

# Unravelling the mass spectrum of destroyed dwarf galaxies with the metallicity distribution function

Alis J. Deason<sup>1,2★</sup>, Sergey E. Koposov<sup>3,4,5★</sup>, Azadeh Fattahi<sup>6,1</sup> and Robert J. J. Grand<sup>6,7</sup>

<sup>1</sup>*Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*

<sup>2</sup>*Centre for Extragalactic Astronomy, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*

<sup>3</sup>*Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK*

<sup>4</sup>*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*

<sup>5</sup>*Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*

<sup>6</sup>*Instituto de Astrofísica de Canarias, Calle Via Lactea s/n, La Laguna, Tenerife E-38205, Spain*

<sup>7</sup>*Departamento de Astrofísica, Universidad de La Laguna, Av. del Astrofísico Francisco Sanchez s/n, La Laguna, Tenerife E-38206, Spain*

Accepted 2023 February 14. Received 2023 February 14; in original form 2023 January 11

## ABSTRACT

Accreted stellar populations are comprised of the remnants of destroyed galaxies, and often dominate the ‘stellar haloes’ of galaxies such as the Milky Way (MW). This ensemble of external contributors is a key indicator of the past assembly history of a galaxy. We introduce a novel statistical method that uses the unbinned metallicity distribution function (MDF) of a stellar population to estimate the mass spectrum of its progenitors. Our model makes use of the well-known mass–metallicity relation of galaxies and assumes Gaussian MDF distributions for individual progenitors: the overall MDF is thus a mixture of MDFs from smaller galaxies. We apply the method to the stellar halo of the MW, as well as the classical MW satellite galaxies. The stellar components of the satellite galaxies have relatively small sample sizes, but we do not find any evidence for accreted populations with  $L > L_{\text{host}}/100$ . We find that the MW stellar halo has  $N \sim 1\text{--}3$  massive progenitors ( $L \gtrsim 10^8 L_{\odot}$ ) within 10 kpc, and likely several hundred progenitors in total. We also test our method on simulations of MW-mass haloes, and find that our method is able to recover the true accreted population within a factor of 2. Future data sets will provide MDFs with orders of magnitude more stars, and this method could be a powerful technique to quantify the accreted populations down to the ultra-faint dwarf mass scale for both the MW and its satellites.

**Key words:** Galaxies: dwarf – Galaxy: halo – Local Group – galaxies: luminosity function.

## 1 INTRODUCTION

Dark matter haloes of all shapes and sizes grow by accumulating lower mass constituents (or subhaloes). The galaxies at the centres of these haloes grow via ongoing star formation, but can also form diffuse ‘stellar haloes’ from the stellar material deposited by the accretion of subhaloes (if they contain stars). Depending on the mass scale, this accreted stellar material can amount to significant (e.g. clusters,  $\sim 20\text{--}30$  per cent) or minuscule (e.g. dwarfs,  $\sim 0\text{--}5$  per cent) fractions of the overall stellar mass of the central galaxy (Purcell, Bullock & Zentner 2007). Despite having a relatively low stellar mass and surface brightness, stellar haloes retain a record of the lower mass systems that have been digested by haloes over time, and quantifying and understanding this accreted relic has been a major research focus in astronomy for several decades (see e.g. Helmi 2008; Belokurov 2013).

The most-studied stellar halo is, unsurprisingly, that of our own Milky Way (MW) galaxy. However, despite significant progress in recent years, we still only have a qualitative view of the mass

spectrum of dwarf galaxies that have been consumed by the MW. Most notably, it has become clear since the game-changing *Gaia* mission (Gaia Collaboration 2016) that the inner stellar halo (within  $\sim 20$  kpc) is dominated by one ancient, massive accretion event, dubbed the *Gaia*-Enceladus-Sausage (GES; Belokurov et al. 2018; Helmi et al. 2018). There is also some evidence that an additional massive structure resides in the very central regions of the galaxy (within  $\sim 4$  kpc), and was accreted even earlier than the GES (Kruijssen et al. 2019; Horta et al. 2021a). However, it is debated whether or not this is really an accreted structure, or rather *in situ* MW material (see e.g. Myeong et al. 2022; Rix et al. 2022). These massive progenitors, join the already discovered streams and substructures, such as the Sagittarius and Orphan streams (e.g. Newberg et al. 2003; Majewski et al. 2004; Belokurov et al. 2007b), and the Virgo (Jurić et al. 2008) and Hercules-Aquila (Belokurov et al. 2007a) clouds (although the latter structures may be related to the GES, see e.g. Simion, Belokurov & Koposov 2019; Chandra et al. 2022), and more stellar structures in the halo are continuously being discovered (e.g. Naidu et al. 2020). The overall inventory of the Galactic stellar halo is evolving, but the picture is far from complete, and we have no quantitative ‘mass-spectrum’ of destroyed dwarfs akin to the surviving satellite dwarf luminosity function (Koposov et al. 2008;

\* E-mail: [alis.j.deason@durham.ac.uk](mailto:alis.j.deason@durham.ac.uk) (AJD); [sergey.koposov@ed.ac.uk](mailto:sergey.koposov@ed.ac.uk) (SEK)

Tollerud et al. 2008; Drlica-Wagner et al. 2020; Nadler et al. 2020), which is a pillar of the field.

Many of the halo structures that have been discovered in the MW are identified in phase-space and/or action-angle space. This, of course, is where an astrometric mission such as *Gaia* has enabled a deeper understanding of the phase-space structure of the halo by providing 6D measurements (at least for the inner halo). However, even with perfect 6D data, robustly identifying *distinct* halo substructures is challenging. Indeed, massive progenitors can have several ‘clumps’ in dynamical spaces which cannot be unambiguously disentangled (e.g. Callingham et al. 2022) and when the stellar material is fully phase-mixed it becomes more difficult to identify from the background (e.g. Johnston et al. 2008). Furthermore, even in the space of conserved quantities the clumps may not stay compact due to perturbations from massive systems such as the Large Magellanic Cloud (LMC, Koposov et al. 2022b). This is where chemical information can be crucial, as galaxies of different mass (and star formation history) can have distinct chemical signatures (e.g. Venn et al. 2004; Tolstoy, Hill & Tosi 2009). Most notable, is the well-known mass–metallicity relation for galaxies, which extends down to the dwarf mass scales (e.g. Skillman, Kennicutt & Hodge 1989; Kirby et al. 2011).

More massive galaxies are, on average, more metal-rich, and the relation between mass and metallicity exists over several orders of magnitude in mass (e.g. Tremonti et al. 2004; Kirby et al. 2013). This relation can, to first order, be explained by the larger gravitational wells of more massive galaxies, which are able to retain metals (Dekel & Silk 1986). Lower mass galaxies lack the gravity to resist the expulsion of metals due to feedback mechanisms. On the dwarf mass scale, not only does the average metallicity vary with mass, but the width of the metallicity distribution function (MDF) also varies, with the lowest mass dwarfs having a wider spread of metallicities (e.g. Kirby et al. 2011). The combined MDF of a population of accreted dwarf galaxies, such as a stellar halo, is therefore the superposition of several individual MDFs. Thus, in principle, metallicity measurements *alone* contain a unique record of the mass spectrum of accreted dwarfs. Indeed, the disentangling of an MDF into its individual components is the main focus of this work. Finally, it is worth noting that previous work on the MDFs of dwarf galaxies has focused on *surviving* dwarfs, which, depending on the largely unknown redshift evolution of the mass–metallicity relation, may or may not be relevant for the destroyed dwarfs that make-up stellar haloes (see e.g. Fattahi et al. 2020; Naidu et al. 2022).

In this work, we consider Galactic-sized stellar haloes as well as the (potential) stellar haloes of dwarf galaxies. In principle, dwarf galaxies themselves can cannibalize lower mass dwarfs, and form what we classically think of as a ‘stellar halo’. However, unlike larger mass scales where the merging dark matter clumps all contain stars, at lower mass scales (below  $\sim 10^9 M_\odot$  in halo mass) dark matter subhaloes may not have any stars at all (e.g. Benítez-Llambay & Frenk 2020). A recent study by Deason et al. (2022) showed that the very existence of a stellar halo around a dwarf galaxy can have important implications for both small-scale galaxy formation and the nature of dark matter. For example, the mass-threshold for galaxy formation, which is largely determined by the epoch of reionization, can have a major effect on the stellar haloes of dwarf galaxies: for models with a high mass threshold for galaxy formation ( $\gtrsim 10^9 M_\odot$ ) dwarf galaxies should not have stellar haloes at all! Thus, the detection or non-detection of lower mass accretion events surrounding dwarf galaxies, particularly at the ultra-faint mass scale ( $M_{\text{star}} \lesssim 10^5 M_\odot$ ), is of utmost importance.

In order to study the MDFs of accreted populations, we need large, ideally unbiased, spectroscopic samples with metallicity measurements. For both the Galactic halo, and dwarf satellite galaxies in the MW, extensive samples are hard to come by, but there has been significant progress in recent years (e.g. Walker et al. 2007; Kirby et al. 2011; Zhao et al. 2012; Kunder et al. 2017; Majewski et al. 2017; Conroy et al. 2019; Taibi et al. 2022). Moreover, and importantly, we are entering a new era of spectroscopic surveys in the MW, with several projects such as DESI, WEAVE, 4MOST, and PFS on the horizon (Takada et al. 2014; de Jong et al. 2019; Cooper et al. 2022; Jin et al. 2022). Thus, with these new surveys in mind, we develop a new modelling method to extract the mass spectrum of accreted components from a sample of [Fe/H] measurements and apply this to current data sets.

The paper is organized as follows. In Section 2, we outline our methodology and introduce the statistical model. This is a fairly technical section that some readers may want to skip over! The method is applied to spectroscopic samples of classical dwarf satellite galaxies, and Galactic halo data in Section 3. We test the method on state-of-the-art cosmological simulations of MW-mass galaxies in Section 4, and discuss caveats and future prospects in Section 5. Finally, we summarize our main findings in Section 6.

## 2 MDF MODELLING

In this section, we present the methodology that allows us to take samples with measured [Fe/H], and some estimate of the total luminosity of the system, and use them to provide constraints on the number of discrete stellar systems of different luminosities that can contribute to a given galaxy.

The next section is fairly technical, so a less statistically minded reader may want to skip it and continue with Section 3. The PYTHON code implementing the inference method presented in this section is released on GitHub.<sup>1</sup>

### 2.1 General statistical model

We construct a generative model that allows us to represent the MDF as a mixture of MDFs from smaller galaxies. Throughout this work, we will assume that the MDF of each smaller galaxy can be represented by a Gaussian.

The generic model, where the sample of stars for the MDF is coming from several galaxies, can be described with these model parameters:

- (i) Number of galaxies  $N$
- (ii)  $L_i$  individual galaxy luminosities (where  $1 < i < N$ )
- (iii)  $\mu_i$  mean galaxy metallicities
- (iv)  $\sigma_i$  widths of MDF of individual galaxies.

