Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **498,** 2472–2491 (2020) doi:10.1093/mnras/staa2452
Advance Access publication 2020 August 17

Kraken reveals itself – the merger history of the Milky Way reconstructed with the E-MOSAICS simulations

J. M. Diederik Kruijssen [1,2]★ Joel L. Pfeffer [3] Mélanie Chevance,[1] Ana Bonaca,[2] Sebastian Trujillo-Gomez [1] Nate Bastian,[3] Marta Reina-Campos [1] Robert A. Crain [3] and Meghan E. Hughes [3]

[1]*Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Mönchhofstraße 12-14, D-69120 Heidelberg, Germany*
[2]*Institute for Theory and Computation, Harvard University, 60 Garden Street, Cambridge, MA 02138, USA*
[3]*Astrophysics Research Institute, Liverpool John Moores University, IC2, Liverpool Science Park, 146 Brownlow Hill, Liverpool L3 5RF, UK*

Accepted 2020 August 5. Received 2020 July 29; in original form 2020 March 1

ABSTRACT

Globular clusters (GCs) formed when the Milky Way experienced a phase of rapid assembly. We use the wealth of information contained in the Galactic GC population to quantify the properties of the satellite galaxies from which the Milky Way assembled. To achieve this, we train an artificial neural network on the E-MOSAICS cosmological simulations of the co-formation and co-evolution of GCs and their host galaxies. The network uses the ages, metallicities, and orbital properties of GCs that formed in the same progenitor galaxies to predict the stellar masses and accretion redshifts of these progenitors. We apply the network to Galactic GCs associated with five progenitors: *Gaia*-Enceladus, the Helmi streams, Sequoia, Sagittarius, and the recently discovered 'low-energy' GCs, which provide an excellent match to the predicted properties of the enigmatic galaxy 'Kraken'. The five galaxies cover a narrow stellar mass range [$M_\star = (0.6\text{–}4.6) \times 10^8$ M$_\odot$], but have widely different accretion redshifts ($z_{\rm acc} = 0.57\text{–}2.65$). All accretion events represent minor mergers, but Kraken likely represents the most major merger ever experienced by the Milky Way, with stellar and virial mass ratios of $r_{M_\star} = 1{:}31^{+34}_{-16}$ and $r_{M_{\rm h}} = 1{:}7^{+4}_{-2}$, respectively. The progenitors match the $z = 0$ relation between GC number and halo virial mass, but have elevated specific frequencies, suggesting an evolution with redshift. Even though these progenitors likely were the Milky Way's most massive accretion events, they contributed a total mass of only $\log(M_{\star,\,\rm tot}/{\rm M}_\odot) = 9.0 \pm 0.1$, similar to the stellar halo. This implies that the Milky Way grew its stellar mass mostly by *in-situ* star formation. We conclude by organizing these accretion events into the most detailed reconstruction to date of the Milky Way's merger tree.

Key words: Galaxy: formation – globular clusters: general – galaxies: evolution – galaxies: formation – galaxies: haloes – galaxies: star formation.

1 INTRODUCTION

It is one of the main goals in modern galaxy formation studies to reconstruct and understand the assembly histories of galaxies (e.g. Eggen, Lynden-Bell & Sandage 1962; Searle & Zinn 1978; Ibata, Gilmore & Irwin 1994; Belokurov et al. 2006; Bell et al. 2008; Johnston et al. 2008; McConnachie et al. 2009; Cooper et al. 2010; Deason et al. 2013; Pillepich et al. 2014; Kruijssen et al. 2019a). Specifically, satellite galaxy accretion histories, often expressed in terms of merger trees, represent a clear and testable prediction of structure formation in the cold dark matter (ΛCDM) cosmology (e.g. Bullock & Johnston 2005; Deason, Belokurov & Weisz 2015; Fattahi et al. 2020). In order to reconstruct these accretion histories, it is necessary to obtain a comprehensive census of the redshifts at which satellites were accreted and the (stellar or halo)[1] masses of these systems at the time of accretion. This can be done by identifying a set of observables that traces the galaxy merger tree of the host galaxy. In the Milky Way, this has recently become possible thanks to two major developments. First, the *Gaia* satellite has provided near-complete six-dimensional (position–velocity) phase–space information for an unprecedented number of stars and stellar clusters in the Milky Way (e.g. Gaia Collaboration 2018; Baumgardt et al. 2019; Vasiliev 2019), which together provides the potential means of inferring the accretion histories of a wide variety of satellite progenitors. Secondly, the modelling frameworks have recently been developed to connect the observed phase–space (e.g. orbital) information to the properties of the progenitor satellites (e.g. Belokurov et al. 2018; Haywood et al. 2018; Helmi et al. 2018; Myeong et al. 2018; Koppelman et al. 2019a; Kruijssen et al. 2019a, b; Massari, Koppelman & Helmi 2019; Helmi 2020).

In particular, the use of globular clusters (GCs) to trace the assembly history of the Milky Way has seen an increase in applications (e.g. Forbes et al. 2018a; Helmi et al. 2018; Myeong et al. 2018, 2019; Kruijssen et al. 2019b). Various combinations of GC energies and angular momenta (i.e. orbits and integrals of motion; Helmi et al. 2018; Myeong et al. 2018), as well as GC ages and metallicities (e.g. Forbes & Bridges 2010; Leaman, VandenBerg & Mendel 2013; Li & Gnedin 2014; Choksi, Gnedin & Li 2018; Kruijssen et al. 2019b),

★ E-mail: kruijssen@uni-heidelberg.de

[1]We use the terms 'halo mass' and 'virial mass' to refer to the sum of the dark matter and baryonic mass of the galaxy within its virial radius.

have provided important constraints on the satellite population that was accreted by the Milky Way. Most recently, these efforts have been aided by the E-MOSAICS project, which is a suite of self-consistent, hydrodynamical cosmological simulations with a complete model for the formation and evolution of the GC population (Pfeffer et al. 2018; Kruijssen et al. 2019a). Because these simulations simultaneously reproduce young and old stellar cluster populations with a single, environmentally dependent cluster formation and disruption model, they enable linking the properties of the cluster population to the assembly history of the host galaxy.

In a recent paper, we used the age–metallicity distribution of Galactic GCs to infer the formation and assembly history of the Milky Way, culminating in the partial reconstruction of its merger tree (Kruijssen et al. 2019b). This work made use of the correlation in the E-MOSAICS simulations between quantities describing the formation assembly histories of galaxies (e.g. the dark matter halo concentration, the number of accretion events, the total number of progenitors, and the number of minor mergers) and the properties of their host GC populations (e.g. the number of GCs, the slope of the age–metallicity distribution, and the median age) to characterize the assembly history of the Milky Way, and additionally used the detailed distribution of Galactic GCs in age–metallicity space to derive the number of accreted satellites and their stellar mass growth histories. The main conclusions of that work are as follows:

(i) The Galactic GC age–metallicity distribution bifurcates into a steep, 'main' branch of GCs that formed *in situ* in the Main progenitor of the Milky Way and a shallow, 'satellite' branch at lower metallicities that is constituted by accreted GCs that formed in low-mass satellite galaxies. A total of $\sim$15 such satellites must have been accreted based on the number of Galactic GCs and the slope of their age–metallicity distribution, even if only a minority of these satellites is expected to have brought in detectable numbers of GCs. At least some of these accretion events may have had no associated GCs at all (also see e.g. Koppelman et al. 2019b).

(ii) The steepness of the main branch implies that the Milky Way assembled quickly for its mass, reaching $\{25,\ 50\}$ per cent of its present-day halo mass already at $z = \{3,\ 1.5\}$ and half of its present-day stellar mass at $z = 1.2$. The growth history of the Milky Way runs ahead of those typical for galaxies of its $z = 0$ mass (e.g. Papovich et al. 2015) by about 1 Gyr.

(iii) There are too many GCs on the satellite branch to be attributable to a single progenitor, because the number of GCs found in this branch is considerably larger than that expected for the low masses of the galaxies (e.g. Harris, Harris & Alessi 2013; Harris, Blakeslee & Harris 2017) forming stars at the low metallicities characterizing the satellite branch. At least two (and preferably three) massive satellites are needed to contribute most of the GCs on the satellite branch, which is supported by the existence of multiple kinematic components that were known pre-*Gaia* (including the Sagittarius dwarf galaxy; Ibata et al. 1994). It is possible that the satellite branch contains traces of a larger number of satellites (e.g. those with only 1–2 GCs, which are not detectable due to Poisson noise), but the age–metallicity distribution does not provide sufficient constraints to tell apart small subgroups. Kinematic information would be necessary to potentially lift this degeneracy.

(iv) The satellite branch in age–metallicity space is quite narrow, with a total metallicity spread of Δ[Fe/H] ≈ 0.3 dex, implying that the masses of the >3 progenitor satellites were similar at any given lookback time or redshift. For this reason, Kruijssen et al. (2019b) do not distinguish between the two most massive satellites, implying that their masses were similar to within a factor of $\sim$2.

(v) Out of the three identified satellite progenitors, the most recent (and at a given lookback time least massive) accretion event corresponds to Sagittarius (Ibata et al. 1994). The next most massive satellite brought in a large number of GCs, many of which were formerly associated with the Canis Major 'mirage', a perceived accretion event that never existed and rather represents a density wave in the Galactic disc (e.g. Martin et al. 2004; Peñarrubia et al. 2005; Deason, Belokurov & Koposov 2018; de Boer, Belokurov & Koposov 2018). The actual satellite that brought in these GCs has since been found in the *Gaia* data (Belokurov et al. 2018) and was dubbed the *Gaia* Sausage (Myeong et al. 2018) or *Gaia*-Enceladus (Helmi et al. 2018).

