraints on the L_6 parameter for different MW halo masses, assuming a νMSM 7.1 keV neutrino composes all of the DM. As before, we exclude with 95 per cent confidence parameter combinations in the blue-shaded regions, and the hatched region provides constraints from X-ray non-detections (Watson et al., 2012; Horiuchi et al., 2014). The vertical purple-shaded region indicates the $9 < L_6 \leq 11.2$ range favoured by X-ray flux observations if the DM comprises 7.1 keV νMSM neutrinos. The unconstrained parameter space between the favoured MW DM halo mass bounds is indicated by the thick solid line.

5.4 Discussion and conclusions

The νMSM is a simple extension to the Standard Model of particle physics that produces a $\mathcal{O}(\,\mathrm{keV})$ mass neutrino with properties that could enable it to act as the DM and evade current astrophysical bounds on the mass of thermal WDM. Such a DM candidate could explain observations of a 3.55 keV X-ray emission line in the spectra of systems dominated by a significant DM component, such as galaxy clusters and in M31. Additionally, the νMSM provides a natural explanation for other open questions in cosmology such as leptogenesis and the matter asymmetry of the Universe. This model is, therefore, an appealing alternative to the standard cosmological paradigm.

The only mechanism to produce the DM candidate of the νMSM is by the mixing of the active neutrinos of the Standard Model of particle physics with a right-chiral, 'sterile' state. This is controlled by the mixing angle, θ_M, which is extremely small such that, alone, this mechanism cannot account for the size of the DM component of the observed matter density of the Universe. However, the neutrino mixing angle can be enhanced when the neutrino passes through a medium, in a process known as the MSW effect. In the νMSM this occurred in the early Universe when the active neutrinos propagated through the primordial plasma in the presence of a significant lepton asymmetry, L_6. This led to the resonant production of DM particles and introduces an additional degree of freedom into the model. The momentum distribution of the resulting particles, which affects the clustering properties of the DM on astrophysical scales, is therefore, a function of two parameters: the particle mass, m_s, and the lepton asymmetry, L_6. Certain values of L_6 produce sterile neutrinos with 'colder' momentum distributions than a thermally-produced DM particle of the same mass. This enables the νMSM to evade existing astrophysical constraints on the thermal DM particle mass.

We showed in Chapter 4 that the satellite galaxies of the MW are a sensitive probe of the underlying DM structure and that this can be used to constrain the viable parameter space of WDM models. Here, we adopt the same approach to place robust lower limits on the m_s–L_6 parameter space of the νMSM by comparing predictions of the abundance of DM

substructure in MW-mass haloes with estimates of the number of satellite galaxies inferred from observations of the MW. Models that produce an insufficient number of subhaloes to host the inferred number of dwarf galaxies are ruled out.

We reject $m_s \leq 1$ keV at 95 per cent confidence for all combinations of $L_6 \geq 0$ and MW halo mass 0.5×10^{12} $M_\odot < M_{200} \leq 2.0 \times 10^{12}$ $M_\odot$; however, constraints on the other values of m_s are not as strong. Assuming a fiducial MW halo mass of $M_{200} = 1.38 \times 10^{12}$ $M_\odot$ and a DM mass, $m_s = 7$ keV, we rule out models with $L_6 \geq 50$ with 95 per cent confidence. Supplemented by constraints from the non-detection of X-ray lines in observations of M31 and nearby dwarf galaxies (Watson et al., 2012; Horiuchi et al., 2014), the viable range of L_6 now lies between 8 and 50. This is consistent with the bounds obtained from direct detection experiments, collider searches, and other astrophysical measurements, and encompasses the $9 < L_6 \leq 11.2$ range favoured by X-ray flux measurements if sterile neutrinos compose all of the DM.

Our results are lower limits on the parameter space permitted by theoretical calculations of the abundance of small-scale DM structure and several constraints derived from observational measurements. Stronger results can be obtained by applying models of baryonic physics to populate the DM subhaloes with luminous galaxies; we defer such analysis to future work. In addition to improvements in theoretical estimates, the constraints will be strengthened further by future improvements to the completeness and depth of surveys of the satellite galaxies of the MW. The planned XRISM mission will also resolve the soft X-ray spectrum sufficiently to confirm or refute a physical origin of the 3.55 keV line.

CHAPTER 6

Summary

The ΛCDM cosmological model and its predictions of numerous properties of the Universe have proved extremely resilient to many detailed observational tests. This relatively simple model is the culmination of decades of effort and relates the structure seen on astronomical scales to the properties of a subatomic fundamental particle that lies beyond the Standard Model of particle physics—the disparity in scale could not be more vast. Despite its success, to date, no such CDM particle has been found either in collider searches, or other direct or indirect detection experiments. Given this, it behoves us to ask: can an alternative DM model explain the nature of the Universe?

Our approach to answering this question comprised two distinct strands. In the first, we addressed a gap in observational capability concerning the nearby dwarf galaxies of the Local Group. These highly DM-dominated objects are abundant in the Universe, making them excellent small-scale cosmological probes of the underlying DM distribution. The number of such objects produced by different DM models depends on the momentum distribution of the DM particle at early times, which is also a function of the particle mass. 'Warmer' models suppress structure formation on small scales and, potentially, the formation of dwarf galaxies. The corollary of this is that the observed abundance of dwarf galaxies can be used to constrain the viable parameter space of alternative DM models.

This was the concern of the second strand of this thesis, in which we focused on the WDM family of models. Of these, we focused on two of the most studied: thermal relics and sterile neutrinos.

6.1 Satellite galaxies of the Milky Way

The cosmological importance of dwarf galaxies has motivated a number of observational campaigns to search for these faint objects. This is an extremely difficult undertaking: at relatively small distances from us the apparent brightness of these objects reduces below background levels, complicating the acquisition of a large statistical sample. Thus far, technological limitations in instrument design and the algorithms used to distinguish dwarf galaxies from the Galactic foreground stars have restricted the sensitivity of surveys. A better prospect to probe the faint end of galaxy formation may be found even closer to home with the satellite galaxies of the MW, the proximity of which allows the population to be characterized at fainter magnitudes. However, distinguishing these objects against observational fore- and backgrounds is still a challenge. The combined effects of the Galactic foregrounds, the disc of the MW itself, and the area of the sky that must be surveyed to sufficient depth, preclude the detection of a large fraction of this population.

In Chapter 2, we developed a Bayesian methodology to infer the luminosity function of the satellite galaxies of the MW from partial observations of this population in surveys whose selection functions have been characterized well. As tracers of the underlying DM substructure, their spatial distribution can tell us a great deal about the formation history of the MW, which must necessarily affect the characteristics and size of the population of Galactic satellites. It is, therefore, a key ingredient in the method which takes as a prior the radial distribution of satellite galaxies. For this, we used fiducial populations of subhaloes taken from the high-resolution DM-only simulations of MW-mass galaxies of the Aquarius Project. We found that selecting subhaloes by the highest maximum circular velocity achieved during their evolution, a property we refer to as $v_{\rm peak}$, produces a good match with the distribution of luminous dwarf galaxies in the APOSTLE hydrodynamic

simulations of the MW and Local Group. It also reproduces well the radial distribution of the satellite galaxies observed in the SDSS and DES, after survey incompleteness has been accounted for.