We can then assume that the number of stars in the sample scales linearly with galaxy luminosity. This assumption is accurate for stellar populations of similar ages. For that assumption to hold, our sample must not be biased towards one galaxy or another (e.g. if our sample comes from a small volume that has an unrepresentative subsample of certain galaxies). If the proportionality holds, one can write the MDF as

$$P(z|N, \{L_i\}, \{\mu_i\}, \{\sigma_i\}) = \frac{1}{\sum L_i} \sum_{i=1}^{i=N} L_i \mathcal{N}(z|\mu_i, \sigma_i). \quad (1)$$

<sup>1</sup>[https://github.com/segasai/mdf\\_modeling\\_paper](https://github.com/segasai/mdf_modeling_paper)

Here, for clarity, we use  $z$  as a short-hand notation of  $[\text{Fe}/\text{H}]$ . Given our expectation that galaxy luminosities and metallicities are correlated (Tremonti et al. 2004; Kirby et al. 2011), we can assume that galaxies follow a mass–metallicity relation (or luminosity–metallicity relation):

$$\mu_i \sim \mathcal{N}(A + B \log L_i | S), \quad (2)$$

where  $A$  and  $B$  are constants i.e. taken from the mass–metallicity relation presented in Kirby et al. (2011) and Simon (2019).  $S$  is a constant representing a scatter in the relation (found to be 0.15 dex by Simon 2019 for MW satellites).

The individual widths  $\sigma_i$  of MDFs differ from galaxy to galaxy but have been approximated to be slowly dependent on the galaxy luminosity  $\sigma = C + D \log L$  (see Simon 2019). If we specify the constants  $A$ ,  $B$ ,  $C$ ,  $D$ , and  $S$  we have a model for the distribution of metallicities, and this model has an integer parameter  $N$  and  $2N$  floating point parameters for luminosities and metallicities of  $N$  individual galaxies.

While this model for the MDF is valid and can be applied to real data, it has the problem of having a variable number of parameters and therefore is difficult to sample in practice (i.e. Green 1995). Therefore, it would be beneficial to reformulate the model in a way that makes the number of parameters fixed.

The first modification we can do is to group galaxies in  $M$  luminosity bins, so that rather than representing their luminosities by discrete parameters we represent the number of galaxies in certain luminosity bins. Now, we define

- (i)  $\hat{L}_j$  are the grid of galaxy luminosities  $1 \leq j \leq M$ .
- (ii)  $N_j$  are the numbers of galaxies with luminosities  $\hat{L}_j$ .
- (iii)  $\mu_{j,k}$  are mean metallicities of  $k$ th galaxy with luminosity  $\hat{L}_j$ .  $1 < k < N_j$

Where due to mass–metallicity relation

$$\mu_{j,k} \sim \mathcal{N}(A + B \log \hat{L}_j | S),$$

or

$$\mu_{j,k} = A + B \log \hat{L}_j + S \epsilon_{j,k},$$

where  $\epsilon_{j,k} \sim \mathcal{N}(0, 1)$ . Here,  $S$  could either be a constant or a deterministic function of  $\hat{L}_j$

The MDF model is now

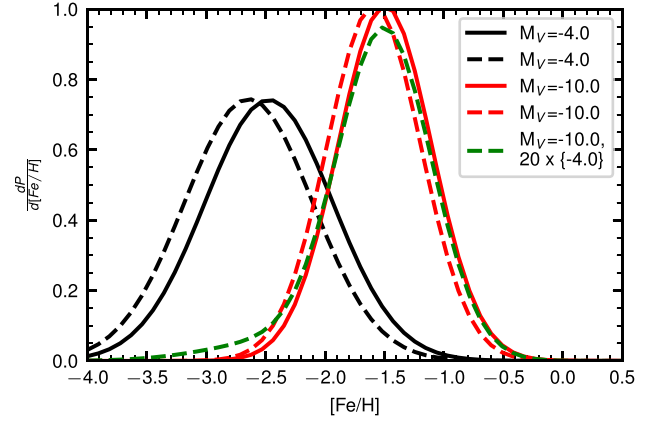
$$P(z | \{N_j\}, \{\epsilon_{j,k}\}) = \frac{1}{\sum_j N_j \hat{L}_j} \sum_{j=1}^M \hat{L}_j \left[ \sum_{k=1}^{N_j} \mathcal{N}(z | \mu_{j,k}, \sigma_{j,k}) \right].$$

The likelihood of the data consisting of (for simplicity) a single star with metallicity  $z$  would be exactly  $P(z | \{N_j\}, \{\epsilon_{j,k}\})$ . The only problem with this formulation is that this likelihood still depends on a variable number of parameters  $\epsilon_{j,k}$ , so one would prefer to marginalize over these.

$$P(z | \{N_j\}) = \int P(z | \{N_j\}, \{\epsilon_{j,k}\}) \mathcal{N}(\{\epsilon_{j,k}\} | 0, 1) d\epsilon_{j,k}$$

While this marginalization is difficult, and may be impossible to do analytically, one can simply perform a Monte-Carlo integration over  $Q$  samples from a normal distribution, where  $\epsilon_{j,k,q}$  are the  $q$ th sample  $1 \leq q \leq Q$  from  $\mathcal{N}(0, 1)$

$$P(z | \{N_j\}) \approx \frac{1}{Q} \sum_{q=1}^Q P(z | \{N_j\}, \{\epsilon_{j,k,q}\}).$$



**Figure 1.** The simulated MDFs for a few systems of different luminosities. The black lines show the expected MDFs in our model for a system with  $M_V = -4$ , with solid and dashed curves showing the MDFs when using a different random seed that controls the offset of the galaxy with respect to the mass–metallicity relation. Red curves similarly show the MDF of a single  $M_V = -10$  galaxy with different random seeds. The green curve shows the MDF for a synthetic galaxy that consists of stars coming from one galaxy with  $M_V = -10$  and 20 galaxies with  $M_V = -4$ .

Finally, instead of directly doing the summation we can simply treat this as likelihood with integer parameter  $q$ :

$$P(z | \{N_j\}, q) = P(z | \{N_j\}, \{\epsilon_{j,k,q}\}), \quad (3)$$

where  $q$  is a nuisance seed parameter that we marginalize over under the uniform prior  $U(1, Q)$ . Here, we assume that  $\epsilon_{j,k,q}$  are coming from a pseudo-random number generator that is seeded by  $q$  and provides normally distributed samples. We then will need to sample the posterior over  $\{N_j\}$  and  $q$ , which gives the model with  $M + 1$  parameters. Armed with equation (3) that specifies the likelihood function for the metallicity distribution, the only missing ingredient for the model are the priors.

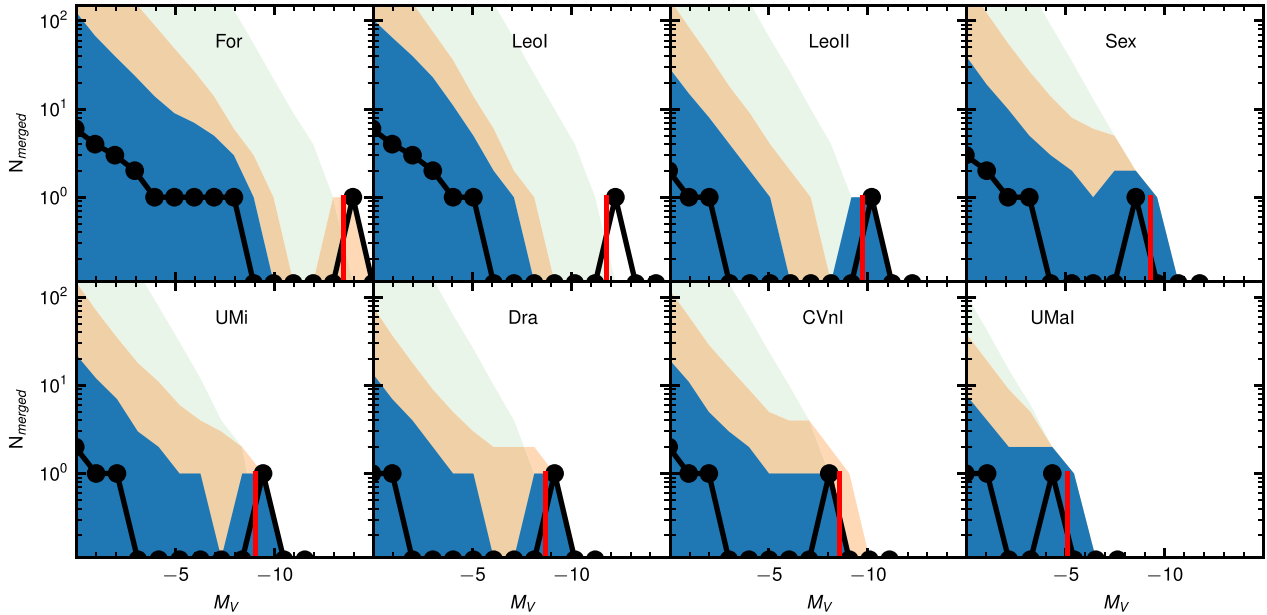
We assume that occupation numbers  $\{N_j\}$  (i.e. numbers of galaxies in luminosity bins) have a prior distribution of  $\lfloor 10^x \rfloor$ , where  $x \sim U(-1, 4)$ . This is essentially the log uniform of integers distribution with 20 per cent prior volume at  $N_i = 0$  and 20 per cent for  $1 \leq N_i \leq 10$ , and  $10 \leq N_i \leq 100$ , etc.

Finally, we complement the model with the constraint on the total luminosity of the system. Specifically, we require that the combined luminosity of multiple galaxies must match certain known total luminosity  $\log L_{\text{tot}}$  with some uncertainty  $\sigma_L$ . This provides a term for the log of the posterior:

$$\log(\sum N_i \hat{L}_i) \sim \mathcal{N}(\log L_{\text{tot}} | \sigma_L).$$

A final remark that despite the introduction of the formalism based on binned number of galaxies, we have found the model is more stable when at least one contributor to the MDF (likely the one being the most massive main progenitor) is represented directly (rather than in a bin) by the satellite luminosity  $L_{\text{main}}$ , metallicity  $z_{\text{main}}$ , and that also obeys the mass–metallicity relation.

To illustrate our modelling approach, in Fig. 1 we show the expected  $[\text{Fe}/\text{H}]$  distributions given our model. Specifically, solid black and red curves show possible MDFs for a single galaxy of  $M_V = -4$  and  $M_V = -10$ , respectively. Dashed lines of the same colours show the MDFs when different random seeds are used. The green curve shows a distribution that we might expect if we observe stars coming from a single  $M_V = -10$  galaxy and 20  $M_V = -4$



**Figure 2.** The inferred contributions from systems to the MDF of different dwarf galaxies from our analysis. In each panel, the black curve shows the median number of galaxies of a given luminosity that could have contributed to the MDF. The blue and orange bands show the 16/84 and 1/99 percentiles, respectively. The green band shows the sampling of the prior with only the constraint on total luminosity of the system. The vertical red line on each panel shows the luminosity of each system. Note that the logarithmic y-axis is cut-off at  $N_{\text{merged}} = 10^{-1}$ , so median values at this level are consistent with zero.

systems. This shows a prominent tail towards low metallicities, and this is exactly what allows us to probe the number of possible mergers with low luminosity systems.