(vi) The third progenitor was dubbed 'Kraken' and must have had a mass very similar to that of *Gaia*-Enceladus. Until recently, it had not been found. However, in a recent paper, Massari et al. (2019) identified a group of GCs at low energies in the *Gaia* DR2 data. These GCs represent a significant fraction of the satellite branch in age–metallicity space and thus likely represent the Kraken progenitor event needed to explain the age–metallicity 'satellite branch' GCs after Sagittarius and *Gaia*-Enceladus have been accounted for.[2] It is one of the main goals of this paper to determine the mass and accretion redshift of the progenitor that brought in the 'low-energy' GCs from Massari et al. (2019) and thus assess whether this progenitor is Kraken.

In addition to Kraken, *Gaia*-Enceladus, and Sagittarius, Massari et al. (2019) find that the satellite branch accommodates GCs from two other accreted satellites, i.e. the progenitor of the 'Helmi streams' (Helmi et al. 1999) and 'Sequoia' (Myeong et al. 2019). In a recent paper, Forbes (2020) used the numbers of GCs that Massari et al. (2019) assign to each of the five satellite progenitors to estimate the galaxy masses, confirming our interpretation that the low-energy GCs match the predicted properties of Kraken. However, as we discuss in Section 3.2.1, the approach of using present-day GC numbers to estimate the host galaxy mass at accretion systematically overestimates the galaxy masses by up to a factor of 3.

In this paper, we use the groups of GCs identified by Massari et al. (2019) as having been accreted from the same satellite progenitors to determine the masses and accretion redshifts of these five galaxies. To do so, we train an artificial neural network to connect the properties of the GCs contributed by individual accreted satellites in the E-MOSAICS simulations to the properties of their host accretion events. Specifically, we use the median and interquartile ranges (IQRs) of the GC apocentre radii, eccentricities, ages, and metallicities as feature variables to predict the target variables of the accretion redshift and the host stellar mass at the time of accretion. The resulting neural network is then applied to the groups of GCs identified by Massari et al. (2019). By combining the resulting predictions with the constraints on the assembly history of the Milky

[2]Massari et al. (2019) caution that the GCs suggested by Kruijssen et al. (2019b) to have been potential members of Kraken do not quite match their kinematic selection. However, Kruijssen et al. (2019b) used kinematic information from the literature that preceded *Gaia* DR2 and therefore explicitly refrained from associating individual GCs with any particular satellites. Instead, we encouraged future studies to look for phase–space correlations between the suggested sets of GCs and those that have been proposed to be associated with particular accretion events. As such, the proposed potential member GCs merely represented 'wish lists' of interesting targets for kinematic follow-up work rather than definitive member lists. As a result, the existence (or not) of Kraken cannot be evaluated based on the possible membership of individual GCs proposed by Kruijssen et al. (2019b).

Way from Kruijssen et al. (2019b), we infer the merger tree of the Milky Way, identifying five specific accretion events. Future work applying a similar methodology to field stars is expected to extend this analysis to additional satellite progenitors.

The structure of this paper is as follows: In Section 2, we describe the procedure followed to train the neural network on the E-MOSAICS simulations. In Section 3, we present the accretion redshifts and stellar masses of the satellite progenitors, and compare the resulting properties of the progenitors to scaling relations describing galaxies and their GC populations in the nearby Universe. In Section 4, we combine this with the stellar mass growth history of the Milky Way to determine the stellar mass ratios of the accretion events and reconstruct the merger tree of the Milky Way. We present our conclusions in Section 5.

2 AN ARTIFICIAL NEURAL NETWORK PREDICTING SATELLITE MASSES AND ACCRETION REDSHIFTS FROM GC DEMOGRAPHICS

2.1 Simulation suite and training set

In this paper, we use the suite of 25 zoom-in simulations from the E-MOSAICS project (Pfeffer et al. 2018; Kruijssen et al. 2019a) to provide a training set based on which the properties of the progenitor galaxies of groups of GCs can be predicted. E-MOSAICS is a suite of hydrodynamical cosmological simulations with a complete, self-consistent model for the formation and evolution of the GC population. This suite of simulations combines the model for galaxy formation and evolution from EAGLE (Crain et al. 2015; Schaye et al. 2015) with the subgrid model for the formation and evolution of the entire stellar cluster population MOSAICS (Kruijssen et al. 2011, 2012; Pfeffer et al. 2018). All simulations adopt a ΛCDM cosmogony, described by the parameters advocated by the Planck Collaboration I (2014), namely $\Omega_0 = 0.307$, $\Omega_\mathrm{b} = 0.04825$, $\Omega_\Lambda = 0.693$, $\sigma_8 = 0.8288$, $n_\mathrm{s} = 0.9611$, $h = 0.6777$, and $Y = 0.248$.

E-MOSAICS reproduces the demographics of young cluster populations in nearby galaxies (Pfeffer et al. 2019b) (as well as predicts those of high-redshift galaxies; see Pfeffer et al. 2019a; Reina-Campos et al. 2019; Keller et al. 2020) and simultaneously reproduces a wide variety of observables describing the old GC population in the local Universe, such as the number of GCs per unit galaxy mass, radial GC population profiles, and the high-mass ($M > 10^5$ M$_\odot$) end of the GC mass function (Kruijssen et al. 2019a), as well as the mass–metallicity relation of metal-poor GCs (the 'blue tilt'; Usher et al. 2018; also see Kruijssen 2019), the GC age–metallicity distribution (Kruijssen et al. 2019b), the kinematics of GC populations (Trujillo-Gomez et al. 2020), the dynamical mass-loss histories of massive GCs (Reina-Campos et al. 2018, 2020; Hughes et al. 2020), and the association of GCs with fossil stellar streams from accreted dwarf galaxies (Hughes et al. 2019).

The goal of this work is to predict the accretion redshifts (z_acc) and the stellar masses ($M_\star$) at the time of accretion on to the Milky Way of the satellite progenitors identified by Massari et al. (2019). To this end, we train an artificial neural network with 'target' variables z_acc and $M_\star$ as a function of eight 'feature' variables that describe the properties of the GCs associated with each of these progenitors. In E-MOSAICS, the accretion redshift is defined as the moment at which SUBFIND (Springel, Yoshida & White 2001; Dolag et al. 2009) can no longer find a bound subhalo and the subhalo is therefore considered to have merged into the halo of the central halo (see Qu et al. 2017, for discussion). As feature variables, we use the medians and IQRs of the GC age (τ), GC metallicity ([Fe/H]), GC orbital apocentre radius (R_a), and GC orbital eccentricity (ϵ).[3]

Across all 25 simulations, we identify all progenitor satellites with stellar masses of $\log(M_\star/\mathrm{M}_\odot) \geq 6.5$ that host GCs and are accreted on to the central galaxy, for a total of $N_\mathrm{sat} = 205$ accretion events, or ∼8 per Milky Way-mass galaxy on average. For each of these satellite progenitors, we tabulate the target variables describing the accretion event and the feature variables describing the properties of the subpopulations of GCs contributed by these satellites. Because E-MOSAICS overpredicts the number of GCs at masses much smaller than $M \sim 10^5$ M$_\odot$ due to underdisruption (Pfeffer et al. 2018; Kruijssen et al. 2019a), we only consider GCs with $z = 0$ masses of $M \geq 5 \times 10^4$ M$_\odot$. We verified that the exact choice of this lower mass limit does not strongly affect the results of this work, because the median and IQR of GC ages, metallicities, apocentre radii, and eccentricities are not strongly correlated with the GC mass. The adopted limit of $M \geq 5 \times 10^4$ M$_\odot$ is found to provide the best compromise between minimizing the effects of underdisruption (requiring high GC masses) and having a sufficient number of GCs (requiring low GC masses). In addition, we restrict the GC metallicities to $-2.5 <$ [Fe/H] < -0.5, in order to match the range of metallicities of Galactic GCs for which age measurements are available (Marín-Franch et al. 2009; Dotter et al. 2010; Forbes & Bridges 2010; Dotter, Sarajedini & Anderson 2011; Leaman et al. 2013; Kruijssen et al. 2019b). Had we included GCs from E-MOSAICS with metallicities outside of that range, this could potentially have introduced a systematic bias in our results.

Of course, the E-MOSAICS simulations themselves only provide a finite (and quite small) sample of 25 Milky Way-mass galaxies and their accretion histories. There is clear evidence that the accretion history of the Milky Way is atypical for a galaxy of its mass (e.g. Deason, Mao & Wechsler 2016; Belokurov et al. 2018; Kruijssen et al. 2019b; Mackereth et al. 2019). Ideally, we would therefore have been able to use many more simulations, in order to ensure that the intricacies of the Milky Way's particular accretion history are captured by at least one of the galaxies in the sample. However, we do not find strong variations in the relations between feature and target variables across the suite of simulations (see Section 2.2 and Pfeffer et al. 2020), which gives some confidence that the application of these relations in this paper is robust.

2.2 Details of the neural network

We train a sequential neural network to predict the accretion redshifts and stellar masses of galaxies accreted by the Milky Way using the PYTHON packages SCIKIT-LEARN (Pedregosa et al. 2011; Buitinck et al. 2013) and KERAS (Chollet et al. 2015), which are application programming interfaces for machine learning and neural networks, respectively. We construct a neural network with 'dense' hidden

[3]These are not the only possible feature variables, but it is the combination that we found performs best. For instance, the GC age–metallicity distribution observed in the Milky Way and simulations exhibits a characteristic shape (Kruijssen et al. 2019a) that could potentially be better described by using feature variables of a higher order than just the (uncorrelated) median and IQR of the age and metallicity. However, as shown in Section 3.1, several of the GC samples contain low numbers of GCs, which prohibits the use of higher order diagnostics. We also experimented with including the vertical angular momentum (L_z) and found that this leads to considerably lower validation scores. In this case, these indicate that the addition of L_z causes the neural network to overfit the training data. We have therefore chosen to omit the L_z information.