We tested the approach by applying it to a large set of partial observations of a mock population of MW satellite galaxies. The method performed extremely well, successfully reproducing the luminosity function of the satellite galaxies and modelling correctly uncertainties in the estimate. This is a significant improvement compared with previous approaches, which underestimated the uncertainties associated with their predictions because they do not account for stochastic effects. These dominate observations of satellites in a given survey volume and are responsible for most of the uncertainty in the estimate of the total satellite galaxy population.

In Chapter 3, we applied this approach to the satellite galaxies observed in the SDSS and DES. These surveys were chosen as their radial completeness has been characterized well, and together, they cover almost half of the sky. From a combined analysis of the satellite populations from both surveys, we predict that there are 124^{+40}_{-27} (68 per cent CL) satellite galaxies brighter than $M_V = 0$ within 300 kpc of the Sun. Almost 40 per cent of these will be ultrafaint dwarf galaxies with $-8 < M_V \leq -3$, and a full half of the predicted population will be hyperfaint galaxies with $-3 < M_V \leq 0$. These estimates are lower limits to the total number of satellites as they do not account for low surface brightness objects that may have been missed in current observations, nor do they account for satellites brought in by the LMC which today lie outside the DES footprint. Future advanced surveys that can probe significantly deeper will see many more satellite galaxies. We predict that the Large Synoptic Survey Telescope (LSST) will see half of the population of satellites that we inferred here.

6.2 Small-scale cosmology: Probing the nature of DM

A requirement of a viable DM model is that it must correctly reproduce the observed structure of the Universe. On large scales such a comparison is relatively straightforward: luminous galaxies trace the DM and comparisons of theoretical and observed clustering properties can be used to assess the viability of a given model. However, on smaller scales this scenario becomes more complex. In this regime structure formation becomes highly non-linear, and numerical simulations become necessary to probe these scales. Additionally, baryons start to play a more important role in the dynamical evolution of galaxies. Both of these effects complicate comparisons of theoretical predictions with observations, as the luminous material may no longer faithfully trace the underlying DM.

The need for visible probes of structure on these scales is a key motivation to understand better the population of MW satellite galaxies as these objects are highly DM-dominated. To be viable, cosmological models must produce at least as many subhaloes in MW-mass DM haloes as there are observed satellite galaxies. While such an approach is appealing in its conceptual simplicity, it is computationally prohibitive to explore the full parameter space of DM models in this way carrying out a large number of numerical simulations to assess the viability of alternative DM models to ΛCDM in this way is computationally prohibitive. To circumvent this restriction, in Chapter 4 we introduced an approach that uses the Extended Press-Schechter formalism to generate analytic estimates of the abundance of DM structure in MW-mass haloes. These are calibrated with numerical simulations of the few DM models that have been carried out. We found that this produced a good match with the results of numerical simulations and that the halo-to-halo scatter in the number of DM subhaloes could be modelled with a negative binomial distribution.

As we mentioned above, galaxy formation processes and baryonic physics decouple the direct correspondence of luminous galaxies to the underlying DM structure. These processes are also not understood in detail, further complicating direct comparisons of DM substructure abundance and the number of Galactic satellite galaxies. We therefore considered two approaches: in the first, we assumed that all DM subhaloes host a galaxy.

This maximizes the probability of a given DM model producing sufficient DM subhaloes to host the estimated population of MW satellite galaxies, and produces extremely robust lower limits on the allowed parameter space of DM models. In the second approach we used the Durham semi-analytic model of galaxy formation, GALFORM, to account for the effects of reionization and supernovae feedback on the number of luminous satellite galaxies in a host DM halo.

We chose to consider one class of cosmological models that have shown promise as an alternative to the standard cosmological paradigm: WDM. The cut-off in the power spectrum on small scales suppresses the formation of structure and could provide a solution to the small-scale challenges to ΛCDM, while its predictions of the large-scale structure are identical to the prevailing standard model. We focus on two WDM models in the sections that follow: thermal relic WDM, and sterile neutrinos from the Neutrino Minimal Standard Model (νMSM).

6.2.1 Thermal relic WDM

In Chapter 4 we introduced an approach to assess the viability of alternative DM models using analytic predictions of the abundance of DM substructure in MW-mass haloes. We applied this to the thermal relic model of WDM, which proposes a DM particle that is relativistic at early times but becomes non-relativistic by matter-radiation equality. This model is parametrized by the mass of the thermal relic particle, to which the momentum distribution of the DM at early times is related: models with a particle that is lower in mass are 'hotter', erasing more structure at small scales than 'colder' ones. Our most conservative constraints place a robust lower limit on the mass of the thermal relic particle, ruling out models with $m_{\rm th} \leq 1.95$ keV with 95 per cent confidence. This is competitive with, and improves upon, existing constraints. We also found that resolution effects in numerical simulations can have an enormous impact on the constraints on DM models. Failing to account for these can lead to artificially restrictive constraints on the DM model parameter space.

We refined our constraints further by using the semi-analytic model GALFORM to incorporate the effects of supernovae feedback and reionization. This generates Monte Carlo merger trees of DM models and applies a galaxy formation prescription to produce predictions of numerous properties of luminous galaxies in DM haloes. The algorithm to generate the merger trees is calibrated to the Millennium simulation, the resolution of which is three orders of magnitude larger than the regime of interest here, and produces predictions of the satellite galaxy luminosity function that are too large when compared with N–body simulations. A partially successful attempt to recalibrate this with the COCO suite of simulations, which are higher resolution, resulted in merger trees that produce 1.6 times more faint galaxies than obtained from the N–body simulations. A more detailed investigation of this will be required for future work.

Despite this, we demonstrated the potential of the GALFORM models to rule out more of the thermal relic parameter space. In our fiducial model where reionization has completed by $z_{\rm reion} = 7$ and which suppresses the cooling of gas into haloes with circular velocity, $v_{\rm circ} \leq 30\ {\rm km\,s^{-1}}$, we rule out thermal relic WDM with mass $m_{\rm th} \leq 2.9$ keV. This is consistent with previous work and with recent observational constraints derived from the Ly α forest. However, we note that these constraints are less restrictive than they otherwise would be if the GALFORM Monte Carlo merger tree algorithm could be calibrated correctly.

6.2.2 Sterile neutrino DM and the νMSM

The second WDM model that we consider is the νMSM. This provides a well-motivated mechanism to produce a candidate DM particle, which takes the form of a right-chiral, 'sterile' neutrino. This particle has $\mathcal{O}(\,{\rm keV})$ mass and is resonantly produced in the early Universe in the presence of a lepton asymmetry. A particle with mass in this range could also explain recent observations of a 3.55 keV X-ray line in the spectra of several galaxy clusters and M31, which might be a decay product of the DM particle. This particular model has also attracted interest as it provides a natural explanation for leptogenesis and

the subsequent matter asymmetry of the Universe. It is therefore an appealing alternative to the standard paradigm.