## 2.2 Sampling

In the previous section, we have introduced the likelihood function for the metallicity distribution that is conditional on the number of different dwarf galaxies  $N_j$  on a grid of luminosities. The model also has an integer seed parameter  $q$ . It is not trivial to sample integer parameters, especially if we expect multiple modes. To perform the sampling, we decide to use the dynamic nested sampling as implemented in the DYNESTY package (Speagle 2020; Kposov et al. 2022a). As nested sampling is technically invalid if the likelihood surface has plateaus (Fowlie, Handley & Su 2020), we add a small level of deterministic noise with standard deviation of 0.01 to the likelihoods, which should not affect the inference.<sup>2</sup>

## 3 APPLICATIONS

We now apply the method described above to observational data. Here, we focus on the classical MW satellites (Section 3.1) and the MW stellar halo (Section 3.2).

### 3.1 Classical dwarf satellite galaxies

We start from the homogeneous sample of dwarf galaxy members presented in Kirby et al. (2011) as provided in the Strasbourg astronomical Data Center (CDS). As mentioned in the previous section, the key assumption that we rely on for our method is that the

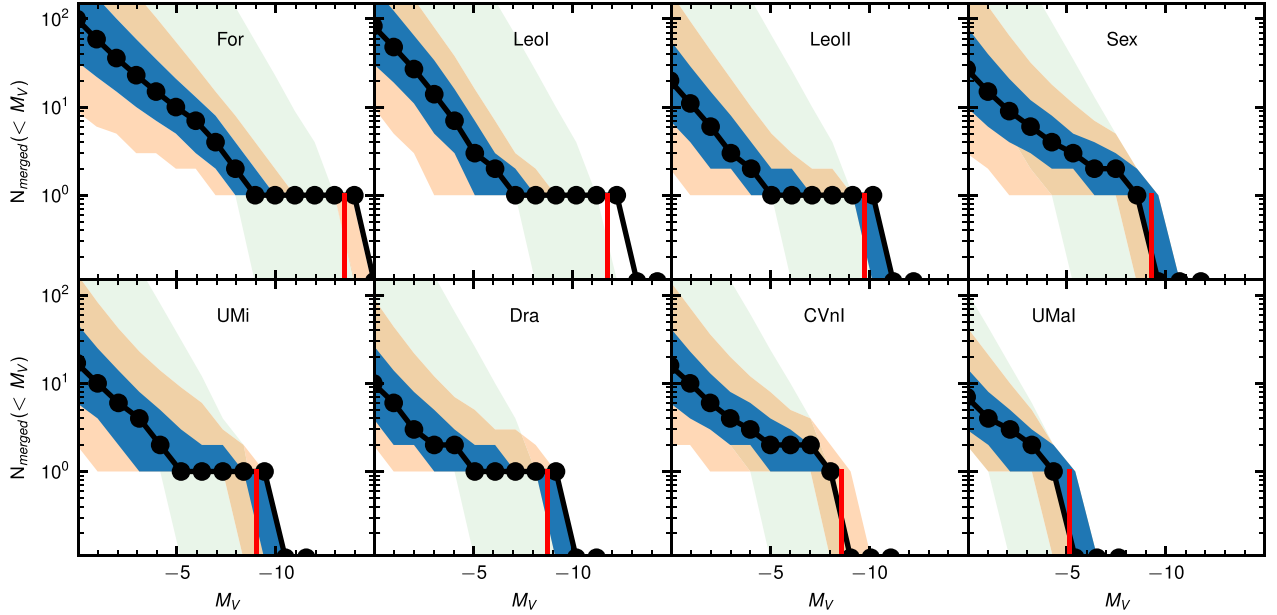
abundances that we model are random samples from the system. This is likely not technically correct for the data at hand since the stellar samples in dwarfs tend to be biased towards the centres of systems (see e.g. Walker & Peñarrubia 2011), and may have slight metallicity biases caused by the colour–magnitude selection of spectroscopic targets. We will, however, proceed ignoring these issues.

We take the sample of stars from Kirby et al. (2011) and only consider stars with small metallicity uncertainty  $\sigma_{[\text{Fe}/\text{H}]} < 0.2$ . This catalogue has measurements of 10 MW satellites with more than 10 stars: Canes Venatici I, Draco, Fornax, Hercules, Leo I, Leo II, Sculptor, Sextans, Ursa Minor, and Ursa Major I. We then proceed to model each of the dwarfs with the machinery presented in Section 2. We take the luminosities of each system from McConnachie (2012) (using an updated catalogue from January 2021) and adopt an  $M_V$  uncertainty for each system of 0.1 mag. For each system, we use the luminosity bins that are 1 mag wide from  $M_V = 0$  to the luminosity of the dwarf itself minus 2.5 mag.

The posterior samples on the number of possible dwarf galaxies that contributed to the systems’ MDF are shown in Figs 2 and 3. We show measurements for 8 out of 10 systems spanning the luminosity range from  $M_V \sim -5$  for Ursa Major I to  $M_V \sim -13$  for Fornax. The panels are ordered by system luminosity. The total number of stars varies from  $N = 15$  for Ursa Major I to  $N = 789$  for Leo I. Fig. 2 shows the constraints on the differential number of systems that have contributed to the dwarfs’ MDF, while Fig. 3 shows constraints on the cumulative counts of the number of systems brighter than a certain value. The blue/orange bands show the 16/84 and 1/99 percentiles, and the black line shows the median of the posterior. The green bands show the constraints if we do not use metallicities at all. This is essentially a prior and corresponds to the case where the only constraint comes from ensuring the combination of galaxies matches the total luminosity of the system. Note that, because we include all the stars in the galaxy, we expect to measure

<sup>2</sup>The recently released 2.1.0 version of *dynesty* lifts the limitation and is now able to sample likelihood functions with plateaus.





**Figure 3.** The inferred contributions to the MDF from our analysis. This is similar to Fig. 2 but shows the cumulative numbers. Each panel shows a different dwarf galaxy. In each panel, the black curve shows the median number of galaxies of a given luminosity or brighter that could have contributed to the MDF. The blue and orange bands show the 16/84 and 1/99 percentiles, respectively. The green band shows the sampling of the prior with only the constraint on total luminosity of the system. The vertical red line on each panel shows the luminosity of each system.

$N_{\text{merged}} = 1$  at around the total luminosity of the dwarf galaxy (shown with the solid red line). Although technically this is an ‘*in situ*’ rather than an accreted component, what we are actually constraining are the contributors to the MDF, regardless of their origin.

We now look at the posteriors in more detail. First, we focus on clear cases where the data is particularly constraining. These are the cases of Fornax, Leo I, Leo II, and Draco, where the spectroscopic samples have hundreds of members. We see that their differential posterior distributions (Fig. 2) have a peak with a value of one next to the system luminosity (highlighted in red) and show the value consistent with zero for  $M_{V,\text{host}} \lesssim M_V \lesssim M_{V,\text{host}} + 5$ . Thus, the data suggests that these systems did not experience a merger with a dwarf that is larger than 1 per cent of the system luminosity. This is also seen in the cumulative plots, where we see the implied number  $N_{\text{merged}}(< M_V)$  is flat and equal to one in the range  $M_{V,\text{host}} \lesssim M_V \lesssim M_{V,\text{host}} + 5$ . Looking at the implications for the number of faint contributors to the MDF for Fornax, Leo I, Leo II, and Draco systems, we can see that our constraints on  $N_{\text{merged}}$  shoot up and become significantly broader. The differential counts are essentially unconstrained. For example, for the Fornax MDF contributors at  $M_V = 0$  (top left-hand panel of Fig. 2) the  $1\sigma$  confidence interval is  $0 < N_{\text{merged}} < 100$  as the data allows many faint dwarfs before the observed MDF is affected significantly. The behaviour of the cumulative counts for the faint MDF contributors is somewhat misleading as it rises at faint  $M_V$  purely because we are summing over bins with non-negative values.

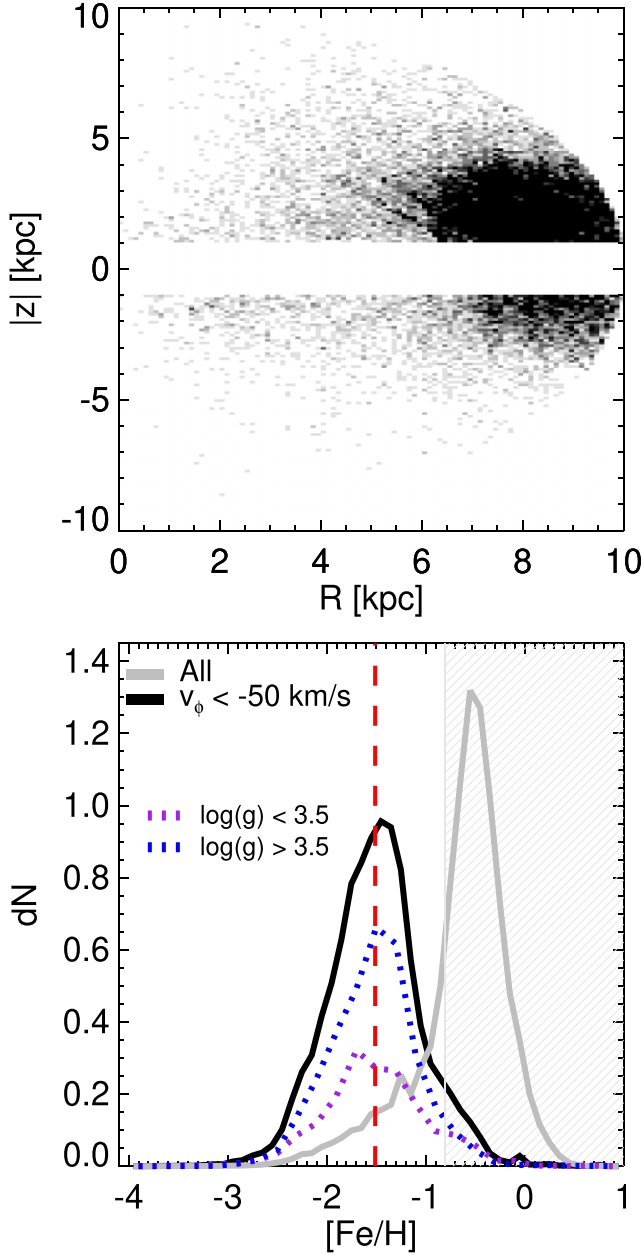
Fainter dwarf galaxies like CVnI or UMa have a smaller number of spectroscopic observations. In Fig. 2, we see that the posteriors on the number of MDF contributors start to rise next to  $M_V = M_{V,\text{host}}$ , which indicates that we cannot even rule out that the galaxy is a product of a merger of two systems with similar luminosities. The constraints on the cumulative number of mergers for fainter dwarfs do not show a flat  $N_{\text{merged}} = 1$  part next to the luminosity

of the system and instead rises to faint luminosities. We also see that for faint systems, the posteriors basically look very close to priors.

### 3.2 Galactic stellar halo

We next apply our method to the Galactic stellar halo. It has been realized for some time that the stellar halo of the MW comprises an assortment of destroyed dwarf debris, and thus the metallicity distribution of these halo stars retains a memory of their dwarf galaxy progenitors.