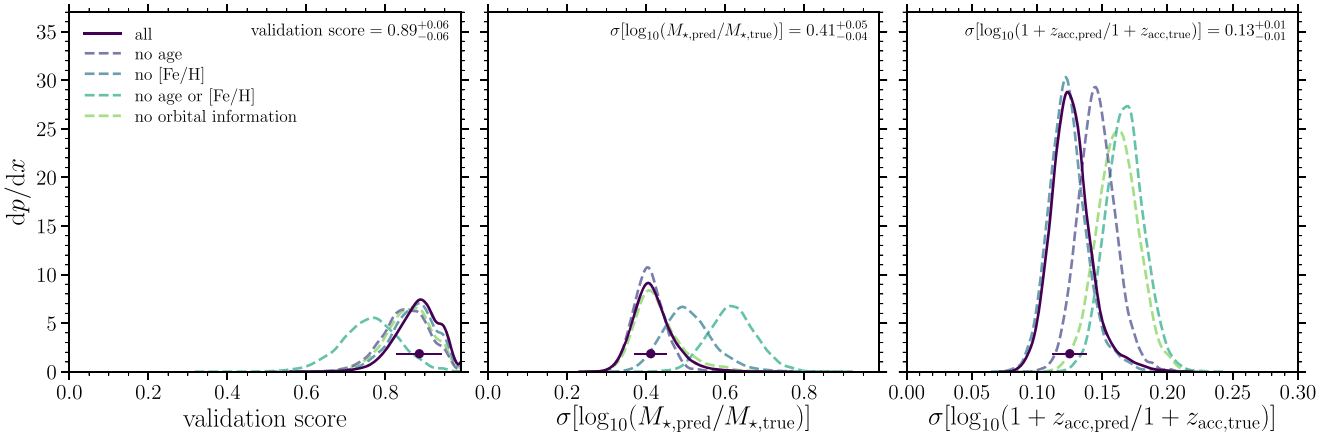

Figure 1. Validation of the neural network. Left: PDF of validation scores across all 10 000 Monte Carlo realizations of the neural network, for models including different subsets of the feature variables as indicated by the legend. The data point with error bar and the annotation in the top right indicate the median and 16–84th percentiles for the experiment including all feature variables. Middle: PDF of the logarithmic scatter of the predicted stellar masses of the satellite progenitors around their true stellar masses. Lines and annotation as in the left-hand panel. Right: PDF of the logarithmic scatter of the predicted accretion redshifts of the satellite progenitors around their true accretion redshifts. Lines and annotation as in the left-hand panel. This figure shows that the neural network performs well, predicting stellar masses ($M_\star$) within a factor of 2.5 and accretion redshifts ($1 + z_{\rm acc}$) within a factor of 1.35.

layers, i.e. one in which all nodes in successive layers are connected. The hidden layers use a Rectified Linear Unit (`ReLU`) activation function, i.e. $f(x) = \max(0, x)$. The architecture of the network is chosen by optimizing the validation score while varying the number of hidden layers in the range $N_{\rm lay} = 1$–10 and the number of nodes per layer in the range $N_{\rm node} = 10$–80. The validation score shows little variation across the parameter space probed, with a slight maximum around $N_{\rm lay} = 4$ and $N_{\rm node} = 50$ (for which the validation score is $0.89^{+0.06}_{-0.06}$; see below). While we adopt these numbers throughout this work, changing them by up to a factor of 2 does not qualitatively affect our results. The hidden layers are connected to an input layer and an output layer. By definition, the input layer consists of eight nodes (reflecting the number of feature variables) and the output layer consists of two nodes (reflecting the number of target variables). We train the neural network 10 000 times, each time adopting a different random seed and varying the hyperparameters of the network as discussed below. This Monte Carlo approach allows us to obtain probability distribution functions (PDFs) of the target variables, where the resulting dispersion reflects the uncertainties of the neural network.

We follow the standard practice of training the network on the scaled feature variables, where across the $N_{\rm sat}$ data points we use a standard scaler to subtract the mean and divide by the standard deviation of each variable. The training set is randomly split into a training subset and a testing subset. For each Monte Carlo realization of the neural network, we set the fraction of the training data that is used for testing the network by randomly drawing $f_{\rm test}$ from a flat distribution between 0.2 and 0.3. The network is compiled using the `Adam` optimizer with the default hyperparameters within KERAS and a mean squared error loss function. The neural network is fitted to the training data for a maximum of 50 epochs, but we use an early stopping monitor with a `patience` value of 3. This means that the fitting loop is stopped when the validation score of the neural network does not improve for three successive epochs, so that in practice 50 fitting epochs are never reached. The model with the highest validation score is saved as a checkpoint and used when applying the model. The validation score is calculated by separating off a validation subset that is a fraction $f_{\rm val}$ of the training subset, where we randomly draw $f_{\rm val}$ from a flat distribution between 0.1 and 0.2 for each Monte Carlo realization of the neural network. This means that the validation subset consists of $f_{\rm val}(1 - f_{\rm test})N_{\rm sat}$ accretion events. We then test the neural network by using it to predict the target variables for the test subset and comparing them to the true values. The above procedure is repeated for each of the 10 000 Monte Carlo realizations. Throughout this paper, uncertainties on quoted numbers reflect the 16th and 84th percentiles of PDFs resulting (or propagated) from these 10 000 realizations. By drawing $f_{\rm test}$ and $f_{\rm val}$ from the aforementioned flat distributions for each Monte Carlo realization, the uncertainties on the target variables account for the effects of changing these hyperparameters. Finally, we repeat the entire process when omitting certain subsets of the feature variables, i.e. when omitting (1) GC age information, (2) GC metallicity information, (3) GC age and metallicity information, and (4) GC orbital information. The goal of carrying out these additional experiments is to identify which of the feature variables are the best predictors of each of the target variables and should thus always be included in observational applications of these neural networks.

Fig. 1 shows the results of the above experiments.[4] When including all feature variables, the neural network achieves a satisfactory validation score of $0.89^{+0.06}_{-0.06}$. This matches the typical training score to within the uncertainties, indicating that the model is not underfitting or overfitting. Omitting any of the feature variables results in lower validation scores, even if their medians are generally consistent with the validation score for all feature variables to within the scatter. The neural network only performs significantly worse when omitting the GC age and metallicity information, with a validation score of $0.76^{+0.07}_{-0.07}$. This shows that the GC age–metallicity distribution encodes crucial information on the stellar masses and accretion redshifts of the Milky Way's satellite progenitor galaxies. This lends further support to the rich body of literature identifying GC age–metallicity space as an important tracer of galaxy assembly (e.g. Forbes & Bridges 2010; Leaman et al. 2013; Choksi et al. 2018; Kruijssen et al. 2019b).

[4]Throughout this paper, PDFs reflect the 10 000 Monte Carlo realizations of the neural network. All PDFs are smoothened using the default kernel density estimator in the SEABORN PYTHON package (Waskom et al. 2020).

We assess the precision with which the neural network predicts the stellar masses and accretion redshifts of the satellite progenitors by calculating the logarithmic standard deviation around the one-to-one relation for the test subset in each Monte Carlo realization. The resulting scatter on the stellar mass is $\sigma[\log_{10}(M_{\star,\mathrm{pred}}/M_{\star,\mathrm{true}})] = 0.41^{+0.05}_{-0.04}$ when including all feature variables. This precision enables making quantitative predictions for the satellite progenitor masses. As before, omitting the age–metallicity information leads to a significantly worse precision, with $\sigma[\log_{10}(M_{\star,\mathrm{pred}}/M_{\star,\mathrm{true}})] = 0.62^{+0.06}_{-0.06}$. Interestingly, this is mostly driven by the GC metallicities, as omitting these leads to a scatter of $\sigma[\log_{10}(M_{\star,\mathrm{pred}}/M_{\star,\mathrm{true}})] = 0.50^{+0.07}_{-0.06}$, whereas omitting only the GC age information barely affects the scatter around the one-to-one relation for true versus predicted satellite progenitor mass. We conclude that the GC metallicities, in combination with either GC ages or orbital information, are critical for constraining the stellar masses of the accreted galaxies.

For the accretion redshifts, we achieve a precision of $\sigma[\log_{10}(1 + z_{\mathrm{acc,pred}}/1 + z_{\mathrm{acc,true}})] = 0.13^{+0.01}_{-0.01}$ or about 30 per cent on $1 + z_{\mathrm{acc}}$. Accurate predictions for the satellite accretion redshift require nearly all information to be used, with the exception of the GC metallicities. Omitting any other feature variable (GC ages or GC orbital information) leads to significantly worse constraints on the accretion redshift, with $\sigma[\log_{10}(1 + z_{\mathrm{acc,\,pred}}/1 + z_{\mathrm{acc,\,true}})] > 0.16$. This shows that the accretion redshift is the most challenging of the two target variables to constrain. The reason it depends on GC ages is obvious – satellites that were accreted early must have older GCs. The strong dependence on the orbital information is a bit more subtle, but also works as expected. Satellites that were accreted early deposit their GCs at small (apocentre) radii and these GCs end up with a large spread in eccentricities by $z = 0$ (Pfeffer et al. 2020).