The νMSM is parametrized by the mass of the sterile neutrino, m_s, and the size of the lepton asymmetry, L_6. The lepton asymmetry has a significant effect on the momentum distribution of the DM particles at fixed m_s, and this behaviour is non-monotonic. Certain values produce sterile neutrinos with 'colder' momentum distributions than a thermally-produced DM particle of the same mass, enabling the νMSM to evade various astrophysical constraints on the mass of thermal sterile neutrino DM. In Chapter 5 we used the method described in Chapter 4 to constrain the νMSM parameter space. We rule out all $m_s \leq 1$ keV at 95 per cent, for all combinations of L_6 and MW halo mass. For the $m_s = 7$ keV model, which is the most likely candidate to explain the 3.55 keV line, only a narrow range of values $9 \leq L_6 < 11.2$ are permitted if the sterile neutrino composes all of the DM. Our lower limits do not rule out any values of L_6 in this range for $m_s = 7$ keV, for any MW halo mass between 0.5×10^{12} $M_\odot$ and 2×10^{12} $M_\odot$. These constraints will tighten considerably with the inclusion of galaxy formation processes and the availability of new observational constraints from advanced instruments such as XRISM, and particle collider searches.

6.3 Concluding remarks

We have come a long way in the last 120 years. At the turn of the 20$^{\text{th}}$ century, scientific understanding held that the Universe comprised solely the Milky Way. Within 30 years this view had been set aside, replaced by the discovery of the existence of other galaxies outside our own. In this context, General Relativity provided the framework to understand the observations that were to come, and its predictions have withstood ever more intense scrutiny flawlessly. Today, the field of cosmology provides an explanation for a Universe containing many billions of galaxies, of which the Milky Way is only one. We now inhabit an era of 'precision cosmology', wherein the fundamental properties of the Universe are measured ever more exquisitely. Set against the astrophysical backdrop of the discovery of

gravitational waves and the first ever 'photo' of a black hole—just two developments of the last five years—these are indeed, exciting times.

There are several milestones on the horizon. Within the next few years, both the LSST and XRISM will be fully functional. These facilities will revolutionize our observational capabilities, probing the faint end of galaxy formation in unprecedented detail and providing much-needed clarity on other observational measurements, such as the nature of the 3.55 keV X-ray line. Gravitational lensing could also play an important role, providing limits on the smallest DM halo and with it, strong constraints on viable DM models. These and other measurements will be crucial to advance our understanding of, and investigate, different cosmological models and the galaxy formation that takes place inside that framework. Within the next decade, therefore, we *might* expect an answer to what the DM is. However, perhaps more realistically, we will certainly know better what it is *not*.

APPENDIX A

Finding satellites in survey data

The response functions presented in Section 2.2 represent simple approximations to the completeness of the survey using a given search algorithm and depend simply only on the absolute $V-$band magnitude of the satellite. In reality, many variables affect the completeness function, which could influence the results produced by the method presented in Chapter 2. Many of these are discussed at great length in the original works that developed and tested search algorithms with well-characterized detection efficiencies (K08; W09). Here, we summarize these contributions and their possible effect on the results of Chapter 3.

Ultra- and hyperfaint dwarf galaxies are identified in survey data as enhancements in the projected spatial density of resolved stars relative to the density of Galactic foreground stars. This approach is subject to confusion with other objects: interloping Galactic foreground stars and distant background galaxies or galaxy clusters, in projection, can artificially enhance the number of apparent stellar overdensities in the observed field. This contamination could obscure the presence of true Galactic satellites and must be accounted for when interpreting raw survey data.

In the two works discussed above, contamination is accounted for by applying 'matched filters' to the survey data—cuts in colour and magnitude chosen using theoretical isochrones

from stellar population synthesis models. In dwarf galaxies, the stellar population is assumed to be composed mainly of old, metal-poor stars. This is not an unreasonable assumption. Theoretical models of structure formation, gas accretion, and reionization suggest a dwarf galaxy archetype of rapid, early star formation that tails off after the onset of reionization (e.g. Bullock et al., 2000; Benson et al., 2002a,b; Somerville, 2002). This model is supported by spectroscopic measurements of several ultrafaint dwarf galaxies in the local Universe, which show that they have very low metallicities, with [Fe / H] $\lesssim -2$ (Muñoz et al., 2006; Martin et al., 2007; Simon & Geha, 2007; Kirby et al., 2008). Their small stellar masses also imply that they have experienced limited supernova activity over their lifetimes, maintaining a stellar population with relatively unpolluted chemical signatures.

The location of the isochrones in colour–magnitude space depends on the distance to the object of interest, and also on the chemical composition of the stellar population. In W09, the scatter in the theoretical isochrones is included in the matched filter, which is recalculated assuming the stellar population lies at various distances from the MW. The photometric cuts derived from these remove a large fraction of the contaminating objects from the sample while preserving most 'genuine' ones that could be members of dwarf galaxies.

After the application of the matched filter, the map of projected positions of the stars is convolved with a spatial kernel which corresponds to the expected surface density profile of a dwarf galaxy—by assuming a Plummer profile—to obtain a smoothed density field. The choice of smoothing kernel can affect the detection of some overdensities as it introduces a preferred scale (at a fixed distance), biasing the algorithm toward objects of this preferred size. Objects falling at or below the scale of the smoothing kernel will be detected more efficiently, while more diffuse or distant objects may not be detectable if the number of stars inside the kernel falls below foreground levels. The physical size of the object and the distance to it are, therefore, degenerate quantities which affect the detectability response of the algorithm. Both of these quantities can be expressed in terms of M_V and the latitude of detection. K08 and W09 adopt different approaches to account for this. In the former, they

apply kernels of different sizes to identify stellar overdensities, the results of which are considered separately in the identification of dwarf galaxy or globular cluster candidates. In W09, the scale is chosen as a compromise between the angular sizes of compact/distant objects, and extended/closer objects. This means that some satellites might be missed, at the expense of being able to characterize well the efficiency of detection of objects by the algorithm. However, the improved understanding of the detection efficiency allows missed detections to be accommodated.

A.1 Characterizing survey response

Characterizing the detection efficiency when using a given search algorithm requires an understanding of the detectability of objects compared with the observational foreground. This can be studied by injecting simulated objects into stellar foreground fields that are understood well and applying the search algorithm to the composite field. Its response depends on the distance to the object, the object's luminosity, its scale length, and the Galactic latitude, b, at which it is observed. Here, we summarize the key findings of W09 and the implications of these for the results of Chapter 3.

The foreground density of stars in the SDSS footprint depends on latitude. At low latitudes close to the Galactic plane, the density of objects in the SDSS is a factor of $\sim$3 times larger than at high Galactic latitudes close to the poles. This could make the detection of faint or very diffuse satellites more difficult as the higher foreground density could obscure the stellar overdensities created by such objects. These considerations are incorporated into the characterization of the response function by W09 when they generate their foreground fields, by matching the density of resolved objects to observed fields at three different latitudes spanning the range encompassed by the SDSS footprint. The stars in the generated fields are distributed randomly, rather than according to the real data, which Willman (2003) showed could lead to a small overestimation of the detection efficiency. Satellite galaxy luminosity functions inferred using overestimated response functions—such as the result we report in Chapter 3—will, therefore, slightly *under*-predict the true number of satellites.

To generate the mock galaxies, W09 draw stars from the resolved stellar populations of the Carina, Draco, and Ursa Minor dwarf galaxies. These are a good match to the theoretical isochrones used elsewhere in their work. The stars are placed randomly within the simulated galaxy assuming a circular Plummer surface brightness profile of a specific physical scale length. However, in reality, many ultrafaint satellites are elliptical (e.g. Martin et al., 2008). Combined with the spherically-symmetric smoothing kernel, this could lead to a small overestimation of the efficiency of detection of some objects, and a more optimistic response function.