Large, homogeneous samples of halo stars with metallicity measurements are hard to come by, and this is a significant limitation of our current study. At present, we build a sample of halo stars based on several spectroscopic surveys and use the latest *Gaia* data (Gaia Collaboration 2021), to help select a clean halo sample. We begin by cross-matching stars with spectroscopic data from SDSS (Abolfathi et al. 2018), RAVE (Kunder et al. 2017), LAMOST (Zhao et al. 2012), APOGEE (Majewski et al. 2017), and GALAH (Buder et al. 2021) with *Gaia* DR3. This results in  $N = 6,560,819$  stars. To estimate distances to the stars, we use the Bailer-Jones et al. (2021) photogeometric distances computed from *Gaia* EDR3. We only consider stars with reasonable parallax  $\sigma_\pi/\pi < 0.5$ , and restrict our sample to  $r < 10$  kpc and  $|z| > 1$  kpc. Finally, to avoid disk contamination, we apply a cut on the rotational velocity of the stars. We impose a fairly strict cut to remove the majority of thick disk and/or *splash* stars (Belokurov et al. 2020), and only include those with retrograde orbits  $v_\phi < -50$  km s $^{-1}$ . The resulting spatial (top-panel) and metallicity distribution (bottom-panel) of the stars are shown in Fig. 4. In the bottom panel, we also show the MDF for the stars without the  $v_\phi$  cut in grey. Our restriction to retrograde orbits is fairly stringent but, as can be seen in the figure, it is effective at removing disk stars, which have prograde orbits and are generally



**Figure 4.** Top panel: The spatial distribution in the  $z$  vs.  $R$  plane of our MW halo sample. Bottom panel: The metallicity distribution of the sample. The grey line indicates the MDF without a cut in  $v_\phi$ , which leads to the inclusion of (metal-rich) disk stars. The dashed red line indicates the median metallicity of our halo sample ( $[\text{Fe}/\text{H}] = -1.5$ ). We also show the MDFs of our halo sample split by  $\log(g)$  with the blue and purple dotted lines, respectively. The gray line-filled region indicates the metal-rich regime ( $[\text{Fe}/\text{H}] > -0.8$ ) that is excluded in our modelling.

more metal-rich. We apply our modelling procedure to stars with  $-4 < [\text{Fe}/\text{H}] < -0.8^3$ , and  $\sigma([\text{Fe}/\text{H}]) < 0.2$ , which results in a sample of  $N = 21\,813$  stars.

Our sample is comprised of five different spectroscopic surveys, with varying selection functions. Here, we aim to maximize the

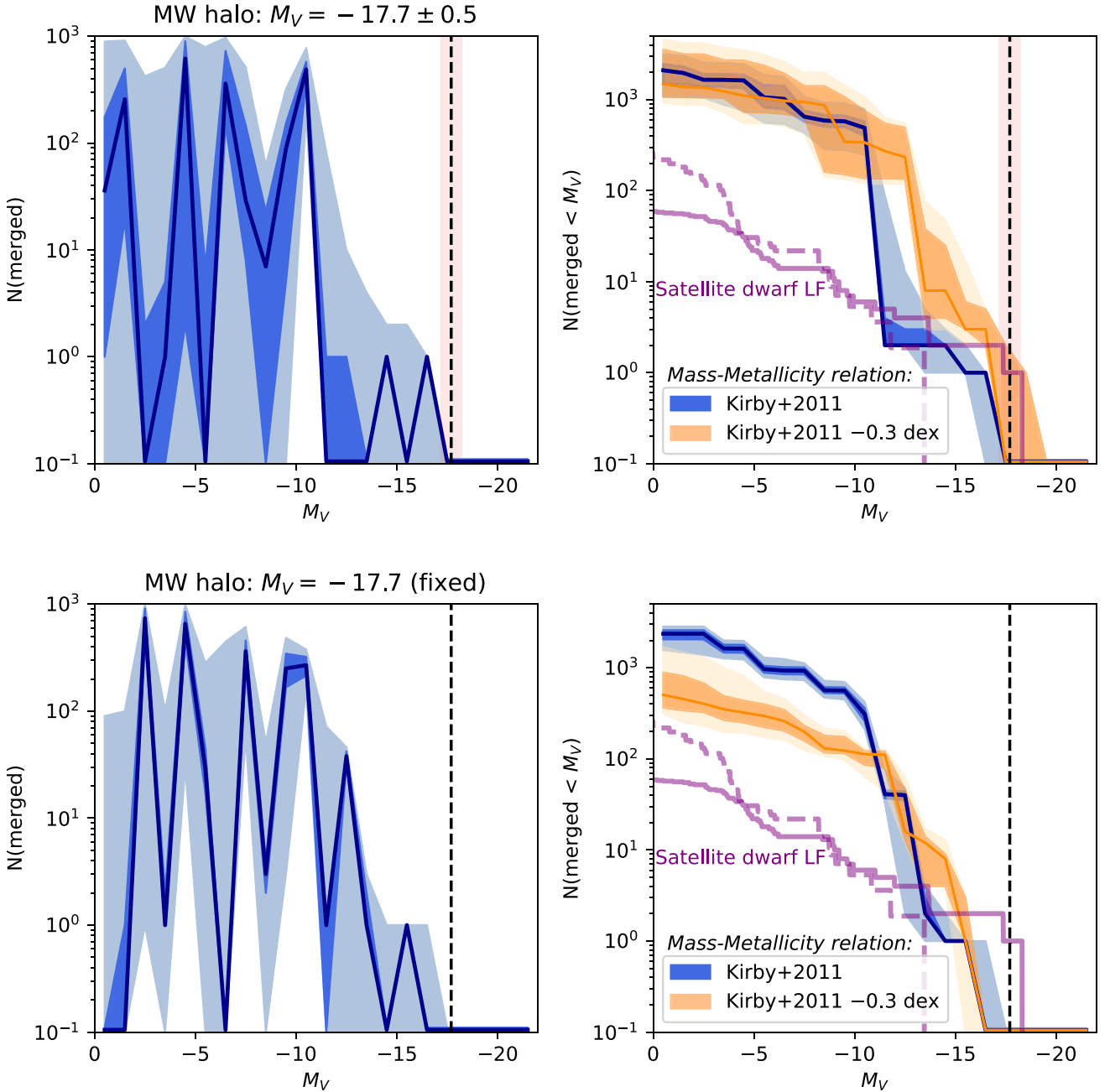
<sup>3</sup>Note that this metallicity cut is made in both the data *and* model, so there is no metallicity bias introduced with our selection.

number of halo stars with metallicity measurements by combining these surveys but note that, ideally, a more homogeneous sample would be used. For now, we continue on, under the assumption that there are no significant metallicity biases in this combined sample. However, we stress that future work with upcoming spectroscopic surveys such as DESI (Cooper et al. 2022) and WEAVE (Dalton et al. 2012) will be much better suited for this type of analysis.

In our analysis, we adopt a total halo luminosity of  $M_V = -17.7 \pm 0.5$ . This is consistent with recent measurements which suggest  $L \sim 1 \times 10^9 L_\odot$ , and also allows a range around this value encompassing the majority of observational constraints and their uncertainties (Deason, Belokurov & Sanders 2019; Mackereth & Bovy 2020; Horta et al. 2021b). The top panels of Fig. 5 show the resulting number of destroyed dwarfs in the MW halo as a function of  $M_V$ . We consider dwarfs with  $0 > M_V > -22$ , using 22 bins with 1 mag bin size. Note that the size of our sample means that we are unlikely constraining dwarfs with  $M_V \gtrsim -8$ , which will not be represented by a large enough number of stars (see e.g. Section 5.2). In blue, we show the results when the Kirby et al. (2011) mass-metallicity relation is used, which is appropriate for surviving dwarf galaxies in the MW. In recent work, Naidu et al. (2022) (see also Fattahi et al. 2020) argue that *destroyed* dwarfs may not lie on this relation, and a relation with  $\sim 0.3$  dex offset to lower metallicities is more appropriate. We show the results with this offset applied in orange.

Our model predicts several hundred ( $N \sim 400$ ) destroyed dwarfs with  $M_V \lesssim -10$ . However, the different mass-metallicity relations (relevant for either ‘surviving’ or ‘destroyed’ dwarfs) predict different distributions of progenitor masses, particularly at larger masses. For example, when using the Kirby et al. (2011) mass-metallicity relation applicable for surviving dwarf galaxies, we estimate  $N = 1$  massive dwarf progenitor with  $L \sim 10^{8.5} L_\odot$ , but this rises to  $N = 3$  when the relation more relevant to destroyed dwarf galaxies is used instead. This seems to be at odds with our adopted *total* halo luminosity of  $L \sim 1 \times 10^9 L_\odot$ . Indeed, by summing the predicted numbers of destroyed dwarfs we find that the total luminosity when the Kirby et al. (2011) relation is used is  $1.1^{+0.2}_{-0.2} \times 10^9 L_\odot$ , but this rises to  $3.4^{+7.2}_{-2.3} \times 10^9 L_\odot$  when an 0.3 dex offset is applied to the mass-metallicity relation. Clearly, in this latter case, the bias in metallicity has pushed the progenitor masses higher, and, because we have allowed a fairly flexible total luminosity, resulted in a high halo luminosity. However, it is still consistent with the input luminosity within  $1-\sigma$ .

In the bottom panel of Fig. 5, we show the results when the total halo luminosity is fixed to  $M_V = -17.7$  (technically, an uncertainty of 0.01 dex is adopted). Here, the ‘fiducial’ result using the Kirby et al. (2011) mass-metallicity relation is only slightly changed. For example, the most massive progenitor is shifted to a slightly lower luminosity (by  $\sim 1$  dex in  $M_V$ ), and the total number of dwarfs with  $M_V < -10$  is reduced ( $N \sim 300$ ). In general, the changes are within the predicted uncertainties. When an 0.3 dex metallicity offset is applied to the mass-metallicity relation, fixing the halo luminosity has a greater effect. This is unsurprising given that allowing for a more flexible halo luminosity favours a higher value than the fiducial  $1 \times 10^9 L_\odot$ . In this case, the most massive progenitor has  $L \sim 10^{8.1} L_\odot$  (compared to  $N \sim 3$  with  $L \sim 10^{8.5} L_\odot$  when the total luminosity is more flexible). The number of low-mass dwarfs is also reduced, with  $N \sim 110$  with  $M_V < -10$ . This exercise emphasizes how important the assumed total halo luminosity, as well as the adopted mass-metallicity relation are for this type of analysis.

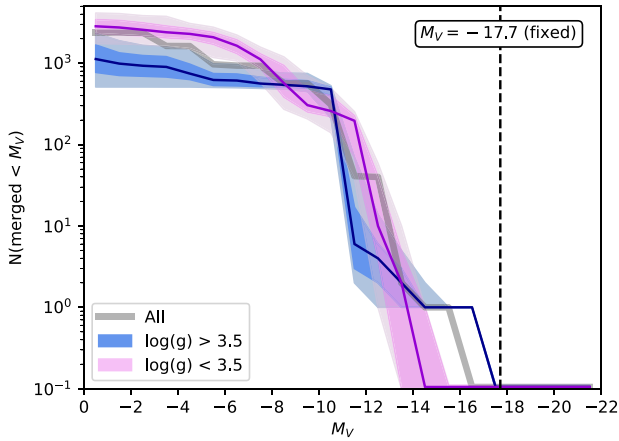


**Figure 5.** The estimated differential (left) and cumulative (right) number of destroyed dwarfs in the MW halo. The dark(light)-shaded regions show the 16–84(1–99) percentiles, and the solid lines are the medians. The dashed black line indicates the assumed total stellar halo luminosity ( $M_V = -17.7$ ). In the top panels, the total luminosity has a flexible uncertainty of  $\pm 0.5$  dex, whereas in the bottom panel the total luminosity is kept fixed. The results in blue are for when the  $z = 0$  mass–metallicity relation for dwarfs is assumed (Kirby et al. 2011). In orange, we show the results when a  $-0.3$  dex offset is applied to the relation, which has been postulated to be more applicable to destroyed dwarfs (Naidu et al. 2022). For comparison, we show the surviving dwarf satellite luminosity function in purple. The dashed line indicates the completeness-corrected LF derived by Drlica-Wagner et al. (2020).