We do not explicitly consider the uncertainties on the observed feature variables, but find that their impact must be small. The GC ages carry the most significant uncertainty, which implies that the impact of the measurement uncertainties must be smaller than omitting the age information altogether (see Fig. 1). We can quantify this as follows. The uncertainty on the median age of a sample of size N is $\sigma_{\mathrm{med}} = [\pi/2(N-1)]^{1/2}\sigma_{\mathrm{age}}$. For a typical uncertainty of $\sigma_{\mathrm{age}} = 0.5$ Gyr and a typical number of $N = 5$–20 GCs per satellite progenitor, we obtain $\sigma_{\mathrm{med}} = 0.1$–$0.3$ Gyr, which is much smaller than the range of median GC ages considered in this work (both in the simulations and the observational sample). The same applies for the IQR, which is larger than 0.5 Gyr for all GC samples. We thus consider the hyperparameters to be the dominant source of uncertainty, as quantified in Fig. 1.

In summary, the inclusion of the complete GC age–metallicity information guarantees the most precise model predictions in general. The orbital information plays an important role in helping constrain the accretion redshift. In the following sections, we apply the model to the GC population of the Milky Way, for which all eight feature variables are known.

3 RECONSTRUCTING THE PROGENITOR SATELLITE POPULATION OF THE MILKY WAY

3.1 Definition of the observational sample

Massari et al. (2019) combine the orbital information of the Galactic GC system with their age–metallicity distribution to identify subsets of GCs that likely originated in a common progenitor. By dividing the GC population in this way, they identify five groups of GCs that plausibly have an extragalactic origin. Four of these they associate with known accretion events, i.e. the *Gaia*-Enceladus–Sausage event (Belokurov et al. 2018; Helmi et al. 2018; Myeong et al. 2018), the progenitor of the Helmi et al. (1999) streams, the Sequoia accretion event (Myeong et al. 2019), and Sagittarius (Ibata et al. 1994). In addition, they identify a group of GCs at low energies (i.e. at high binding energies), which is similar in number to the GCs associated with the *Gaia*-Enceladus event.[5] While Massari et al. (2019) are unable to draw firm conclusions regarding the origin of these low-energy GCs, we demonstrate below that this group represents the Kraken accretion event predicted by Kruijssen et al. (2019b). After identifying these five groups, Massari et al. (2019) combine the remaining GCs with known ages, metallicities, and orbits in a 'high-energy' group, which are distributed across parameter space and are unlikely to have originated in a common progenitor. Instead, they could represent an ensemble of low-mass accretion events that each contributed one or two GCs. We therefore omit the high-energy group from our further analysis.

We use the GC ages and metallicities from the compilation of Kruijssen et al. (2019b, who combined literature measurements from Forbes & Bridges 2010, Dotter et al. 2010, 2011, and VandenBerg et al. 2013) and the orbital properties from Baumgardt et al. (2019). For the five satellite progenitors, we adopt the GC membership selection from Massari et al. (2019), with a small number of changes. First, they consider Pal 1 to have formed *in situ* (i.e. in the 'Main progenitor'). However, given its position in age–metallicity space, it unambiguously belongs to the satellite branch, having a relatively low metallicity of $[\mathrm{Fe/H}] = -0.7$ at an age of only 7.3 Gyr. In order to have formed within the Main progenitor, Pal 1 would have needed to have had roughly solar metallicity (Haywood et al. 2013). Because it is not known to which satellite progenitor Pal 1 should be attributed, we omit if from our analysis. Conversely, Massari et al. (2019) associate NGC 6441 and E3 with the low-energy group and (possibly) the progenitor of the Helmi streams, respectively. However, based on their high metallicities ($[\mathrm{Fe/H}] = -0.6$ and -0.83) at old ages ($\tau = 11.3 \pm 0.9$ and 12.8 ± 1.4 Gyr) these GCs must have formed in the Main progenitor. We therefore classify these GCs as ambiguous, possibly being members of the low-energy group or the Main progenitor, and the Helmi streams or the Main progenitor, respectively. In practice, this means that we exclude E3 from our analysis (also because of its low mass), and consider versions of the low-energy group both including and excluding NGC 6441. Finally, Horta et al. (2020b) argue that NGC 6121, which Massari et al. (2019) associate with the low-energy group, has an *in-situ* origin. Based on the age and metallicity of NGC 6121 ($\tau = 12.2 \pm 0.5$ Gyr and $[\mathrm{Fe/H}] = -1.14$), we consider it too metal poor to have formed *in situ* and follow the choice of Massari et al. (2019) to associate it with the low-energy group, even if we acknowledge that this is an edge case. The results presented in this paper are unaffected by this choice.

Before proceeding, we note that not all memberships proposed by Massari et al. (2019) are unambiguous. Several GCs have dual

[5] At first inspection, the critical binding energy used by Massari et al. (2019) to separate the GCs belonging to *Gaia*-Enceladus and the low-energy group may seem somewhat arbitrary. However, Horta et al. (2020a) recently showed that the accreted population of metal-poor ($[\mathrm{Fe/H}] < -0.8$) stars in the bulge shows a clear dearth of stars between the *Gaia*-Enceladus and low-energy groups in energy–angular momentum space, exactly at the point where Massari et al. (2019) separated both groups of GCs. Furthermore, the satellite branch of GCs in age–metallicity space (see the discussion below) contains too many GCs at comparatively low metallicities to have been populated by a single massive satellite progenitor (Kruijssen et al. 2019b). Therefore, we conclude that the division between the *Gaia*-Enceladus and low-energy GCs is well motivated empirically.

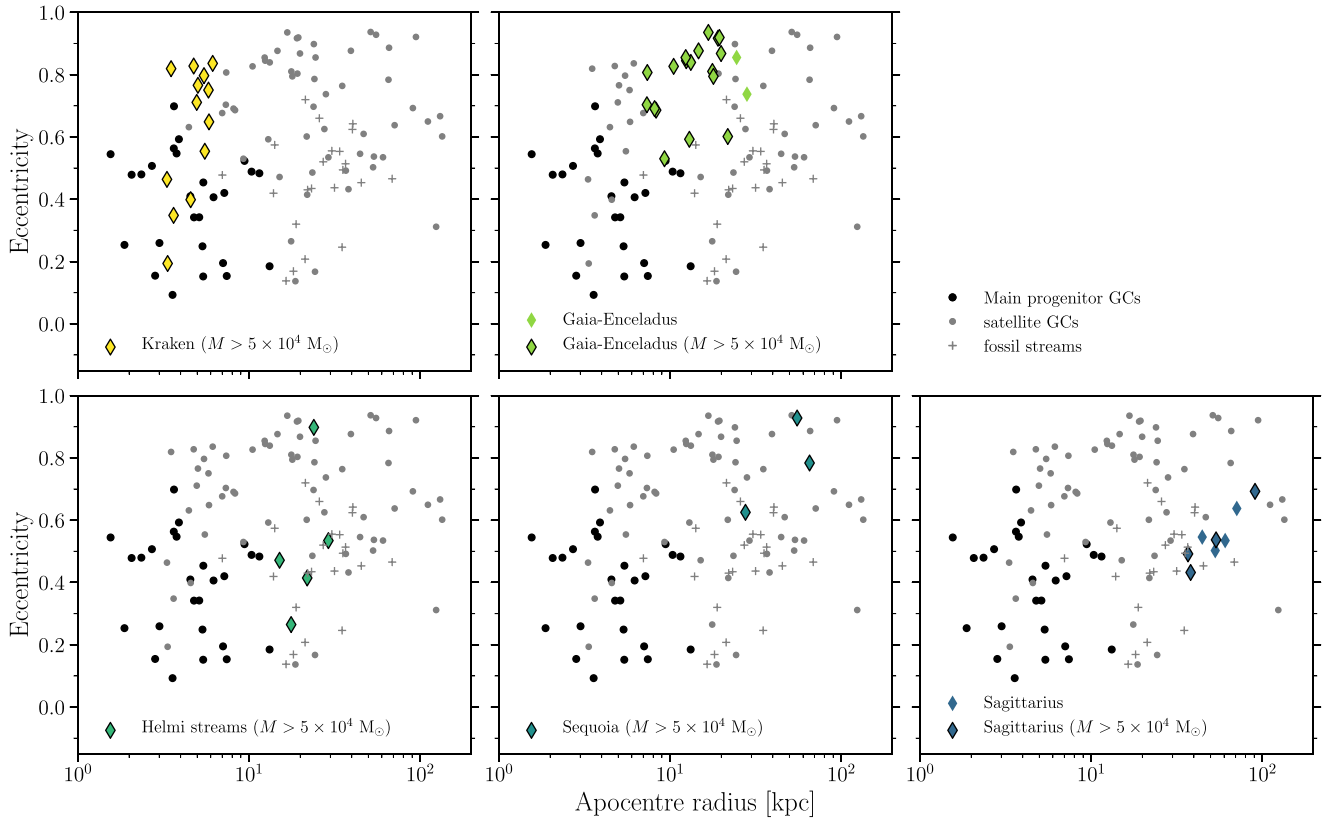

Figure 2. Orbital properties of the Galactic GC population, expressed in terms of their apocentre radii and eccentricities. In all the panels, we distinguish GCs that formed in the Main progenitor and those that formed in satellites (largely following Massari et al. 2019, with changes as discussed in the text), as well as fossil streams that are likely relics of GCs (taken from Bonaca & Kruijssen, 2020), which we do not associate to any specific accretion event here. In each of the panels, the unambiguous members of each satellite progenitor are highlighted, demonstrating that GCs associated with different satellite progenitors have different orbital properties.

associations (e.g. 'low-energy/Sequoia'). We restrict our fiducial analysis to the GCs that are unambiguously associated with a single satellite progenitor. However, throughout this paper we systematically consider all possible membership permutations, i.e. all possible assignments of ambiguous GCs to the satellite progenitors that Massari et al. (2019) identify as possible hosts. The goal of this procedure is to demonstrate how the uncertainties in GC membership propagate into uncertainties on the stellar masses and accretion redshifts of the satellite progenitors. From here on, we also refer to the low-energy GCs of Massari et al. (2019) as Kraken GCs.