When seeding the galaxies at a given distance, W09 adjust the photometry of the stellar population to the correct distance modulus. They also add photometric scatter to account for measurement uncertainty, which is especially prevalent at faint magnitudes. In the faintest objects, stochasticity in the stellar luminosity function becomes more important, introducing stochasticity in the distances that such objects can be observed at. These effects smooth out the sharp transition in detection efficiency from $\epsilon = 1.0$ to 0.0 with increasing distance.

The final key component of an assessment of the response function of the survey is the observing strategy of the survey—in particular, the length of time certain regions of the sky have been observed for compared to others in the survey footprint. In the SDSS, certain areas of the sky have deeper exposures, enabling the detection of fainter satellites than in other SDSS fields. Therefore, in detail, the detection efficiency is not uniform across the survey footprint. However, for the purposes of the methodology presented in Chapter 2, we neglect such a detailed treatment and account for these effects by taking an average sensitivity across the survey area, corresponding to an average depth and background number of stars.

Bibliography

Aaboud M., et al., 2018, JHEP, 2018, 126

Abazajian K. N., et al., 2012, preprint (arXiv:1204.5379)

Ade P. a. R., et al., 2016, A&A, 594, A13

Adelman-McCarthy J. K., et al., 2007, ApJS, 172, 634

Adhikari R., et al., 2017, JCAP, 2017, 025

Aharonian F. A., et al., 2017, ApJL, 837, L15

Alam S., et al., 2015, ApJS, 219, 12

Alpher R. A., Herman R. C., 1948a, Phys. Rev., 74, 1737

Alpher R. A., Herman R., 1948b, Nature, 162, 774

Alpher R. A., Bethe H., Gamow G., 1948, Phys. Rev., 73, 803

Amorisco N. C., Evans N. W., 2012, MNRAS, 419, 184

Anderson M. E., Churazov E., Bregman J. N., 2015, MNRAS, 452, 3905

Angulo R. E., Hahn O., Abel T., 2013, MNRAS, 434, 3337

Arraki K. S., Klypin A., More S., Trujillo-Gomez S., 2014, MNRAS, 438, 1466

Asaka T., Shaposhnikov M., 2005, Phys. Lett. B, 620, 17

Asaka T., Blanchet S., Shaposhnikov M., 2005, Phys. Lett. B, 631, 151

Avila-Reese V., Colín P., Valenzuela O., D'Onghia E., Firmani C., 2001, ApJ, 559, 516

Babcock H. W., 1939, Lick Observatory Bulletin, 498, 41

Bahcall J. N., 1964, Phys. Rev., 12, L300

Bañados E., et al., 2018, Nature, 553, 473

Barbieri R., Nanopoulos D. V., Morchio G., Strocchi F., 1980, Phys. Lett. B, 90, 91

Battaglia G., Helmi A., Tolstoy E., Irwin M., Hill V., Jablonka P., 2008, ApJ, 681, L13

Baur J., Palanque-Delabrouille N., Yèche C., Magneville C., Viel M., 2016, JCAP, 2016, 012

Bechtol K., et al., 2015, ApJ, 807, 50

Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015, MNRAS, 447, 3402

Belokurov V., et al., 2006a, ApJ, 642, L137

Belokurov V., et al., 2006b, ApJ, 647, L111

Benítez-Llambay A., Frenk C. S., Ludlow A. D., Navarro J. F., 2019, MNRAS

Benson A. J., Lacey C. G., Baugh C. M., Cole S., Frenk C. S., 2002a, MNRAS, 333, 156

Benson A. J., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2002b, MNRAS, 333, 177

Benson A. J., et al., 2013, MNRAS, 428, 1774

de Bernardis P., et al., 2000, Nature, 404, 955

Blumenthal G. R., Faber S. M., Primack J. R., Rees M. J., 1984, Nature, 311, 517

Bode P., Ostriker J. P., Turok N., 2001, ApJ, 556, 93

Bœhm C., Schewtschenko J. A., Wilkinson R. J., Baugh C. M., Pascoli S., 2014, MNRAS, 445, L31

Boldrini P., Mohayaee R., Silk J., 2019, MNRAS, 485, 2546

Bolton J. S., Haehnelt M. G., Warren S. J., Hewett P. C., Mortlock D. J., Venemans B. P., McMahon R. G., Simpson C., 2011, MNRAS, 416, L70

Bond J. R., Szalay A. S., 1983, ApJ, 274, 443

Bond J. R., Efstathiou G., Silk J., 1980, Phys. Rev., 45, L1980

Bond J. R., Szalay A. S., Turner M. S., 1982, Phys. Rev., 48, L1636

Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440

Bondi H., Gold T., 1948, MNRAS, 108, 252

van den Bosch F. C., Ogiya G., 2018, MNRAS, 475, 4066

Bose S., Hellwing W. A., Frenk C. S., Jenkins A., Lovell M. R., Helly J. C., Li B., 2016, MNRAS, 455, 318

Bose S., et al., 2017, MNRAS, 464, 4520

Bose S., Deason A. J., Frenk C. S., 2018, ApJ, 863, 123

Bose S., et al., 2019, MNRAS, 486, 4790

Bower R. G., 1991, MNRAS, 248, 332

Boyarsky A., Ruchayskiy O., Shaposhnikov M., 2009, Annu. Rev. Nucl. Part. Sci., 59, 191

Boyarsky A., Franse J., Iakubovskyi D., Ruchayskiy O., 2014a, preprint (arXiv:1408.4388)

Boyarsky A., Ruchayskiy O., Iakubovskyi D., Franse J., 2014b, Phys. Rev., 113, L251301

Boyarsky A., Franse J., Iakubovskyi D., Ruchayskiy O., 2015, Phys. Rev., 115, L161301

Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., 2010, MNRAS, 406, 896

Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2011, MNRAS, 415, L40

Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2012, MNRAS, 422, 1203

Breddels M. A., Helmi A., van den Bosch R. C. E., van de Ven G., Battaglia G., 2013, MNRAS, 433, 3173

Brook C. B., Di Cintio A., 2015, MNRAS, 450, 3920

Brooks A. M., Zolotov A., 2014, ApJ, 786, 87

Bulbul E., Markevitch M., Foster A. R., Smith R. K., Loewenstein M., Randall S. W., 2014a, preprint (arXiv:1409.4143)

Bulbul E., Markevitch M., Foster A., Smith R. K., Loewenstein M., Randall S. W., 2014b, ApJ, 789, 13

Bulbul E., Markevitch M., Foster A., Miller E., Bautz M., Loewenstein M., Randall S. W., Smith R. K., 2016, ApJ, 831, 55

Bullock J. S., Kravtsov A. V., Weinberg D. H., 2000, ApJ, 539, 517

Callingham T. M., et al., 2019, MNRAS, 484, 5453

Canetti L., Drewes M., Shaposhnikov M., 2012, New J. Phys., 14, 095012

Canetti L., Drewes M., Frossard T., Shaposhnikov M., 2013a, Phys. Rev. D, 87, 093006

Canetti L., Drewes M., Shaposhnikov M., 2013b, Phys. Rev., 110, L061801

Cappelluti N., et al., 2018, ApJ, 854, 179

Carlin J. L., et al., 2017, AJ, 154, 267

Caruana J., Bunker A. J., Wilkins S. M., Stanway E. R., Lacy M., Jarvis M. J., Lorenzoni S., Hickey S., 2012, MNRAS, 427, 3055