We also show the surviving dwarf satellite luminosity function for comparison in Fig. 5. Here, we show the observed (solid purple) and completeness-corrected (dashed purple) cumulative number counts given by Drlica-Wagner et al. (2020). Note that the completeness-corrected counts differ from the ‘observed’ counts at the bright end because it does not include the massive satellites (Sagittarius, SMC and LMC). The numbers of low luminosity satellite systems are much lower than the predicted number of destroyed dwarfs. This is perhaps unsurprising given that our estimates are likely overestimated at

low luminosities, owing both to sample size and our assumption of Gaussian MDFs (see Section 5.1). Interestingly, the (cumulative) number counts are similar at intermediate luminosities ( $-16 \lesssim M_V \lesssim -12$ ) but destroyed dwarfs as massive as the LMC are not favoured unless the adopted mass–metallicity relation is adjusted from the fiducial  $z = 0$  form.

Fattahi et al. (2020) show using the Auriga simulation suite that the number of destroyed dwarfs in MW-mass haloes is larger than the number of surviving satellites, at least down to  $M_V \sim -8$ . This is



**Figure 6.** The estimated cumulative number of destroyed dwarfs in the MW halo. Same as Fig. 5, but split into two bins with low ( $\log(g) < 3.5$ ) and high ( $\log(g) > 3.5$ ) surface gravity stars. The thick grey line indicates the overall sample. The stellar halo luminosity is fixed ( $M_V = -17.7$ ).

in agreement with our results, however, our estimated *total* number of destroyed dwarfs is far higher than these models (by a factor of  $\sim 3$ – $10$ , see also Fig. 7). This could be a genuine tension with the models, but it is worth stressing that our number estimates at low luminosities are likely biased high, and the numbers could be reduced if we had larger sample sizes and/or the metal-poor tails of higher mass systems are taken into account (see Section 5.1).

Finally, given the heterogeneous nature of our sample of halo stars, we consider how different cuts in surface gravity affect the results. Namely, dwarf stars and giants can have different metallicity biases, and probe different volumes in magnitude-limited surveys. The MDF of our halo sample split by  $\log(g)$  was shown in Fig. 4. Here, we can see there are slight differences for low and high  $\log(g)$ , and now we consider how our inferred number counts of destroyed dwarfs are affected. The cumulative number of destroyed dwarfs is shown in Fig. 6 with two different bins of  $\log(g)$ , appropriate for dwarf stars ( $\log(g) > 3.5$ ) and giants ( $\log(g) < 3.5$ ). It is worth bearing in mind that our overall sample is dominated by the high surface gravity dwarf stars (approximately  $\sim 2/3$  have  $\log(g) > 3.5$ ). Note that here we only use the Kirby et al. (2011) mass–metallicity relation, and the total halo luminosity is fixed. Encouragingly, the total number of progenitors (for  $M_V \lesssim -10$ ) is very similar for the two bins of  $\log(g)$ . However, massive progenitors ( $L \gtrsim 10^8 L_\odot$ ) are only favoured in the high  $\log(g)$  sample. This is likely because the MDF is biased towards lower metallicities for the giant star sample (see Fig. 4). Moreover, the giant and dwarfs are probing slightly different volumes, with the high surface gravity dwarfs more concentrated around the solar neighbourhood. This exercise highlights the difficulty of using a ‘hodge-podge’ of halo stars for our analysis, and it will clearly be preferable for future work to have a more homogeneous sample, where the selection function is clearly defined.

#### 4 AURIGA SIMULATIONS

Our modelling procedure makes various assumptions and simplifications. For example, it assumes each progenitor galaxy is sampled in a representative way, and that their MDFs are adequately described by a Gaussian distribution. In reality, this may not be the case, particularly for volume-limited Galactic-sized stellar haloes. To this end, we test

our model on simulated MW stellar haloes, which are representative of ‘realistic’ accreted populations. We apply our modelling procedure to halo stars in the Auriga simulations (Grand et al. 2017); these cosmological hydrodynamical simulations are a suite of  $N \sim 30$  high resolution ( $m_p \sim 5 \times 10^4 M_\odot$ ) MW–mass ( $1 - 2 \times 10^{12} M_\odot$ ) haloes. In this work, we make use of the  $N = 28$  haloes studied in Fattahi et al. (2019), which omits two haloes currently undergoing major mergers. We only consider accreted halo stars, which are identified in Fattahi et al. (2019) as those that formed in subhaloes other than the main progenitor galaxy.

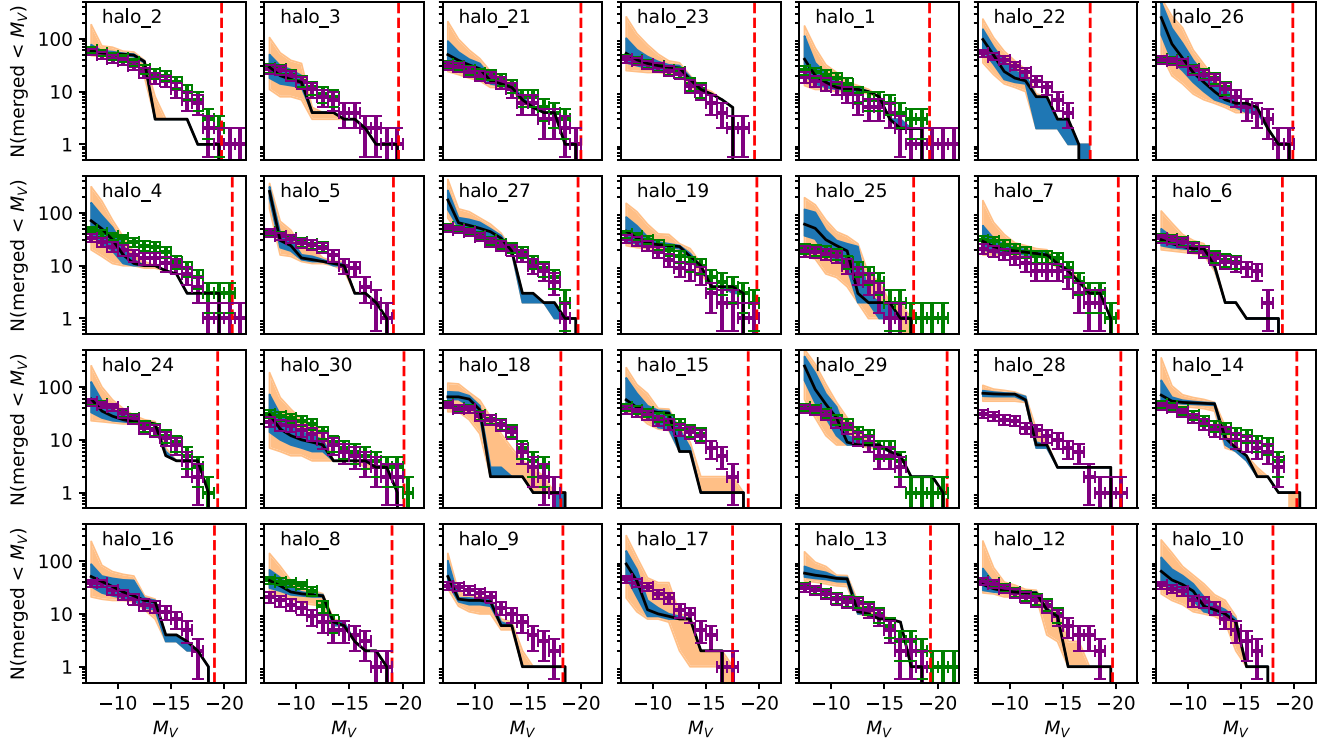
For each halo, we construct a sample of halo star particles within  $r < 20$  kpc. This is chosen to roughly mimic the volume limit of current observations, and ensure large enough sample sizes. The input into the model is the  $[\text{Fe}/\text{H}]$  values of the stellar particles. Of course, in the simulations, we also know the progenitor galaxy of each star particle, and can thus test the estimated mass spectrum of accreted dwarfs from our modelling procedure. The final ingredient we need to define is the mass–metallicity relation for the Auriga simulations. Grand et al. (2021) show that the mass–metallicity relation for dwarf galaxies in Auriga is in good agreement with low-mass dwarfs ( $M_{\text{star}} \sim 10^6 M_\odot$ ), but is too metal-rich by  $\sim 0.5$  dex for more massive dwarfs (see fig. 13 in Grand et al. 2021). We use all the destroyed dwarf progenitors across the  $N = 28$  Auriga haloes to calibrate this relation.<sup>4</sup> However, we do exclude dwarfs that are accreted recently (less than 5 Gyr ago) as these can have significantly different metallicities due to ongoing star formation. The debris from these events is still included in the analysis, but our calibration is only based on the relatively old dwarf galaxies. Note that we only consider dwarf progenitors with  $M_V > -7$ , which corresponds to a stellar mass of  $M_{\text{star}} > 10^5 M_\odot$  or  $N > 2$  star particles. We use the ‘peak’ stellar mass of each dwarf, which corresponds to the maximum stellar mass the progenitor has reached (note that we get similar results if the stellar mass at infall is used instead). The resulting mass–metallicity<sup>5</sup> relation for Auriga is  $[\text{Fe}/\text{H}] = -1.69 + 0.39 \times (\log_{10} L - 6)$ . To estimate the scatter around this mean relation, we calculate the scatter for each individual halo, and use the median value across all haloes. This results in a scatter around the mean  $[\text{Fe}/\text{H}]$  relation of 0.3 dex. Finally, we consider the spread in  $[\text{Fe}/\text{H}]$  for individual dwarfs. Unlike the observations, we find no strong evidence for a variation with dwarf mass, so instead adopt a constant dispersion of 0.4 dex of the MDF for all dwarfs. Armed with the mass–metallicity relation appropriate for Auriga, we can now test our modelling procedure on these cosmological haloes.

When applying our method to the Auriga haloes, we assume the total luminosity of the halo is known. This of course results in additional uncertainty in the real observations, but we particularly want to investigate the systematic influences present in the cosmological simulations. We consider accreted dwarfs in the range  $-7 > M_V > -22$ , and estimate the number of dwarfs in 15 bins with bin size of 1 mag. Fig. 7 shows the resulting cumulative number of destroyed dwarfs in the Auriga haloes. Each panel shows a different halo, and our estimated numbers are shown with the solid black lines (median), and blue/orange shaded regions (16–84/1–99 percentiles). The points with error bars are the true values, with Poisson noise adopted for the uncertainties in each  $M_V$  bin. Note that the ‘true’ values include all dwarfs that have deposited *any* material within

<sup>4</sup>To clarify, all destroyed dwarfs are used, not just those that have debris within 20 kpc of the host halo

<sup>5</sup>Note that we assume a stellar mass-to-light ratio of  $(M/L) = 2$  to convert stellar mass to luminosity.