We show the distributions of Galactic GCs in the plane spanned by apocentre radius and eccentricity in Fig. 2 and in the plane spanned by age and metallicity in Fig. 3, colour coded by their memberships of each of the five satellite progenitors. In the colour coding, we only include the unambiguous memberships. In addition, here and in the subsequent analysis we include observed GCs with masses of $M < 5 \times 10^4$ M$_\odot$, contrary to the selection of the training set. This mass cut only needs to be applied to the simulations, because it compensates the underdisruption of GCs in E-MOSAICS – real-Universe GCs do not suffer from this problem, implying that a mass cut is not required. The inclusion of low-mass GCs has the additional benefit that the statistics of the observational samples improve somewhat. We do not expect the inclusion of low-mass GCs to bias the predictions of the neural network, because the masses of Galactic GCs are not correlated with age, metallicity, or orbital parameters.

In Fig. 2, the five groups of GCs occupy distinct parts of apocentre–eccentricity space. The Kraken GCs occupy the smallest radii, but a considerably wider range of eccentricities than the other satellites or even than the Main progenitor. This suggests that Kraken was massive and accreted early (Pfeffer et al. 2020). The *Gaia*-Enceladus GCs occupy intermediate radii and high eccentricities, whereas the GCs associated with the Helmi streams orbit at intermediate radii and (mostly) intermediate eccentricities. Finally, the Sequoia and Sagittarius GCs have large apocentre radii, with high and intermediate eccentricities, respectively. These differences are suggestive of differences in origin – the orbital characteristics of each group of GCs reflect the orbital properties of the accretion events, which in turn trace the masses and accretion redshifts of the satellite progenitors.

For reference, Fig. 2 also includes the orbital properties of several fossil stellar streams in the Galactic halo, taken from Bonaca & Kruijssen (2020, and originally discovered by Balbinot et al. 2016, Shipp et al. 2018, and Ibata, Malhan & Martin 2019), which plausibly originated from disrupted GCs. The reasonable correspondence between the apocentre–eccentricity distribution of these streams and that of the GCs associated with each of the five progenitor satellites implies that several of these streams may be associated with the same progenitors. In general, the streams seem to be relics of GCs with an *ex-situ* origin, as there is little correspondence with the orbital properties of GCs that formed in the Main progenitor. The orbital eccentricities of most streams do not reach values as high as those of *Gaia*-Enceladus and Sequoia, but instead occupy the low-to-intermediate eccentricity range. We find one stream (Fimbulthul) that may possibly be a relic of a GC that formed in Kraken (here

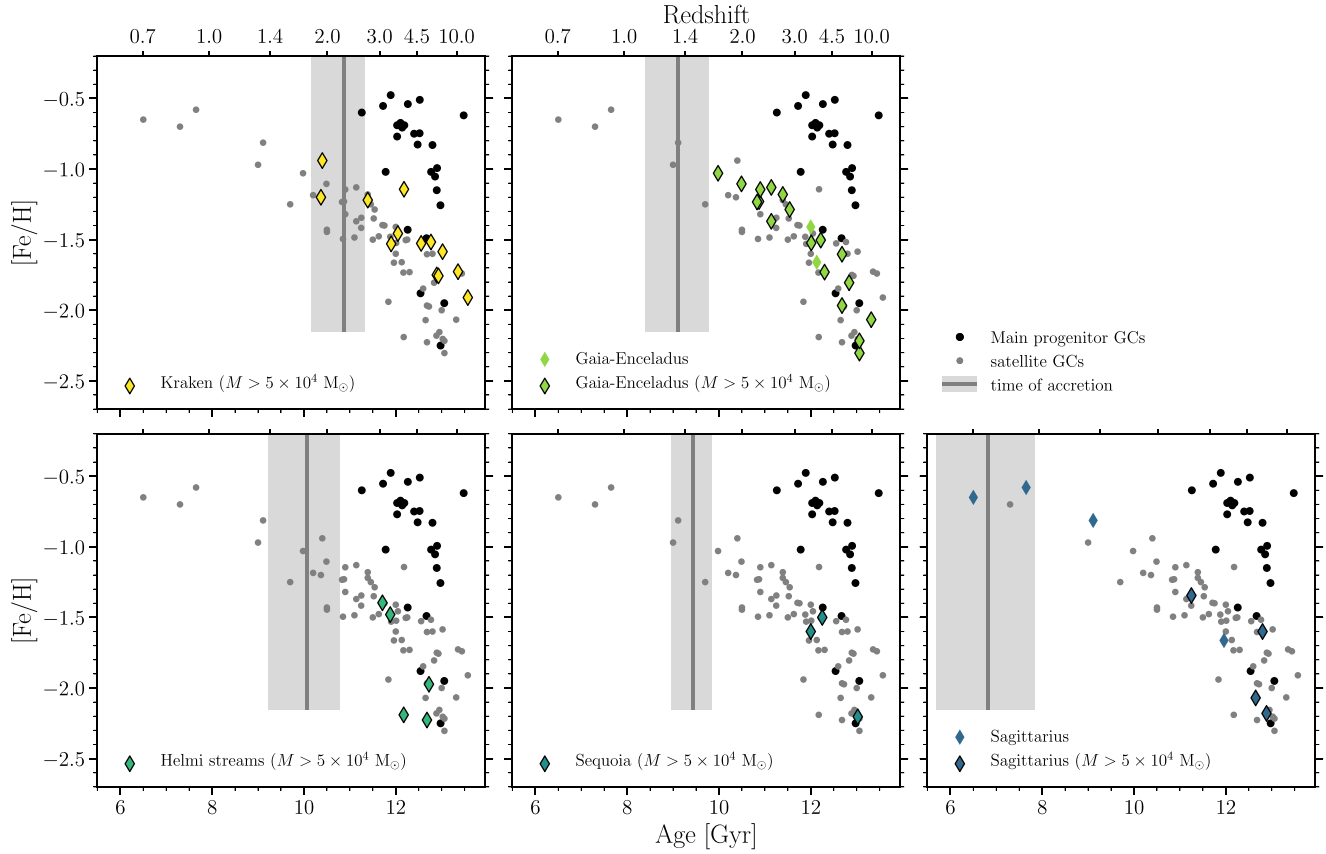

Figure 3. Age–metallicity distribution of the Galactic GC population. In all the panels, we distinguish GCs that formed in the Main progenitor and those that formed in satellites (largely following Massari et al. 2019, with changes as discussed in the text). In each of the panels, the unambiguous members of each satellite progenitor are highlighted, demonstrating that GCs associated with different satellite progenitors follow different tracks in age–metallicity space. Each panel also includes a vertical line indicating the time of accretion inferred in Section 3.2, with the shaded band representing the 1σ uncertainty.

selected using $R_a \lesssim 7$ kpc). Three streams (Ophiuchus, Phlegethon, and Wambelong) may be associated with *Gaia*-Enceladus, given their high orbital eccentricities ($\epsilon \gtrsim 0.55$) and intermediate radii ($7 \lesssim R_a/\mathrm{kpc} \lesssim 30$). A further five streams (Fjörm, GD-1, Ravi, Sylgr, and Ylgr) orbit at radii ($15 \lesssim R_a/\mathrm{kpc} \lesssim 30$) and eccentricities ($0.2 \lesssim \epsilon \lesssim 0.55$) similar to those of the GCs associated with the Helmi streams, suggesting that these five streams could have originated from GCs that were once part of the Helmi et al. (1999) satellite progenitor. There are no streams unambiguously associated with the Sequoia GCs, except possibly Gjöll, Jhelum, Phlegethon, Slidr, and Turranburra (this depends critically on the membership of IC 4499). Finally, nine streams (ATLAS, Aliqa Uma, Elqui, Gjöll, Jhelum, Leiptr, Slidr, Turbio, and Turranburra) have apocentre radii of $R_a \gtrsim 30$ kpc and eccentricities of $0.4 \lesssim \epsilon \lesssim 0.7$, suggesting that they could be disrupted GCs that were brought in by Sagittarius. We refer to Bonaca & Kruijssen (2020) for more details on the properties of the fossil streams and conclude this brief discussion by emphasizing that the selection performed here is deliberately generous – the inclusion of additional selection criteria may further trim the sample of streams associated with the satellite progenitors.

In Fig. 3, we show the age–metallicity distribution of the GC sample. As for Fig. 2, it is immediately clear that the GCs associated with the five satellite progenitors occupy a different part of the plane than those that formed *in situ* within the Main progenitor. The figure also shows that there are quantitative differences between the individual satellites. At any given age, the Kraken GCs are the most metal rich, which suggests that Kraken was the most massive satellite at any given moment in time prior to its accretion on to the Milky Way (in terms of both its stellar and halo mass).[6] The second most massive satellite at any given time is *Gaia*-Enceladus, with the progenitor of the Helmi streams, Sequoia, and Sagittarius, likely having had lower masses at any time, based on the fact that the metallicities of their GCs are generally lower and their GC populations are less numerous at young ages ($\tau < 11$ Gyr). Finally, we see no GCs with ages younger than the accretion redshifts inferred in Section 3.2, which is an important consistency check. We do note that the most massive satellites (Kraken, *Gaia*-Enceladus, and Sagittarius; see Section 3.2.1) are able to form GCs all the way up to their time of accretion, whereas the lower mass satellites (the progenitor of the Helmi streams and Sequoia; also see Section 3.2.1) have GC formation truncated earlier. This may happen because the low-mass satellites are getting disrupted more rapidly after they enter the Galactic halo, whereas the massive satellites survive longer.