Caruana J., Bunker A. J., Wilkins S. M., Stanway E. R., Lorenzoni S., Jarvis M. J., Ebert H., 2014, MNRAS, 443, 2831

Cautun M., Hellwing W. A., van de Weygaert R., Frenk C. S., Jones B. J. T., Sawala T., 2014a, MNRAS, 445, 1820

Cautun M., Frenk C. S., van de Weygaert R., Hellwing W. A., Jones B. J. T., 2014b, MNRAS, 445, 2049

Cautun M., Bose S., Frenk C. S., Guo Q., Han J., Hellwing W. A., Sawala T., Wang W., 2015, MNRAS, 452, 3838

Chan T. K., Kereš D., Oñorbe J., Hopkins P. F., Muratov A. L., Faucher-Giguère C.-A., Quataert E., 2015, MNRAS, 454, 2981

Cole S., Aragón-Salamanca A., Frenk C. S., Navarro J. F., Zepf S. E., 1994, MNRAS, 271, 781

Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168

Cole D. R., Dehnen W., Read J. I., Wilkinson M. I., 2012, MNRAS, 426, 601

Colombi S., Dodelson S., Widrow L. M., 1996, ApJ, 458, 1

Conn A. R., et al., 2013, ApJ, 766, 120

Contenta F., et al., 2018, MNRAS, 476, 3124

Cowsik R., McClelland J., 1972, Phys. Rev., 29, L669

Cowsik R., McClelland J., 1973, ApJ, 180, 7

Crain R. A., et al., 2015, MNRAS, 450, 1937

D'Onghia E., Springel V., Hernquist L., Keres D., 2010, ApJ, 709, 1138

Davies F. B., et al., 2018, ApJ, 864, 142

Davis R., 1964, Phys. Rev., 12, L303

Davis M., Peebles P. J. E., 1983, ApJ, 267, 465

Davis M., Huchra J., Latham D. W., Tonry J., 1982, ApJ, 253, 423

Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371

Davis M., Summers F. J., Schlegel D., 1992, Nature, 359, 393

Dekel A., Silk J., 1986, ApJ, 303, 39

Dodelson S., Widrow L. M., 1994, Phys. Rev., 72, L17

Doroshkevich A. G., Zeldovich Y. B., Syunyaev R. A., Khlopov M. Y., 1980, SvA, 6, L252

Drlica-Wagner A., et al., 2015, ApJ, 813, 109

Drlica-Wagner A., et al., 2016, ApJ, 833, L5

Dubinski J., Carlberg R. G., 1991, ApJ, 378, 496

Dutton A. A., Macciò A. V., Frings J., Wang L., Stinson G. S., Penzo C., Kang X., 2016, MNRAS, 457, L74

Dyson F. W., Eddington A. S., Davidson C., 1920, Philosophical Transactions of the Royal Society of London Series A, 220, 291

Efstathiou G., 1992, MNRAS, 256, 43P

Efstathiou G., Davis M., White S. D. M., Frenk C. S., 1985, The Astrophysical Journal Supplement Series, 57, 241

Efstathiou G., Sutherland W. J., Maddox S. J., 1990, Nature, 348, 705

Efstathiou G., Bond J. R., White S. D. M., 1992, MNRAS, 258, 1P

Einasto J., 1965, Trudy Inst. Astroz. Alma-Ata, 5, 87

Einasto J., Kaasik A., Saar E., 1974a, Nature, 250, 309

Einasto J., Saar E., Kaasik A., Chernin A. D., 1974b, Nature, 252, 111

Einstein A., 1915, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften, p. 844

Einstein A., 1917, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften, p. 142

Einstein A., 1930, Math. Ann., 102, 685

El-Zant A., Shlosman I., Hoffman Y., 2001, ApJ, 560, 636

Errani R., Peñarrubia J., Laporte C. F. P., Gómez F. A., 2017, MNRAS, 465, L59

Escudero M., Lopez-Honorez L., Mena O., Palomares-Ruiz S., Villanueva-Domingo P., 2018, JCAP, 2018, 007

Faber S. M., Gallagher J. S., 1979, ARA&A, 17, 135

Fan X., et al., 2006, AJ, 132, 117

Fattahi A., et al., 2016, MNRAS, 457, 844

Fitts A., et al., 2017, MNRAS, 471, 3547

Flores R. A., Primack J. R., 1994, ApJ, 427, L1

Font A. S., et al., 2011, MNRAS, 417, 1260

Franse J., et al., 2016, ApJ, 829, 124

Freeman K. C., 1970, ApJ, 160, 811

Frenk C. S., White S. D. M., 2012, Ann. Phys., 524, 507

Frenk C. S., White S. D. M., Efstathiou G., Davis M., 1985, Nature, 317, 595

Friedman A., 1922, Z. Physik, 10, 377

Fukugita M., Yanagida T., 1986, Phys. Lett. B, 174, 45

Gamow G., 1948a, Phys. Rev., 74, 505

Gamow G., 1948b, Nature, 162, 680

Garrison-Kimmel S., Boylan-Kolchin M., Bullock J. S., Lee K., 2014, MNRAS, 438, 2578

Garrison-Kimmel S., et al., 2017, MNRAS, 471, 1709

Gelato S., Sommer-Larsen J., 1999, MNRAS, 303, 321

Gell-Mann M., Ramond P., Slansky R., 1979, preprint (arXiv:1306.4669)

Genina A., et al., 2018, MNRAS, 474, 1398

Gershtein S. S., Zel'dovich Y. B., 1966a, Soviet Journal of Experimental and Theoretical Physics Letters, 4, 120

Gershtein S. S., Zel'dovich Y. B., 1966b, ZhETF Pisma Redaktsiiu, 4, 174

Ghigna S., Moore B., Governato F., Lake G., Quinn T., Stadel J., 1998, MNRAS, 300, 146

Giocoli C., Pieri L., Tormen G., 2008, MNRAS, 387, 689

Goerdt T., Moore B., Read J. I., Stadel J., Zemp M., 2006, MNRAS, 368, 1073

Governato F., et al., 2012, MNRAS, 422, 1231

Grand R. J. J., Springel V., Gómez F. A., Marinacci F., Pakmor R., Campbell D. J. R., Jenkins A., 2016, MNRAS, 459, 199

Greig B., Mesinger A., Haiman Z., Simcoe R. A., 2017, MNRAS, 466, 4239

Gu L., Kaastra J., Raassen A. J. J., Mullen P. D., Cumbee R. S., Lyons D., Stancil P. C., 2015, A&A, 584, L11

Guo Q., Cole S., Eke V., Frenk C., 2012, MNRAS, 427, 428

Guth A. H., 1981, Phys. Rev. D, 23, 347

Guth A. H., Pi S.-Y., 1982, Phys. Rev., 49, L1110

Han J., Cole S., Frenk C. S., Jing Y., 2016, MNRAS, 457, 1208

Hanany S., et al., 2000, ApJ, 545, L5

Hargis J. R., Willman B., Peter A. H. G., 2014, ApJ, 795, L13

Haslbauer M., Dabringhausen J., Kroupa P., Javanmardi B., Banik I., 2019, A&A, 626, A47