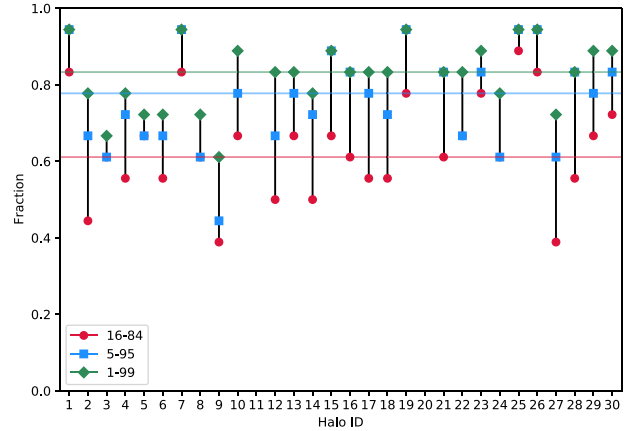


**Figure 7.** The estimated cumulative number of destroyed dwarfs for the  $N = 28$  Auriga haloes. The solid black line shows the median, and the shaded blue/orange regions the 16–84(1–99) percentiles. The red dashed line indicates the assumed total luminosity of the halo. For each halo, accreted star particles are selected within  $r < 20$  kpc. The points with (Poisson) error bars indicate the ‘truth’, with all progenitors shown in green, and only those accreted  $>5$  Gyr ago in purple. The latter are shown because recently accreted dwarfs are likely (i) not fully phase-mixed, and (ii) can significantly deviate from the mass–metallicity relation for (old) dwarf galaxies in Auriga.

20 kpc of the host halo. Thus, there can be cases where only a small fraction of a destroyed dwarf is included in the sample (see below). The green values in Fig. 7 are for *all* progenitors, while the purple are only those accreted earlier than 5 Gyr ago. In many cases, there is little difference between the green and purple values, because most dwarfs are accreted at earlier times. However, we highlight the most recently accreted dwarfs because these are likely not fully phase-mixed, and can significantly deviate from the mass–metallicity relation for (old) dwarf galaxies in Auriga (see above). In reality, we find that these recently accreted dwarfs only cause a significant effect if the progenitors are relatively massive (e.g. Halo 25).

We discuss these results more quantitatively below, but first cast a qualitative eye on Fig. 7. In general, our estimates agree well with the true mass spectrum of accreted dwarfs. However, in some cases, there can be notable differences. We find that the most significant deviations are due to the following: (1) relatively massive progenitors that lie off the mass–metallicity relation (e.g. Halo 2, 6) and/or (2) progenitors with a low fraction of their material within the given radial range (e.g. Halo 15, 27). These systematics, and sometimes the combination of both, are most likely to cause our method to fail. On the other hand, there are a significant number of haloes for which we recover the mass spectrum very well, which is encouraging given the complexity of these hydrodynamic simulations, and the cosmological nature of their assembly histories.

In Fig. 8, we give a more quantitative summary of our tests of the Auriga haloes. Here, for each halo (identified in the  $x$ -axis) we show the fraction of  $M_V$  bins that have number estimates that



**Figure 8.** Quantifying the test with Auriga haloes. For each halo, we show the fraction of  $M_V$  bins (1 mag wide) that have estimated numbers that agree within the 16–84 (red-filled circles), 5–95 (blue-filled squares), and 1–99 (green-filled diamonds) percentage confidence limits, respectively. The median values are shown with the horizontal coloured lines.

agree within the 16–84, 5–95, and 1–99 percentile confidence limits. The median recovery fractions across all haloes are 0.61, 0.77, and 0.83, respectively. These fractions are below the expected fractions for a ‘perfect’ procedure, but this is unsurprising given the various systematic influences present in the simulations, such as deviations from the adopted mass–metallicity relation and the presence of stellar

debris that does not fully occupy the available phase space. These, of course, are realistic effects that could be present in the observational data.

In Fig. 9, we explore the halo-to-halo scatter more closely. In the left-hand panel, we show the difference between the estimated and true cumulative numbers of destroyed dwarfs as a function of  $M_V$ . The black line shows the median of the  $N = 28$  haloes, and the blue and orange shaded regions show the 16–84 and 1–99 percentiles, respectively. The deviation from the true numbers is fairly symmetrical and only starts to shift from zero for very low-mass progenitors. It is worth noting that there is a trend towards overestimating the number of accreted dwarfs at lower masses. This could be a real effect, caused by e.g. the assumption of Gaussian MDFs, however, this low-mass regime may also be affected by resolution limitations in the simulations, as the MDFs of these dwarfs are only represented by a handful of star particles. In the right-hand panel, we show the estimated vs. true cumulative number of progenitors in four different magnitude ranges. Here, we can see that there is considerable scatter around the 1-to-1 line, but the spread is fairly symmetrical. Finally, to quantify these findings we compute the typical accuracy of our  $N(< M_V)$  estimates (averaged over all  $M_V$  bins); we find that  $N(< M_V)_{\text{est}}/N(< M_V)_{\text{true}} = 0.9^{+0.6}_{-0.4}$ . Thus, we estimate that our method is able to recover the true  $N(< M_V)$  within 50 per cent for most  $M_V$  bins.

## 5 DISCUSSION

### 5.1 Caveats and potential improvements

Probably the most significant caveat in our modelling approach is the assumption of Gaussian MDFs. We know that galaxies are expected to have metallicity distributions that are not Gaussian (Revaz et al. 2009; Kirby et al. 2011, 2013). The details of non-Gaussianity heavily depend on the star formation history, the time-scale and intensity of gas inflows (Lanfranchi & Matteucci 2004; Romano & Starkenburg 2013), and likely other processes. The non-Gaussianity is likely a bigger problem for more luminous systems, as they have more extended star formation histories compared to faint systems, which, in some cases, are consistent with a single burst of star formation before reionization (Weisz et al. 2014). What is the possible systematic effect of neglecting the non-Gaussianity? Assuming that the non-Gaussianity is not caused by accreted systems, but is intrinsic, that would lead us to overestimate the number of accreted fainter systems. Thus, our constraints would be upper limits on the number of accreted events. However, the Gaussian assumption is something that can be potentially fixed in our formalism. For example, it could be done by assuming parametric MDF families from Kirby et al. (2011), where one would need to assume some dependence of the MDF parameters on galaxy luminosity.

Another key assumption is that all of the accreted stars are coming from dwarf galaxies. However, it is likely that some fraction of stars (at least in the MW) are coming from disrupted globular clusters. If trends seen in more massive galaxies extend to faint dwarfs (Forbes et al. 2018; Huang & Koposov 2021; Eadie, Harris & Springford 2022), we may expect that 0.1–1 per cent of stars come from disrupted GCs. Note, however, that other works have argued for much higher fractions (e.g. Martell et al. 2011). The metallicity distribution of clusters is poorly understood, and since individual GCs have extremely narrow MDFs it is unclear if there is a solution to take GCs into account in our model.

### 5.2 Future prospects

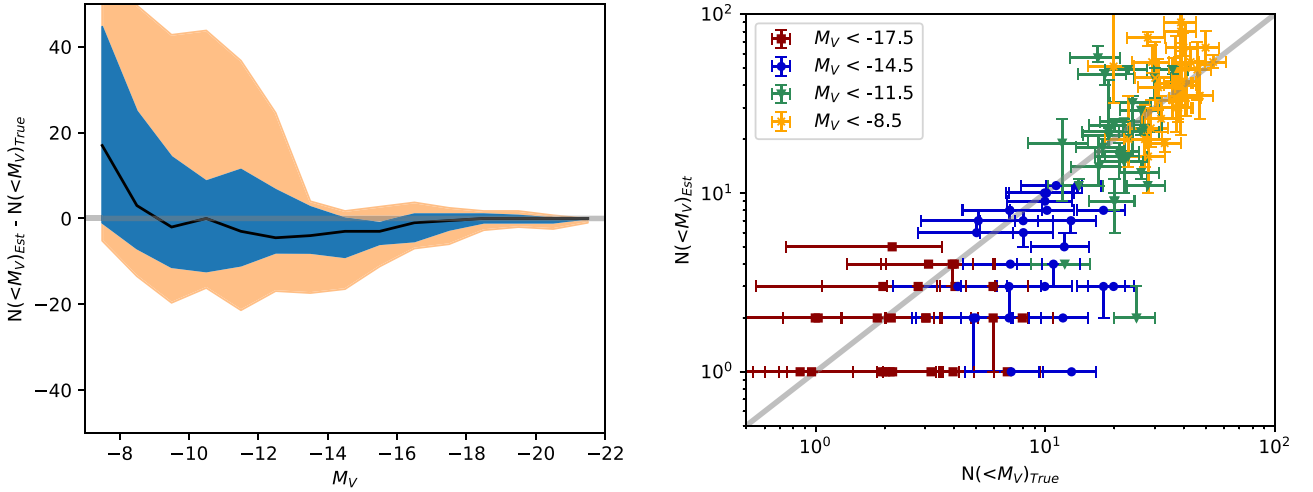
The MDF modelling procedure we have outlined in this work has compelling potential when applied to future spectroscopic data sets. In particular, the availability of much larger numbers of stars with metallicity measurements will allow us to probe to lower dwarf mass scales, and potentially constrain the number of destroyed ultra-faint dwarfs. These latter measurements would not only inform us about the low-mass accretion history of galaxies but could also be used to constrain small-scale galaxy formation and the nature of dark matter (Deason et al. 2022).

Here, we use toy models to estimate the sample sizes needed to probe down to the ultra-faint mass scale ( $M_V \gtrsim -8$ ). Note here we focus on the *ideal* case and ignore the potential caveats discussed in the previous subsection and elsewhere. We generate Gaussian MDFs that follow the Kirby et al. (2011) mass–metallicity relation, with varying sample sizes. We consider two example cases, one similar to a classical dwarf galaxy ( $M_V = -13.5$ ), and another akin to a Galactic stellar halo with one main progenitor ( $M_V = -17.5$ ). For each case, we generate the central MDFs with no lower mass progenitors, or with an additional  $N = 10$ –50 low luminosity systems ( $M_V = -7.5$ ). The results of this exercise are shown in Figs 10 and 11.