[6]Despite possibly being the most massive at any time prior to its accretion, Kraken need not be the most massive satellite that the Milky Way ever accreted, if it was accreted significantly earlier than the other satellites. This would truncate its growth, whereas satellites that were accreted later would be able to continue growing their masses. See Section 3.2 for details.

Table 1. GC membership adopted in this work. For each satellite progenitor, we consider all possible membership permutations throughout our analysis. The 'Abbreviation' column lists the shorthand used to refer to these subsets of GCs in the figure legends.

Possible progenitors	Abbreviation	GCs
Kraken	–	NGC 5946, NGC 5986, NGC 6093, NGC 6121, NGC 6144, NGC 6254, NGC 6273, NGC 6287, NGC 6541, NGC 6544, NGC 6681, NGC 6712, NGC 6809
Kraken/Main progenitor	Kraken/Main	NGC 6441
Kraken/Sequoia	Kraken/Seq	NGC 6535
Gaia-Enceladus	G-E	NGC 288, NGC 362, NGC 1261, NGC 1851, NGC 1904, NGC 2298, NGC 2808, NGC 4147, NGC 4833, NGC 5286, NGC 5897, NGC 6205, NGC 6235, NGC 6284, NGC 6341, NGC 6779, NGC 6864, NGC 7089, NGC 7099, NGC 7492
Gaia-Enceladus/Sequoia	G-E/Seq	NGC 5139
Helmi streams	H99	NGC 4590, NGC 5024, NGC 5053, NGC 5272, NGC 6981
Helmi streams/*Gaia*-Enceladus	H99/G-E	NGC 5634, NGC 5904
Sequoia	Seq	NGC 5466, NGC 7006, IC4499
Sequoia/*Gaia*-Enceladus	Seq/G-E	NGC 3201, NGC 6101
Sagittarius	–	NGC 2419, NGC 5824, NGC 6715, Pal 12, Terzan 7, Terzan 8, Arp 2, Whiting 1

3.2 Application of the neural network

We take the neural network described in Section 2 and apply it to the GC populations associated with each of the satellite progenitors to obtain their stellar masses and accretion redshifts. We list the adopted memberships in Table 1. As discussed above, some of the memberships are ambiguous. Throughout the majority of the following discussion, we consider all possible permutations that at least include the unambiguous members of each satellite progenitor. For progenitors that are listed $\{1, 2, 3, 4\}$ times in Table 1, this means that we need to consider $\{1, 2, 4, 8\}$ possible permutations, for a total of 23 permutations across all five satellite progenitors.

For each Monte Carlo realization of the neural network and for each membership permutation, we subject the eight feature variables (medians and IQRs of the GC age, metallicity, apocentre radius, and eccentricity) to the same scaler transform as we did for the training set in Section 2.2, i.e. we subtract the mean of the training set and divide by the standard deviation of the training set. This transforms the GC properties to the same coordinate space used to train the network. We then use the neural network to predict the two target variables (satellite progenitor stellar mass and its accretion redshift). This results in 10 000 predictions for each target variable and for each GC membership permutation.

3.2.1 Stellar masses at the time of accretion

Fig. 4 shows the PDFs of the stellar mass of the satellite progenitors at the time of accretion (see Section 2.1 for its definition). The satellite progenitors span a relatively narrow range of stellar masses at the time of accretion, of a factor of 3–4. As expected, Kraken and *Gaia*-Enceladus are the main accretion events, with stellar masses of $M_\star = 1.9^{+1.0}_{-0.6} \times 10^8$ and $2.7^{+1.1}_{-0.8} \times 10^8$ $\mathrm{M}_\odot$, respectively. These masses exceed those of the progenitor of the Helmi streams and Sequoia, which had stellar masses of $M_\star = 0.9^{+0.5}_{-0.3} \times 10^8$ and $0.8^{+0.2}_{-0.2} \times 10^8$ $\mathrm{M}_\odot$, respectively. Somewhat surprisingly, Sagittarius is predicted to have had a mass of $M_\star = 2.8^{+1.8}_{-1.1} \times 10^8$ $\mathrm{M}_\odot$, similar to Kraken and *Gaia*-Enceladus when they merged, despite the Sagittarius GCs having lower metallicities when considering the same GC age interval. This high mass is enabled by the late time of its accretion, giving it more time to grow its stellar mass than Kraken and *Gaia*-Enceladus had before they got cannibalized (see Section 3.2.2).

The stellar masses that we infer for each of the satellite progenitors are only weakly affected by the GC membership permutation. The total spread of the median falls within the uncertainties on the prediction for all progenitors except Sequoia, for which the inclusion of ambiguous GCs can increase its mass by up to 0.2 dex. This is not surprising, because a large fraction (40 per cent) of the GCs potentially associated with Sequoia is ambiguous. For all other progenitors, the impact of the membership ambiguity is smaller and usually of the order of 0.1 dex.

Relative to other measurements in the literature, we find a lower mass for *Gaia*-Enceladus than Helmi et al. (2018), who provide a very rough estimate of $M_\star \sim 6 \times 10^8$ $\mathrm{M}_\odot$ from the star formation rate ($\sim$0.3 $\mathrm{M}_\odot$ yr^{-1}) and duration ($\sim$2 Gyr) necessary to reproduce the α-poor stellar population (Fernández-Alvar et al. 2018). The back-of-the-envelope nature of this estimate makes it quite uncertain and, depending on the membership of the α-poor stellar population, it may represent an upper limit. We therefore consider our predicted stellar mass broadly compatible with the estimate of Helmi et al. (2018), but point out that the prediction made here is likely to be more accurate, as well as more meaningful thanks to the inclusion of error bars on the predictions. In Kruijssen et al. (2019b), we estimated that both Kraken and *Gaia*-Enceladus had masses as high as 10^9 $\mathrm{M}_\odot$ at the time of accretion, based on the age–metallicity–mass distribution of central galaxies in the EAGLE simulation. In this paper, we place them at a lower mass, because the satellites that are accreted on to Milky Way-mass centrals have lower masses than field dwarf galaxies, as their star formation may be halted soon after falling into the halo. The neural network trained here automatically accounts for this bias, which was left unaccounted for by Kruijssen et al. (2019b).

The stellar masses inferred for the final three satellite progenitors also agree with previous literature results. Koppelman et al. (2019a) estimate that the progenitor of the Helmi streams had a mass of $M_\star \sim 10^8$ $\mathrm{M}_\odot$ when it accreted, which is entirely consistent with our prediction. Likewise, Myeong et al. (2019) estimate that Sequoia had a stellar mass at the time of accretion of $M_\star = 1.7 \times 10^8$ $\mathrm{M}_\odot$, with a factor-of-few uncertainty, again consistent with the mass reported above. Finally, Niederste-Ostholt et al. (2010) and Niederste-Ostholt, Belokurov & Evans (2012) find that the stellar mass of Sagittarius at the time of accretion was $M_\star = (2.0\text{–}2.9) \times 10^8$ $\mathrm{M}_\odot$ (assuming a mass-to-light ratio of $M/L = 2$ $\mathrm{M}_\odot$ $\mathrm{L}_\odot^{-1}$), compatible with our prediction. Sagittarius is by far the best-studied accretion event out of the five considered here. The fact that the neural network's prediction for the mass of this galaxy is in such good agreement with independent mass estimates from the literature adds credence to the network's predictions for the masses of the