Hawking S. W., 1982, Phys. Lett. B, 115, 295

Hayashi K., Fabrizio M., Łokas E. L., Bono G., Monelli M., Dall'Ora M., Stetson P. B., 2018, MNRAS, 481, 250

Hellwing W. A., Frenk C. S., Cautun M., Bose S., Helly J., Jenkins A., Sawala T., Cytowski M., 2016, MNRAS, 457, 3492

Hernquist L., 1990, ApJ, 356, 359

Hinshaw G., et al., 2013, ApJS, 208, 19

Hofmann F., Sanders J. S., Nandra K., Clerc N., Gaspari M., 2016, A&A, 592, A112

Homma D., et al., 2016, ApJ, 832, 21

Homma D., et al., 2018, PASJ, 70

Horiuchi S., Humphrey P. J., Oñorbe J., Abazajian K. N., Kaplinghat M., Garrison-Kimmel S., 2014, Phys. Rev. D, 89, 025017

Hoyle F., 1948, MNRAS, 108, 372

Hsueh J.-W., Enzi W., Vegetti S., Auger M., Fassnacht C. D., Despali G., Koopmans L. V. E., McKean J. P., 2019, preprint (arXiv:1905.04182)

Hubble E., 1924, The New York Times

Hubble E. P., 1925, Popular Astronomy, 33, 252

Hubble E., 1929, PNAS, 15, 168

Iakubovskyi D., Bulbul E., Foster A. R., Savchenko D., Sadova V., 2015, preprint (arXiv:1508.05186)

Ibata R., Lewis G. F., Irwin M., Totten E., Quinn T., 2001, ApJ, 551, 294

Ibata R. A., et al., 2013, Nature, 493, 62

Iršič V., et al., 2017, Phys. Rev. D, 96, 023522

Irwin M. J., et al., 2007, ApJ, 656, L13

Jeltema T., Profumo S., 2015, MNRAS, 450, 2143

Jethwa P., Erkal D., Belokurov V., 2016, MNRAS, 461, 2212

Jethwa P., Erkal D., Belokurov V., 2018, MNRAS, 473, 2060

Kahn F. D., Woltjer L., 1959, ApJ, 130, 705

Kallivayalil N., van der Marel R. P., Besla G., Anderson J., Alcock C., 2013, ApJ, 764, 161

Kang X., Mao S., Gao L., Jing Y. P., 2005, A&A, 437, 383

Kapteyn J. C., 1922, ApJ, 55, 302

Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201

Kaviraj S., Darg D., Lintott C., Schawinski K., Silk J., 2012, MNRAS, 419, 70

Kelvin W. T., 1904, Baltimore Lectures on Molecular Dynamics and the Wave Theory of Light. London : C. J. Clay and sons; Baltimore, Publication agency of the Johns Hopkins university

Kennedy R., Frenk C., Cole S., Benson A., 2014, MNRAS, 442, 2487

Khachatryan V., et al., 2016, JHEP, 2016, 83

Kim D., Jerjen H., Mackey D., Costa G. S. D., Milone A. P., 2015, ApJ, 804, L44

Kim D., et al., 2016, ApJ, 833, 16

Kirby E. N., Simon J. D., Geha M., Guhathakurta P., Frebel A., 2008, ApJ, 685, L43

Kirby E. N., Bullock J. S., Boylan-Kolchin M., Kaplinghat M., Cohen J. G., 2014, MNRAS, 439, 1015

Klypin A., Kravtsov A. V., Valenzuela O., Prada F., 1999, ApJ, 522, 82

Komatsu E., et al., 2011, ApJS, 192, 18

Koposov S., et al., 2008, ApJ, 686, 279

Koposov S. E., Belokurov V., Torrealba G., Evans N. W., 2015a, ApJ, 805, 130

Koposov S. E., et al., 2015b, ApJ, 811, 62

Kowalczyk K., Łokas E. L., Kazantzidis S., Mayer L., 2013, MNRAS, 431, 2796

Kroupa P., Theis C., Boily C. M., 2005, A&A, 431, 517

Kuhlen M., Diemand J., Madau P., Zemp M., 2008, J. Phys.: Conf. Ser., 125, 012008

Kunkel W. E., Demers S., 1976, The Galaxy and the Local Group, 182, 241

Küpper A. H. W., Balbinot E., Bonaca A., Johnston K. V., Hogg D. W., Kroupa P., Santiago B. X., 2015, ApJ, 803, 80

LUX Collaboration et al., 2014, Phys. Rev., 112, L091303

Lacey C., Cole S., 1993, MNRAS, 262, 627

Lacey C. G., et al., 2016, MNRAS, 462, 3854

Laevens B. P. M., et al., 2015, ApJ, 813, 44

Laine M., Shaposhnikov M., 2008, JCAP, 2008, 031

Larson R. B., 1974, MNRAS, 169, 229

Lemaître G., 1927, Annales de la Société Scientifique de Bruxelles, 47, 49

Lemaître G., 1931, Nature, 127, 706

Leo M., Baugh C. M., Li B., Pascoli S., 2018, JCAP, 2018, 010

Li Y.-S., White S. D. M., 2008, MNRAS, 384, 1459

Li T. S., et al., 2017, ApJ, 838, 8

Libeskind N. I., Frenk C. S., Cole S., Helly J. C., Jenkins A., Navarro J. F., Power C., 2005, MNRAS, 363, 146

Libeskind N. I., Frenk C. S., Cole S., Jenkins A., Helly J. C., 2009, MNRAS, 399, 550

Linde A. D., 1982a, Phys. Lett. B, 108, 389

Linde A. D., 1982b, Phys. Lett. B, 116, 335

Lisenfeld U., Braine J., Duc P. A., Boquien M., Brinks E., Bournaud F., Lelli F., Charmandaris V., 2016, A&A, 590, A92

Liu J., Chen X., Ji X., 2017, Nature Physics, 13, 212

Lovell M. R., et al., 2012, MNRAS, 420, 2318

Lovell M. R., Frenk C. S., Eke V. R., Jenkins A., Gao L., Theuns T., 2014, MNRAS, 439, 300

Lovell M. R., et al., 2016, MNRAS, 461, 60

Lovell M. R., et al., 2017, MNRAS, 468, 4285

Lovell M. R., et al., 2019a, MNRAS, 485, 4071

Lovell M. R., Iakubovskyi D., Barnes D., Bose S., Frenk C. S., Theuns T., Hellwing W. A., 2019b, ApJ, 875, L24

Lubimov V. A., Novikov E. G., Nozik V. Z., Tretyakov E. F., Kosik V. S., 1980, Phys. Lett. B, 94, 266

Ludlow A. D., Bose S., Angulo R. E., Wang L., Hellwing W. A., Navarro J. F., Cole S., Frenk C. S., 2016, MNRAS, 460, 1214

Lynden-Bell D., 1967, MNRAS, 136, 101

Lynden-Bell D., 1976, MNRAS, 174, 695

Macciò A. V., Fontanot F., 2010, MNRAS, 404, L16

Macciò A. V., Dutton A. A., Bosch V. D., C F., Moore B., Potter D., Stadel J., 2007, MNRAS, 378, 55

Madau P., Shen S., Governato F., 2014, ApJ, 789, L17

Maddox S. J., Efstathiou G., Sutherland W. J., Loveday J., 1990, MNRAS, 242, 43P

Manohar A., 1987, Phys. Lett. B, 186, 370

Marsh D. J. E., 2016, Phys. Rep., 643, 1

Martin N. F., Ibata R. A., Chapman S. C., Irwin M., Lewis G. F., 2007, MNRAS, 380, 281