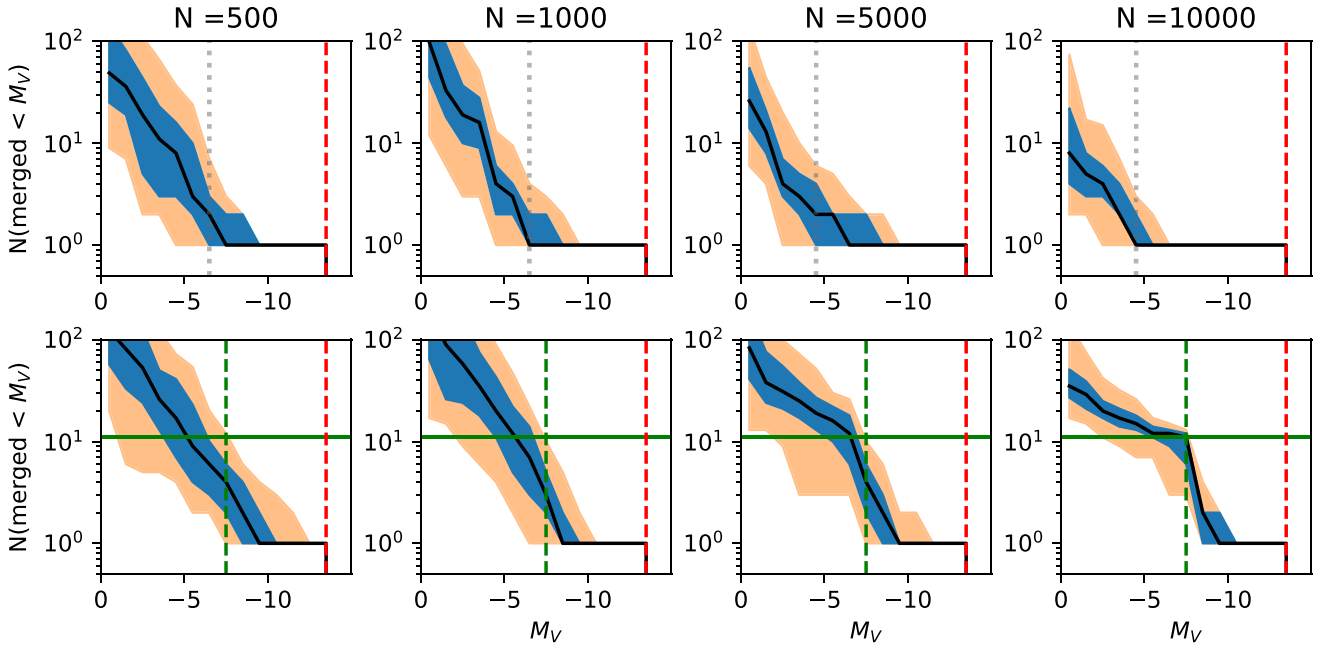
It is immediately clear that as the sample sizes increase, we can probe to lower mass scales. The grey dotted line in the top-row of Figs 10 and 11 indicates where the estimated number of contributors starts to become less reliable (i.e. when  $N_{\text{Merged}} > 1$  for fake tests with no merger events). For the typical sample size of the classical MW satellites ( $N \sim 500$ ), we can currently only reliably probe down to  $M_V \gtrsim -7$ . For Galactic haloes with  $N \sim 10^4$  tracers, we can likely probe down to  $M_V \gtrsim -8$ . In order to probe down to the ultra-faint regime ( $M_V \gtrsim -5$ ) requires significant sample sizes that are not currently available. For example, for a typical classical dwarf  $N \gtrsim 5000$  stars are needed to unambiguously detect low-mass progenitors. On the other hand, for Galactic stellar haloes the sample sizes likely need to exceed  $N \gtrsim 10^5$ . Although these numbers are larger than the sample sizes currently available, they are achievable with upcoming spectroscopic surveys. Indeed, the large field-of-view and copious number of fibres available in the DESI, WEAVE, and 4MOST instruments, make them ideal tools for this task. Dedicated programs focusing on classical dwarf satellite galaxies could yield thousands of member stars with spectroscopic measurements. Furthermore, the MW surveys planned with these facilities are predicted to obtain measurements for  $N \sim 10^6$  halo stars between 10 and 30 kpc (e.g. Cooper et al. 2022). These survey data will not only provide significant numbers of dwarf members and halo stars with metallicity measurements, but will also provide more homogeneous sampling, and well-defined selection functions. This latter point is a particular downside of the current implementation in this work, which relies on a combination of data samples with ill-defined selection functions. In summary, the method we propose here is poised to exploit upcoming data sets to robustly quantify the accreted populations of stars in the MW and its dwarf galaxies.

## 6 CONCLUSIONS

We have introduced a new statistical method to model the MDF of a stellar population as an ensemble of individual components. These components follow the galaxy mass–metallicity relation and are assumed to be Gaussian distributed around their mean values (with a mass-dependent spread). We apply the method to observations of the MW halo and classical dwarf satellites, and we also test the



**Figure 9.** Comparing the estimated and true numbers of destroyed progenitors in the ( $N = 28$ ) Auriga haloes. Left-hand panel: We show the difference between the estimated and true (cumulative) numbers as a function of  $M_V$ . The solid black line shows the median, and the shaded blue (orange) regions the 16–84 (1–99) percentiles. Right-hand panel: The estimated versus the true number in different ranges of  $M_V$ . Error bars show the Poisson errors in the true numbers and the 16–84 percentiles in the estimated numbers.



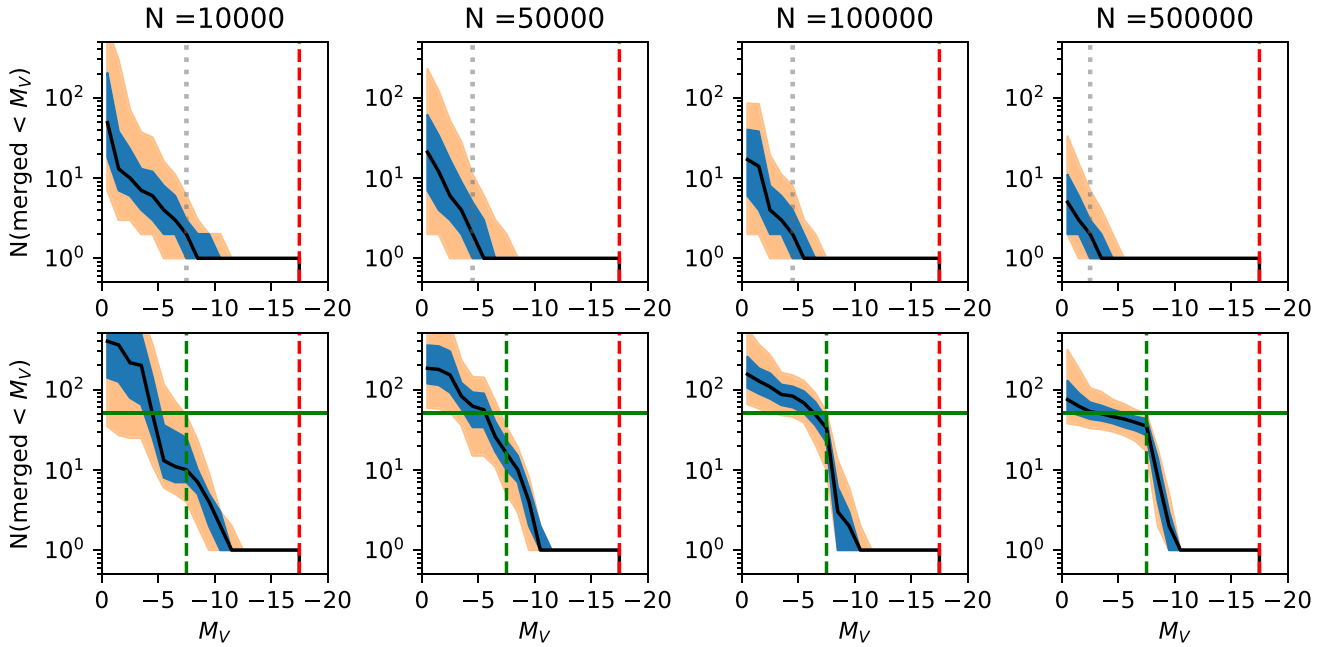
**Figure 10.** Testing the method on dwarf galaxies with toy fake data. Here, dwarfs are generated with Gaussian MDFs following the adopted  $z = 0$  mass–metallicity relation. In the top row, there are no merger events (just the MDF of the central galaxy,  $M_V = -13.5$ ). The dotted grey line indicates the approximate  $M_V$  value where the estimated number of contributors becomes less reliable. As the sample sizes increase we can probe to fainter luminosity systems. In the bottom row,  $N = 10$  low mass ( $M_V = -7.5$ ) systems are included. The size of the samples generated increases with each column. The solid black line shows the median, and the shaded blue (orange) regions the 16–84 (1–99) percentiles. The vertical red dashed line indicates the  $M_V$  of the central galaxy, and the green dashed line shows the  $M_V$  of the accreted system (if included). The true number of lower mass systems is shown with the solid horizontal green line.

procedure on cosmological hydrodynamical simulations of MW–mass haloes. Our main conclusions are summarized as follows:

(i) Most samples of stars associated with MW dwarf satellites are too small to robustly probe lower mass accretion events. However, we do not find any evidence for significant mergers, and can indeed in some cases (e.g. Fornax, Leo I), rule out accreted components more massive than  $M_{V,\text{host}} + 5$  (or  $L_{\text{host}}/100$ ).

(ii) We constructed a sample of MW halo stars within  $r < 10$  kpc using several spectroscopic surveys and *Gaia* data. By adopting the mass–metallicity relation applicable to surviving dwarf galaxies we find that one massive progenitor is favoured with  $L \sim 10^{8.5} L_\odot$ , and there are several hundred ( $N \sim 400$ ) progenitors in total down to  $M_V < -10$ .

(iii) We also consider a mass–metallicity relation more appropriate for *destroyed* dwarf galaxies for the MW stellar halo, as suggested by



**Figure 11.** Same as Fig. 10 but for MW haloes. Here, one massive progenitor is generated ( $M_V = -17.5$ ) with no other progenitors (top panel), and with  $N = 50$  additional low-mass progenitors (bottom panel).

Naidu et al. (2022). Here,  $N = 3$  massive progenitors are favoured, but the total number of progenitors down to  $M_V < -10$  is similar to the fiducial case. By placing a stringent constraint on the total halo luminosity ( $L_{\text{tot}} = 10^9 M_\odot$ ), the two different mass–metallicity relations give more similar results for massive progenitors, but the total number of progenitors differs more significantly (by a factor of 3).

(iv) We find that the total halo luminosity in our model, and the adopted mass–metallicity relation, are both important assumptions. The former can be constrained by other means (e.g. Deason et al. 2019; Mackereth & Bovy 2020), and more work needs to be done to understand the redshift evolution of the stellar mass–metallicity relation<sup>6</sup>

(v) Our modelling procedure is applied to the hydrodynamic cosmological Auriga simulations, a suite of  $N \sim 30$  MW–mass haloes. Here, many of our assumptions (e.g. phase-mixed material, Gaussian MDFs) are unlikely to hold, so this provides a strong test for our method. We find that, in many cases, our procedure works well, and most failures come from scatter in the mass–metallicity relation and/or recent accretion events not fully occupying the phase-space we are probing. In general, we find that we can recover the true luminosity function ( $N(< M_V)$ ) of destroyed dwarfs to within 50 per cent for most  $M_V$  bins.