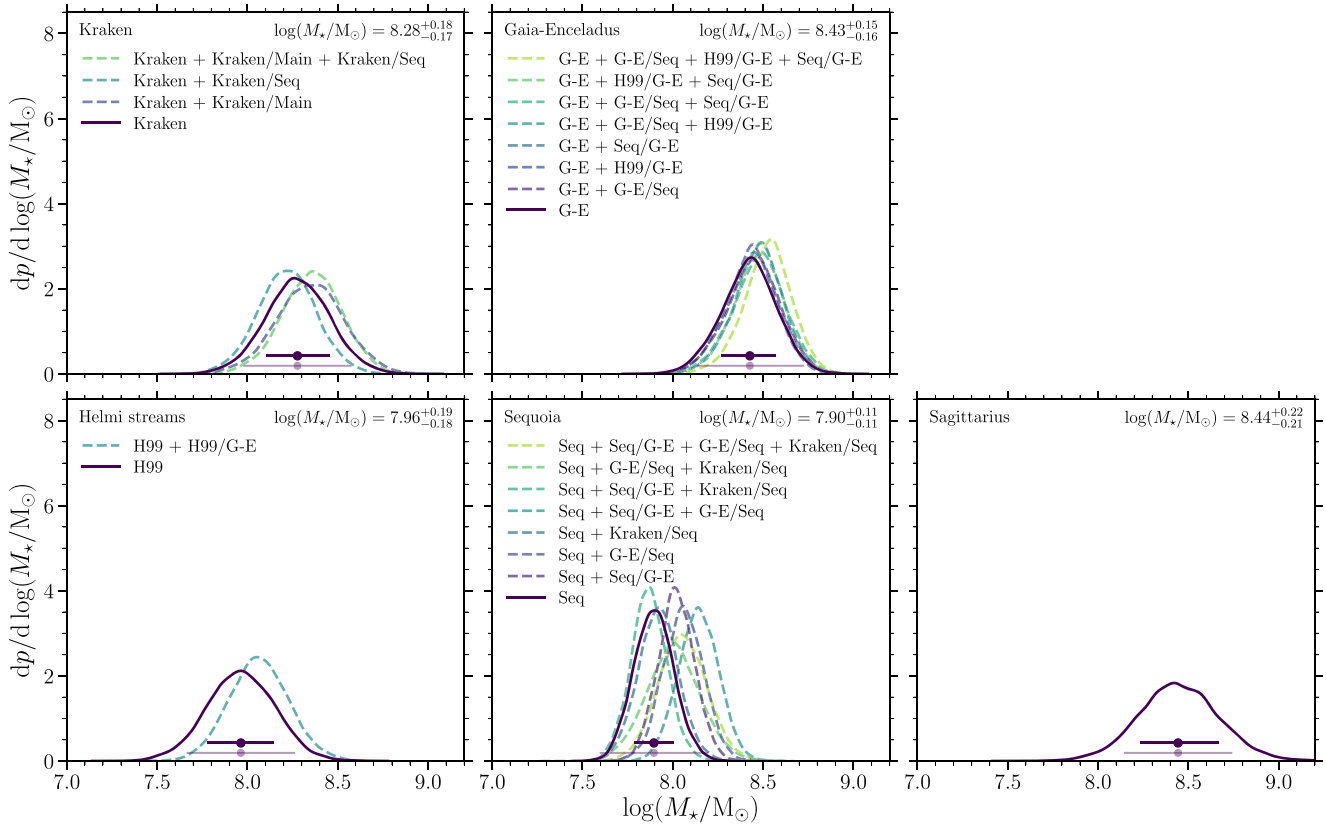

Figure 4. PDFs of the predicted stellar masses of the satellite progenitors, inferred by applying a neural network trained on the E-MOSAICS simulations to their GC populations. Each PDF shows the distribution across all 10 000 Monte Carlo realizations of the neural network, applied to different GC membership permutations as indicated by the legend (also see Table 1). In each panel, the solid data point with error bar and the annotation in the top right indicate the median and 16–84th percentiles for the experiment using the unambiguous member GCs (corresponding to the solid line in that panel). The transparent data point with error bar indicates the systematic uncertainty that applies to the full sample of progenitors. This figure shows that the five satellite progenitors span a relatively narrow stellar mass range of $M_\star = (0.6\text{–}4.6) \times 10^8$ M$_\odot$.

other satellite progenitors (as well as their accretion redshifts; see Section 3.2.2).

Forbes (2020) estimates the stellar masses of the satellite progenitors considered here by using the total number of GCs (including ambiguous members) as a probe of the halo mass and converting it to a stellar mass by adopting a stellar mass–halo mass relation. This makes the strong assumption that the relation between the number of GCs and the halo mass at $z = 0$ does not evolve with redshift. If the number of GCs per unit halo mass is higher at high redshift (as suggested by, e.g. Kruijssen 2015; Choksi & Gnedin 2019; El-Badry et al. 2019; Bastian et al. 2020), this assumption overestimates the galaxy mass. Likewise, the inclusion of ambiguous GCs as members also maximizes the galaxy mass. For these reasons, the resulting estimates of the stellar masses represent (quite uncertain) upper limits, because they effectively represent projected masses at $z = 0$ rather than at the time of accretion. For Kraken, *Gaia*-Enceladus, the progenitor of the Helmi streams, Sequoia, and Sagittarius, Forbes (2020) estimates $\log(M_\star/\mathrm{M}_\odot) = \{8.7, 8.9, 7.9, 7.9, 7.9\}$. Out of these, the masses of the progenitor of the Helmi streams and Sequoia are consistent with the masses derived here. The masses of Kraken and *Gaia*-Enceladus are larger, most likely due to the biases described above. The mass of Sagittarius is lower than both the result of Niederste-Ostholt et al. (2012) and the mass derived in this paper, possibly because the sample of GCs attributed to Sagittarius may be incomplete.

We conclude the discussion of Fig. 4 by pointing out that the quoted error bars on the predictions reflect random uncertainties. A comparison to the standard deviation shown in the middle panel of Fig. 1 (which is larger than the typical error bar) suggests the existence of an additional systematic uncertainty that may affect all stellar mass measurements by up to 0.3 dex.

3.2.2 *Accretion redshifts*

Fig. 5 shows the PDFs of the accretion redshifts of the satellite progenitors, which is defined in E-MOSAICS as the moment at which we can no longer detect a gravitationally bound subhalo (see Section 2.1). As is immediately obvious from the figure, the satellite progenitors span a wide range of accretion redshifts. Chronologically, Kraken was the first galaxy to be accreted ($z_\mathrm{acc} = 2.26^{+0.39}_{-0.45}$ or $t_\mathrm{acc} = 10.9^{+0.4}_{-0.7}$ Gyr), followed by the progenitor of the Helmi streams ($z_\mathrm{acc} = 1.75^{+0.42}_{-0.37}$ or $t_\mathrm{acc} = 10.1^{+0.7}_{-0.9}$ Gyr). Sequoia ($z_\mathrm{acc} = 1.46^{+0.17}_{-0.17}$ or $t_\mathrm{acc} = 9.4^{+0.4}_{-0.5}$ Gyr) and *Gaia*-Enceladus ($z_\mathrm{acc} = 1.35^{+0.26}_{-0.23}$ or $t_\mathrm{acc} = 9.1^{+0.7}_{-0.7}$ Gyr) accreted at approximately the same time, but still well before Sagittarius ($z_\mathrm{acc} = 0.76^{+0.22}_{-0.19}$ or $t_\mathrm{acc} = 6.8^{+1.1}_{-1.1}$ Gyr), which is the final accretion event considered here. For all five satellite progenitors, the lookback time of accretion is consistent with (or larger than) the age of the youngest associated GC, which provides an important consistency check (see Fig. 3).

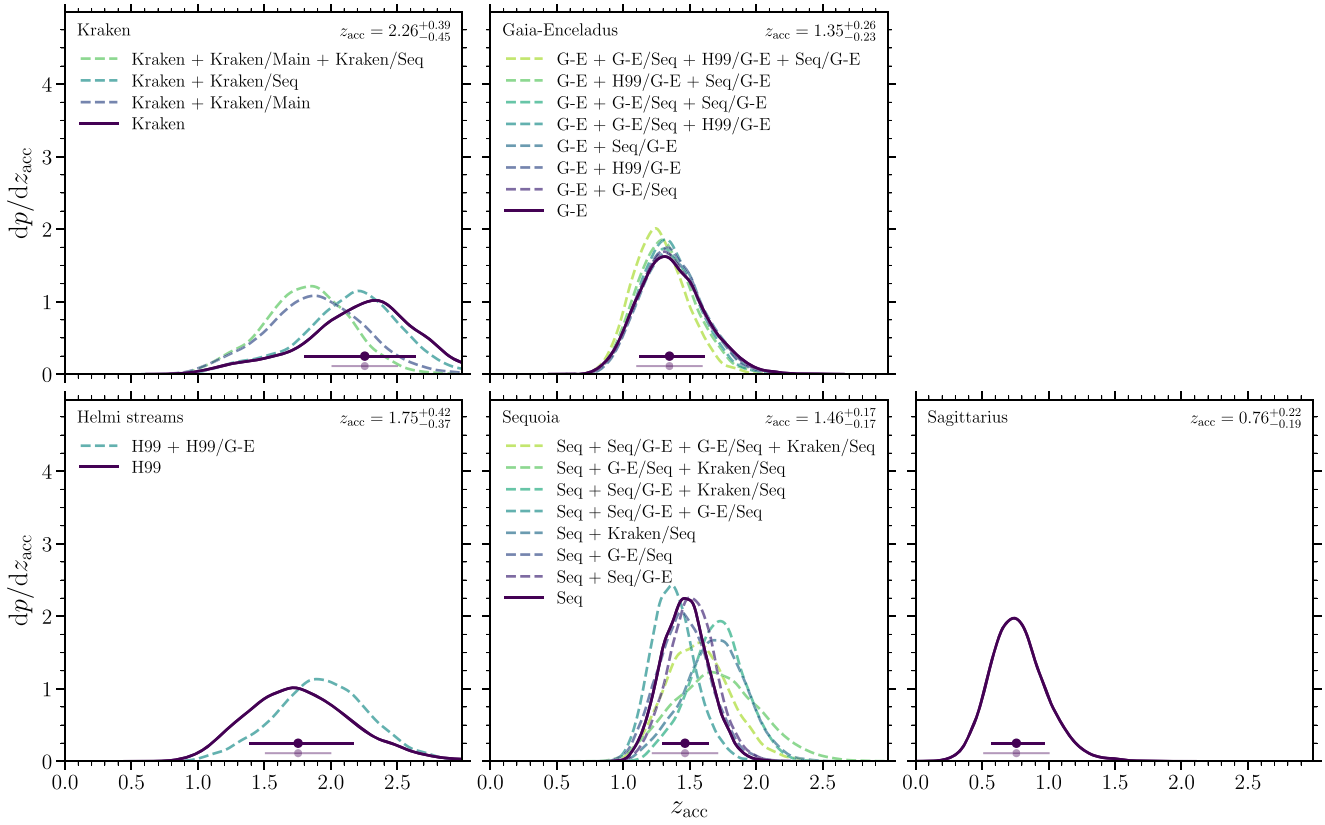

Figure 5. PDFs of the predicted accretion redshifts of the satellite progenitors, inferred by applying a neural network trained on the E-MOSAICS simulations to their GC populations. Each PDF shows the distribution across all 10 000 Monte Carlo realizations of the neural network, applied to different GC membership permutations as indicated by the legend (also see Table 1). In each panel, the solid data point with error bar and the annotation in the top right indicate the median and 16–84th percentiles for the experiment using the unambiguous member GCs (corresponding to the solid line in that panel). The transparent data point with error bar indicates the systematic uncertainty that applies to the full sample of progenitors. This figure shows that the five satellite progenitors accreted over a wide redshift range of $z_{\rm acc} = 0.57$–2.65, corresponding to lookback times of $t_{\rm acc} = 5.7$–11.3 Gyr.