Martin N. F., de Jong J. T. A., Rix H.-W., 2008, ApJ, 684, 1075

Martin N. F., et al., 2015, ApJ, 804, L5

Martínez-Delgado D., Gómez-Flechoso M. Á., Aparicio A., Carrera R., 2004, ApJ, 601, 242

Mashchenko S., Wadsley J., Couchman H. M. P., 2008, Science, 319, 174

Mason C. A., Treu T., Dijkstra M., Mesinger A., Trenti M., Pentericci L., de Barros S., Vanzella E., 2018, ApJ, 856, 2

McConnachie A. W., 2012, AJ, 144, 4

Merritt D., Graham A. W., Moore B., Diemand J., Terzić B., 2006, AJ, 132, 2685

Mikheyev S. P., Smirnov A. Y., 1985, Yadernaya Fizika, 42, 1441

Milne E. A., 1935, Relativity, Gravitation and World-Structure, first edn. The International Series of Monographs on Physics, Oxford University Press, Oxford

Minkowski P., 1977, Phys. Lett. B, 67, 421

Mohapatra R. N., Senjanović G., 1980, Phys. Rev., 44, L912

Moore B., 1994, Nature, 370, 629

Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, ApJ, 524, L19

Müller O., Pawlowski M. S., Jerjen H., Lelli F., 2018, Science, 359, 534

Muñoz R. R., Carlin J. L., Frinchaboy P. M., Nidever D. L., Majewski S. R., Patterson R. J., 2006, ApJ, 650, L51

Munshi F., et al., 2013, ApJ, 766, 56

Nadler E. O., Gluscevic V., Boddy K. K., Wechsler R. H., 2019, ApJ, 878, L32

Navarro J. F., Frenk C. S., White S. D. M., 1995, MNRAS, 275, 720

Navarro J. F., Eke V. R., Frenk C. S., 1996a, MNRAS, 283, L72

Navarro J. F., Frenk C. S., White S. D. M., 1996b, ApJ, 462, 563

Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493

Navarro J. F., et al., 2004, MNRAS, 349, 1039

Newton O., Cautun M., 2018, MW Satellite LF: V1.0.0 Release, Zenodo, doi:10.5281/zenodo.1205622

Newton O., Cautun M., Jenkins A., Frenk C. S., Helly J. C., 2018, MNRAS, 479, 2853

Okamoto T., Eke V. R., Frenk C. S., Jenkins A., 2005, MNRAS, 363, 1299

Okamoto T., Gao L., Theuns T., 2008, MNRAS, 390, 920

Oldengott I. M., Schwarz D. J., 2017, EPL, 119, 29001

Olive K. A., Turner M. S., 1982, Phys. Rev. D, 25, 213

Olling R. P., Merrifield M. R., 2000, MNRAS, 311, 361

Oman K. A., Marasco A., Navarro J. F., Frenk C. S., Schaye J., Benítez-Llambay A., 2019, MNRAS, 482, 821

Onions J., et al., 2012, MNRAS, 423, 1200

Oñorbe J., Boylan-Kolchin M., Bullock J. S., Hopkins P. F., Kereš D., Faucher-Giguère C.-A., Quataert E., Murray N., 2015, MNRAS, 454, 2092

Oort J. H., 1932, Bulletin of the Astronomical Institutes of the Netherlands, 6, 249

Ostriker J. P., Peebles P. J. E., 1973, ApJ, 186, 467

Ostriker J. P., Peebles P. J. E., Yahil A., 1974, ApJ, 193, L1

Pagels H., Primack J. R., 1982, Phys. Rev., 48, L223

Papastergis E., Shankar F., 2016, A&A, 591, A58

Parkinson H., Cole S., Helly J., 2008, MNRAS, 383, 557

Pascale R., Posti L., Nipoti C., Binney J., 2018, MNRAS, 480, 927

Pawlowski M. S., Kroupa P., 2013, MNRAS, 435, 2116

Pawlowski M. S., Pflamm-Altenburg J., Kroupa P., 2012, MNRAS, 423, 1109

Peebles P. J. E., 1982a, ApJ, 258, 415

Peebles P. J. E., 1982b, ApJ, 263, L1

Penzias A. A., Wilson R. W., 1965, ApJ, 142, 419

Perez K., Ng K. C. Y., Beacom J. F., Hersh C., Horiuchi S., Krivonos R., 2017, Phys. Rev. D, 95, 123002

Perlmutter S., et al., 1999, ApJ, 517, 565

Piffl T., et al., 2014, A&A, 562, A91

Planck Collaboration et al., 2018, preprint (arXiv:1807.06209)

Ploeckinger S., Sharma K., Schaye J., Crain R. A., Schaller M., Barber C., 2018, MNRAS, 474, 580

Poincaré H., 1906, Popular Astronomy, 14, 475

Poincaré H., Vergne H., 1911, Leçons Sur Les Hypothèses Cosmogoniques Professées à La Sorbonne

Polisensky E., Ricotti M., 2011, Phys. Rev. D, 83, 043506

Pontzen A., Governato F., 2012, MNRAS, 421, 3464

Press W. H., Schechter P., 1974, ApJ, 187, 425

Quinn P. J., Salmon J. K., Zurek W. H., 1986, Nature, 322, 329

Read J. I., Gilmore G., 2005, MNRAS, 356, 107

Read J. I., Walker M. G., Steger P., 2018, MNRAS, 481, 860

Richardson T., Fairbairn M., 2014, MNRAS, 441, 1584

Riess A. G., et al., 1998, AJ, 116, 1009

Riess A. G., et al., 2016, ApJ, 826, 56

Ritondale E., Vegetti S., Despali G., Auger M. W., Koopmans L. V. E., McKean J. P., 2019, MNRAS, 485, 2179

Roberts M. S., 1966, ApJ, 144, 639

Roberts M. S., Rots A. H., 1973, A&A, 26, 483

Robertson B. E., Ellis R. S., Furlanetto S. R., Dunlop J. S., 2015, ApJ, 802, L19

Rogstad D. H., Shostak G. S., 1972, ApJ, 176, 315

Rubin V. C., Ford Jr. W. K., 1970, ApJ, 159, 379

Rubin V. C., Ford W. K., Thonnard N., 1980, ApJ, 238, 471

Ruchayskiy O., et al., 2016, MNRAS, 460, 1390

SNO Collaboration et al., 2001, Phys. Rev., 87, L071301

SNO Collaboration et al., 2002, Phys. Rev., 89, L011301

Safarzadeh M., Scannapieco E., Babul A., 2018, ApJ, 859, L18

Sakharov A., 1967a, Journal of Experimental and Theoretical Physics Letters, 5, 24

Sakharov A. D., 1967b, ZhETF Pisma Redaktsiiu, 5, 32

Sakharov A. D., 1991, Sov. Phys. Usp., 34, 392

Sales L. V., Navarro J. F., Lambas D. G., White S. D. M., Croton D. J., 2007, MNRAS, 382, 1901

Sales L. V., Navarro J. F., Cooper A. P., White S. D. M., Frenk C. S., Helmi A., 2011, MNRAS, 418, 648