(vi) Finally, we consider how the increase in sample sizes from future spectroscopic surveys can allow us to probe down to the ultra-faint dwarf mass scale ( $M_V > -10$ ). We find that MW stellar halo samples with  $N \sim 10^6$  tracers will allow us to probe down to  $M_V > -10$ ; encouragingly, this should be feasible with upcoming surveys such as DESI and WEAVE. Moreover, with sample sizes exceeding  $N \sim 5000$ , we should be able to probe the lower mass accretion events associated with classical dwarf satellites in the MW. Our ability to

probe down to these puny stellar systems will enable us to address fundamental questions about galaxy formation at the lowest mass scales and, potentially, the nature of dark matter.

We have shown that using *only* the MDF of an (accreted) stellar population, the mass spectrum of its progenitors can be uncovered. This is encouraging for the upcoming generation of spectroscopic surveys of the MW. However, a possible extension of this work would be to combine the MDF modelling with phase-space data and/or additional chemical dimensions (see e.g. Cunningham et al. 2022). The addition of dynamical information could provide tighter constraints on the luminosity function of destroyed dwarfs. In particular, where the MDF modelling is weakest, i.e. when the stellar material is un-mixed in phase space, is likely where the dynamical data is the most informative. Moving forward, modelling in the chemodynamical space is the next logical step, and, importantly, we will have the data to do this. Thus, it is clear that future data sets combined with modelling methods such as that presented here will provide all the tools needed to finally quantify the accretion history of the Galaxy and its satellite population.

## ACKNOWLEDGEMENTS

We thank an anonymous referee for providing useful comments that helped improve the paper.

AD is supported by a Royal Society University Research Fellowship. AD acknowledges support from the Leverhulme Trust and the Science and Technology Facilities Council (STFC; grant numbers ST/P000541/1 and ST/T000244/1). AF is supported by a UKRI Future Leaders Fellowship (grant number MR/T042362/1). RG acknowledges financial support from the Spanish Ministry of Science and Innovation (MICINN) through the Spanish State Research Agency, under the Severo Ochoa Program 2020–2023 (CEX2019-000920-S).

This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC

<sup>6</sup>Although the study of the evolution of the mass–metallicity relation is in its infancy, there are several efforts in this direction (e.g. Choi et al. 2014; Leethochawalit et al. 2018, 2019; Beverage et al. 2021; Zhuang et al. 2022).



DiRAC HPC Facility ([www.dirac.ac.uk](http://www.dirac.ac.uk)). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1, and ST/S002502/1, Durham University and STFC operations grant number ST/R000832/1. DiRAC is part of the National e-Infrastructure. This work was performed in part at Aspen Center for Physics, which is supported by National Science Foundation grant number PHY-1607611. This work was partially supported by a grant from the Simons Foundation.

For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any author-accepted manuscript version arising from this submission.

AD thanks Ethan Nadler for providing the completeness-corrected estimates of the MW dwarf satellite luminosity function.

## DATA AVAILABILITY

The data analysed in this article can be made available upon reasonable request to the corresponding authors.

The code used to perform the MDF modelling is available on Github<sup>7</sup>

## REFERENCES

- Abolfathi B. et al., 2018, *ApJS*, 235, 42
- Bailer-Jones C. A. L., Rybizki J., Foesneau M., Demleitner M., Andrae R., 2021, *AJ*, 161, 147
- Belokurov V. et al., 2007a, *ApJ*, 657, L89
- Belokurov V. et al., 2007b, *ApJ*, 658, 337
- Belokurov V., 2013, *New Astron. Rev.*, 57, 100
- Belokurov V., Erkal D., Evans N. W., Koposov S. E., Deason A. J., 2018, *MNRAS*, 478, 611
- Belokurov V., Sanders J. L., Fattahi A., Smith M. C., Deason A. J., Evans N. W., Grand R. J. J., 2020, *MNRAS*, 494, 3880
- Benitez-Llambay A., Frenk C., 2020, *MNRAS*, 498, 4887
- Beverage A. G., Kriek M., Conroy C., Bezanson R., Franx M., van der Wel A., 2021, *ApJ*, 917, L1
- Buder S. et al., 2021, *MNRAS*, 506, 150
- Callingham T. M., Cautun M., Deason A. J., Frenk C. S., Grand R. J. J., Marinacci F., 2022, *MNRAS*, 513, 4107
- Chandra V. et al., 2022, preprint ([arXiv:2212.00806](https://arxiv.org/abs/2212.00806))
- Choi J., Conroy C., Moustakas J., Graves G. J., Holden B. P., Brodwin M., Brown M. J. I., van Dokkum P. G., 2014, *ApJ*, 792, 95
- Conroy C. et al., 2019, *ApJ*, 883, 107
- Cooper A. P. et al., 2022, preprint ([arXiv:2208.08514](https://arxiv.org/abs/2208.08514))
- Cunningham E. C. et al., 2022, *ApJ*, 934, 172
- Dalton G. et al. 2012, in McLean I. S., Ramsay S. K., Takami H. eds, Proc. SPIE Conf. Ser. Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV. SPIE, Bellingham, p. 84460P
- de Jong R. S. et al. 2019, *Messenger*, 175, 3
- Deason A. J., Belokurov V., Sanders J. L., 2019, *MNRAS*, 490, 3426
- Deason A. J., Bose S., Fattahi A., Amorisco N. C., Hellwing W., Frenk C. S., 2022, *MNRAS*, 511, 4044
- Dekel A., Silk J., 1986, *ApJ*, 303, 39
- Drlica-Wagner A. et al., 2020, *ApJ*, 893, 47
- Eadie G. M., Harris W. E., Springford A., 2022, *ApJ*, 926, 162
- Fattahi A. et al., 2019, *MNRAS*, 484, 4471
- Fattahi A. et al., 2020, *MNRAS*, 497, 4459
- Forbes D. A., Read J. I., Gieles M., Collins M. L. M., 2018, *MNRAS*, 481, 5592
- Fowlie A., Handley W., Su L., 2020, *MNRAS*, 497, 5256
- Gaia Collaboration, 2016, *A&A*, 595, A1
- Gaia Collaboration, 2021, *A&A*, 649, A1
- Grand R. J. J. et al., 2017, *MNRAS*, 467, 179
- Grand R. J. J. et al., 2021, *MNRAS*, 507, 4953
- Green P. J., 1995, *Biometrika*, 82, 711
- Helmi A., 2008, *A&AR*, 15, 145
- Helmi A., Babusiaux C., Koppelman H. H., Massari D., Veljanoski J., Brown A. G. A., 2018, *Nature*, 563, 85
- Horta D. et al., 2021a, *MNRAS*, 500, 1385
- Horta D. et al., 2021b, *MNRAS*, 500, 5462
- Huang K.-W., Koposov S. E., 2021, *MNRAS*, 500, 986
- Jin S. et al., 2022, preprint ([arXiv:2212.03981](https://arxiv.org/abs/2212.03981))
- Johnston K. V., Bullock J. S., Sharma S., Font A., Robertson B. E., Leitner S. N., 2008, *ApJ*, 689, 936
- Jurić M. et al., 2008, *ApJ*, 673, 864
- Kirby E. N., Cohen J. G., Guhathakurta P., Cheng L., Bullock J. S., Gallazzi A., 2013, *ApJ*, 779, 102
- Kirby E. N., Lanfranchi G. A., Simon J. D., Cohen J. G., Guhathakurta P., 2011, *ApJ*, 727, 78
- Koposov S. E. et al., 2022b, preprint ([arXiv:2211.04495](https://arxiv.org/abs/2211.04495))
- Koposov S. et al., 2008, *ApJ*, 686, 279
- Koposov S. et al., 2022a, joshspeagle/dynesty: v1.2.3., <https://doi.org/10.5281/zenodo.6609296>
- Kruijsen J. M. D., Pfeffer J. L., Reina-Campos M., Crain R. A., Bastian N., 2019, *MNRAS*, 486, 3180
- Kunder A. et al., 2017, *AJ*, 153, 75
- Lanfranchi G. A., Matteucci F., 2004, *MNRAS*, 351, 1338
- Leethochawalit N., Kirby E. N., Ellis R. S., Moran S. M., Treu T., 2019, *ApJ*, 885, 100
- Leethochawalit N., Kirby E. N., Moran S. M., Ellis R. S., Treu T., 2018, *ApJ*, 856, 15
- Mackereth J. T., Bovy J., 2020, *MNRAS*, 492, 3631
- Majewski S. R. et al., 2004, *AJ*, 128, 245
- Majewski S. R. et al., 2017, *AJ*, 154, 94
- Martell S. L., Smolinski J. P., Beers T. C., Grebel E. K., 2011, *A&A*, 534, A136
- McConnachie A. W., 2012, *AJ*, 144, 4
- Myeong G. C., Belokurov V., Aguado D. S., Evans N. W., Caldwell N., Bradley J., 2022, *ApJ*, 938, 21
- Nadler E. O. et al., 2020, *ApJ*, 893, 48
- Naidu R. P. et al., 2022, preprint ([arXiv:2204.09057](https://arxiv.org/abs/2204.09057))
- Naidu R. P., Conroy C., Bonaca A., Johnson B. D., Ting Y.-S., Caldwell N., Zaritsky D., Cargile P. A., 2020, *ApJ*, 901, 48
- Newberg H. J. et al., 2003, *ApJ*, 596, L191
- Purcell C. W., Bullock J. S., Zentner A. R., 2007, *ApJ*, 666, 20
- Revaz Y. et al., 2009, *A&A*, 501, 189
- Rix H.-W. et al., 2022, *ApJ*, 941, 45
- Romano D., Starkenburg E., 2013, *MNRAS*, 434, 471
- Simion I. T., Belokurov V., Koposov S. E., 2019, *MNRAS*, 482, 921
- Simon J. D., 2019, *ARA&A*, 57, 375
- Skillman E. D., Kennicutt R. C., Hodge P. W., 1989, *ApJ*, 347, 875
- Speagle J. S., 2020, *MNRAS*, 493, 3132
- Taibi S., Battaglia G., Leaman R., Brooks A., Riggs C., Munshi F., Revaz Y., Jablonka P., 2022, *A&A*, 665, A92
- Takada M. et al., 2014, *PASJ*, 66, R1
- Tollerud E. J., Bullock J. S., Strigari L. E., Willman B., 2008, *ApJ*, 688, 277
- Tolstoy E., Hill V., Tosi M., 2009, *ARA&A*, 47, 371
- Tremonti C. A. et al., 2004, *ApJ*, 613, 898
- Venn K. A., Irwin M., Shetrone M. D., Tout C. A., Hill V., Tolstoy E., 2004, *AJ*, 128, 1177
- Walker M. G., Mateo M., Olszewski E. W., Gnedin O. Y., Wang X., Sen B., Woodroffe M., 2007, *ApJ*, 667, L53
- Walker M. G., Peñarrubia J., 2011, *ApJ*, 742, 20
- Weisz D. R., Dolphin A. E., Skillman E. D., Holtzman J., Gilbert K. M., Dalcanton J. J., Williams B. F., 2014, *ApJ*, 789, 148
- Zhao G., Zhao Y.-H., Chu Y.-Q., Jing Y.-P., Deng L.-C., 2012, *Res. Astron. Astrophys.*, 12, 723
- Zhuang Z. et al., 2022, preprint ([arXiv:2212.04731](https://arxiv.org/abs/2212.04731))

<sup>7</sup>[https://github.com/segasai/mdf\\_modeling\\_paper](https://github.com/segasai/mdf_modeling_paper)