Sánchez-Salcedo F. J., Reyes-Iturbide J., Hernandez X., 2006, MNRAS, 370, 1829

Sawala T., et al., 2015, MNRAS, 448, 2941

Sawala T., et al., 2016a, MNRAS, 456, 85

Sawala T., et al., 2016b, MNRAS, 457, 1931

Sawala T., Pihajoki P., Johansson P. H., Frenk C. S., Navarro J. F., Oman K. A., White S. D. M., 2017, MNRAS, 467, 4383

Schaye J., et al., 2015, MNRAS, 446, 521

Schewtschenko J. A., Wilkinson R. J., Baugh C. M., Bœhm C., Pascoli S., 2015, MNRAS, 449, 3587

Schmidt B. P., et al., 1998, ApJ, 507, 46

Schneider A., 2015, MNRAS, 451, 3117

Schneider A., 2016, JCAP, 2016, 059

Schneider A., Smith R. E., Reed D., 2013, MNRAS, 433, 1573

Schramm D. N., Steigman G., 1981, ApJ, 243, 1

Shafi Q., Stecker F. W., 1984, Phys. Rev., 53, L1292

Shah C., Dobrodey S., Bernitt S., Steinbrügge R., López-Urrutia J. R. C., Gu L., Kaastra J., 2016, ApJ, 833, 52

Shao S., Cautun M., Frenk C. S., Gao L., Crain R. A., Schaller M., Schaye J., Theuns T., 2016, MNRAS, 460, 3772

Shaposhnikov M., 2008, JHEP, 2008, 008

Shen S., Madau P., Conroy C., Governato F., Mayer L., 2014, ApJ, 792, 99

Shu F. H., 1978, ApJ, 225, 83

Simha V., Cole S., 2017, MNRAS, 472, 1392

Simon J. D., Geha M., 2007, ApJ, 670, 313

Sirunyan A. M., et al., 2018, JHEP, 2018, 166

Smoot G. F., et al., 1992, ApJ, 396, L1

Somerville R. S., 2002, ApJ, 572, L23

Sonbas E., Rangelov B., Kargaltsev O., Dhuga K. S., Hare J., Volkov I., 2016, ApJ, 821, 54

Spergel D. N., et al., 2003, ApJS, 148, 175

Springel V., et al., 2005, Nature, 435, 629

Springel V., et al., 2008, MNRAS, 391, 1685

Starobinsky A. A., 1982, Phys. Lett. B, 117, 175

Strigari L. E., Frenk C. S., White S. D. M., 2017, ApJ, 838, 123

Super-Kamiokande Collaboration et al., 1998, Phys. Rev., 81, L1562

Szalay A. S., Marx G., 1976, A&A, 49, 437

Tashiro M., et al., 2018, in Space Telescopes and Instrumentation 2018: Ultraviolet to Gamma Ray. International Society for Optics and Photonics, p. 1069922, doi:10.1117/12.2309455

Tollerud E. J., Bullock J. S., Strigari L. E., Willman B., 2008, ApJ, 688, 277

Tollerud E. J., Boylan-Kolchin M., Bullock J. S., 2014, MNRAS, 440, 3511

Tollet E., et al., 2016, MNRAS, 456, 3542

Tomozeiu M., Mayer L., Quinn T., 2016, ApJ, 827, L15

Torrealba G., Koposov S. E., Belokurov V., Irwin M., 2016a, MNRAS, 459, 2370

Torrealba G., et al., 2016b, MNRAS, 463, 712

Torrealba G., et al., 2018, MNRAS, 475, 5085

Tully R. B., Libeskind N. I., Karachentsev I. D., Karachentseva V. E., Rizzi L., Shaya E. J., 2015, ApJ, 802, L25

Urban O., Werner N., Allen S. W., Simionescu A., Kaastra J. S., Strigari L. E., 2015, MNRAS, 451, 2447

Vegetti S., Despali G., Lovell M. R., Enzi W., 2018, MNRAS, 481, 3661

Vera-Ciro C. A., Helmi A., Starkenburg E., Breddels M. A., 2013, MNRAS, 428, 1696

Verbeke R., Papastergis E., Ponomareva A. A., Rathi S., Rijcke S. D., 2017, A&A, 607, A13

Viel M., Lesgourgues J., Haehnelt M. G., Matarrese S., Riotto A., 2005, Phys. Rev. D, 71, 063534

Viel M., Becker G. D., Bolton J. S., Haehnelt M. G., 2013, Phys. Rev. D, 88, 043502

Walker M. G., Peñarrubia J., 2011, ApJ, 742, 20

Walker M. G., et al., 2016, ApJ, 819, 53

Walsh S. M., Jerjen H., Willman B., 2007, ApJ, 662, L83

Walsh S. M., Willman B., Jerjen H., 2009, AJ, 137, 450

Wang J., White S. D. M., 2007, MNRAS, 380, 93

Wang W., White S. D. M., 2012, MNRAS, 424, 2574

Wang J., Frenk C. S., Navarro J. F., Gao L., Sawala T., 2012, MNRAS, 424, 2715

Wang J., Frenk C. S., Cooper A. P., 2013, MNRAS, 429, 1502

Wang W., Han J., Cooper A. P., Cole S., Frenk C., Lowing B., 2015, MNRAS, 453, 377

Warren M. S., Quinn P. J., Salmon J. K., Zurek W. H., 1992, ApJ, 399, 405

Watkins L. L., et al., 2009, MNRAS, 398, 1757

Watkins L. L., Evans N. W., An J. H., 2010, MNRAS, 406, 264

Watson C. R., Li Z., Polley N. K., 2012, JCAP, 2012, 018

Weinberg M. D., Katz N., 2002, ApJ, 580, 627

Weinberg D. H., Bullock J. S., Governato F., de Naray R. K., Peter A. H. G., 2015, PNAS, 112, 12249

Wetzel A. R., Hopkins P. F., Kim J.-h., Faucher-Giguère C.-A., Kereš D., Quataert E., 2016, ApJ, 827, L23

Wheeler C., Oñorbe J., Bullock J. S., Boylan-Kolchin M., Elbert O. D., Garrison-Kimmel S., Hopkins P. F., Kereš D., 2015, MNRAS, 453, 1305

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

White S. D. M., Frenk C. S., Davis M., 1983, ApJ, 274, L1

White S. D. M., Navarro J. F., Evrard A. E., Frenk C. S., 1993, Nature, 366, 429

Willman B., 2003, PhD thesis, University of Washington

Willman B., et al., 2005, ApJ, 626, L85

Wolfenstein L., 1978, Phys. Rev. D, 17, 2369

Wright E. L., et al., 1992, ApJ, 396, L13

XENON Collaboration et al., 2017, Phys. Rev., 119, L181301

XENON Collaboration et al., 2018, Phys. Rev., 121, L111302

Yanagida T., 1980, Prog. Theor. Phys., 64, 1103

Zentner A. R., Kravtsov A. V., Gnedin O. Y., Klypin A. A., 2005, ApJ, 629, 219

Zucker D. B., et al., 2006a, ApJ, 643, L103

Zucker D. B., et al., 2006b, ApJ, 650, L41

Zwicky F., 1929, PNAS, 15, 773

Zwicky F., 1933, Helvetica Physica Acta, 6, 110

Zwicky F., 1937, ApJ, 86, 217

Index

H

I

L

M

N

P

R

S

T

V

W