The large variety of accretion redshifts explains an apparent inconsistency that appeared above. In the discussion of Fig. 3, we previously inferred the mass-ranked order of satellites at a given age from their metallicity offsets and suggested that Kraken was the most massive satellite, whereas Sagittarius was one of the lowest mass ones. As discussed above, this differs from their inferred mass-ranked order at the time of their accretion (see Fig. 4). We now see that this difference arises because satellites that were accreted early had their mass growth cut short, whereas those that were accreted late continued to grow long after the other; initially, more massive satellites were disrupted.

For a subset of the satellite progenitors, the GC membership permutation influences the accretion redshift more strongly than it affects the stellar masses (see the discussion of Fig. 4). When expanding the sample of unambiguous Kraken GCs with NGC 6441, which most likely formed *in situ* in the Main progenitor (see Section 3.1), the accretion redshift decreases considerably, to $z_{\rm acc} \approx 1.8$. However, this is extremely unlikely to be accurate given the high metallicity ($[\mathrm{Fe/H}] = -0.6$) and old age ($\tau = 11.3 \pm 0.9$ Gyr) of NGC 6441. When omitting NGC 6441 from the *ex-situ* sample, the GC membership permutation has no significant influence on the accretion redshift of Kraken. Likewise, the accretion redshifts of *Gaia*-Enceladus and the progenitor of the Helmi streams are not significantly affected by the GC membership selection, as the shifts of the median fall well within the quoted uncertainties. As for its stellar mass, the accretion redshift of Sequoia changes for different GC membership permutations. Specifically, when attributing the ambiguous GC NGC 6535 to Sequoia, it systematically has a higher accretion redshift of $z_{\rm acc} \approx 1.7$. While the position of NGC 6535 in age–metallicity space ($[\mathrm{Fe/H}] = -1.73$ and $\tau = 12.2 \pm 0.6$ Gyr) does not allow distinguishing between Kraken and Sequoia, its orbital properties ($R_{\rm a} \approx 4.5$ kpc and $\epsilon \approx 0.63$) clearly place it in a part of the orbital parameter space that is not covered by any of the Sequoia GCs (which occupy $R_{\rm a} > 25$ kpc in the selection of Massari et al. 2019) and is rather consistent with the Kraken GCs (which have $R_{\rm a} < 7$ kpc).[7] When omitting NGC 6535 from the Sequoia membership permutations, the accretion redshift is always consistent with the value reported for the unambiguous GC membership assignment.

[7]We note that Myeong et al. (2019) associate additional GCs with Sequoia that also have apocentres $R_{\rm a} < 10$ kpc (NGC 5139 and NGC 6388). The first of these is included in our other permutations for Sequoia (see Table 1), implying that its inclusion or omission does not affect our results, whereas we omit NGC 6388 altogether, because its age ($\tau = 12.0 \pm 1.0$ Gyr) and metallicity ($[\mathrm{Fe/H}] = -0.77$) imply that it likely formed *in situ*.

Relative to other measurements in the literature, our predicted accretion redshifts largely satisfy the constraints obtained through independent methods. Helmi et al. (2018) estimate that *Gaia*-Enceladus was accreted ∼10 Gyr ago, whereas Belokurov et al. (2018) estimate a range of 8–11 Gyr ago, and Mackereth et al. (2019) provide an upper limit on the accretion redshift of $z_{\rm acc} < 1.5$ (or $t_{\rm acc} < 9.5$ Gyr). All three of these constraints are consistent with our predicted range of $t_{\rm acc} = 9.1^{+0.7}_{-0.7}$ Gyr. The same applies for Sequoia, which has been proposed to have been accreted 9–11 Gyr ago (Myeong et al. 2019), consistent with our estimate of $t_{\rm acc} = 9.4^{+0.4}_{-0.5}$ Gyr.

By contrast, Koppelman et al. (2019a) propose that the progenitor of the Helmi streams was accreted 5–8 Gyr ago, whereas our analysis predicts $t_{\rm acc} = 10.1^{+0.7}_{-0.9}$ Gyr. As discussed above, extending the GC membership leads to even later accretion redshifts. The estimate of Koppelman et al. (2019a) is based on collisionless *N*-body simulations of the Helmi streams, obtaining a best kinematic match for accretion times of 5–8 Gyr. However, they do find that the stellar age range of the Helmi streams is $\tau = 11$–13 Gyr. The agreement of the lower bound of the age range with our predicted accretion time suggests that star formation in the progenitor of the Helmi streams may have been truncated by its tidal disruption in the Galactic halo. If this interpretation is correct, then it remains an important open question how the dynamical constraints from Koppelman et al. (2019a) can be reconciled with this picture. There are several possible explanations. The upper limit on the accretion time reported by Koppelman et al. (2019a) of 8 Gyr reflects the maximum of the range of accretion times considered in their dynamical models, indicating that earlier accretion times may be dynamically possible, but have not been explored. The dynamical models also assume that the orbits do not evolve in time and neglect dynamical friction. Finally, Helmi (2020) notes that the orbit is close to a resonance, which would slow the spreading of the stars away from the stream, thereby biasing the analysis of Koppelman et al. (2019a) towards later accretion times. In view of these considerations, a plausible solution would be to extend the orbital parameter space surveyed by Koppelman et al. (2019a) to look for a kinematic match that also satisfies the prediction of our model, accounting for the effects of resonances if necessary.

For Sagittarius, the simulations of Law & Majewski (2010) and Niederste-Ostholt et al. (2012) suggest that it has been undergoing intense tidal disruption for the past 4–7 Gyr. Which moment in this interval corresponds to our time of accretion (as obtained from the E-MOSAICS simulations) depends quite sensitively on its definition. The large time interval over which the disruption of Sagittarius has been taking place greatly complicates this interpretation. At face value, the time over which Sagittarius has been strongly disrupted is consistent with our prediction of $t_{\rm acc} = 6.8^{+1.1}_{-1.1}$ Gyr. Using the association of GCs with stellar streams from disrupted dwarf satellites in E-MOSAICS, Hughes et al. (2019) find a relation between the age range of accreted GCs and the stellar mass of the satellite progenitor, where excursions from that relation are strongly correlated with the infall time, i.e. the time at which the progenitor crosses the virial radius of the halo. Applying these relations to Sagittarius, Hughes et al. (2019) predict that it entered the halo of the Milky Way $t_{\rm infall} = 9.3 \pm 1.8$ Gyr ago. This upper limit on the time of accretion is consistent with our prediction, as well as with the long time (at least 4–7 Gyr) spent by Sagittarius in the Galactic halo.

Finally, we point out that the quoted error bars on the predictions reflect random uncertainties. A comparison to the standard deviation shown in the right-hand panel of Fig. 1 (which is larger than the typical error bar) suggests the existence of an additional systematic uncertainty that may affect all accretion redshift measurements by up to 0.25 points in redshift.

3.2.3 Scaling relations between GC subpopulations and their host satellite progenitor masses

There exist well-documented relations between the total number ($N_{\rm GC}$) or mass ($M_{\rm GCS}$) of GCs and the stellar ($M_\star$) or halo virial ($M_{\rm h}$) mass of the host galaxy (e.g. Spitler & Forbes 2009; Durrell et al. 2014; Hudson, Harris & Harris 2014; Harris et al. 2017; Forbes et al. 2018b; Burkert & Forbes 2020). Because these relations are close to linear, they are often expressed in terms of number ratios, such as the specific frequency normalized by stellar mass ($T_{\rm N} \equiv N_{\rm GC}/M_\star$) or by halo mass ($\eta_{\rm N} \equiv N_{\rm GC}/M_{\rm h}$), or in terms of the ratio between the total GC system mass and the halo virial mass ($\eta_{\rm M} \equiv M_{\rm GCS}/M_{\rm h}$). It is an important question whether these relations are fundamental and were imprinted at the time of GC formation (Spitler & Forbes 2009; Boylan-Kolchin 2017; Harris et al. 2017; Burkert & Forbes 2020) or if they result from a combination of ongoing baryonic processes and further linearization by hierarchical galaxy assembly (Kruijssen 2015; Choksi et al. 2018; van Dokkum et al. 2018; El-Badry et al. 2019; Bastian et al. 2020). The key underlying question is how these metrics evolve with redshift. If they are imprinted at birth, they should exhibit little redshift evolution. However, if they are the outcome of gradual galaxy formation processes, then they should evolve with redshift. It is hard to avoid this interpretation – the fact that GCs are disrupted and galaxies grow with time means that the above metrics ($T_{\rm N}$, $\eta_{\rm N}$, and $\eta_{\rm M}$) should be higher at higher redshifts. Now that we have obtained the stellar masses, accretion redshifts, and number of GCs of the satellite progenitors, we can infer the above metrics at the times of satellite accretion and compare them to the relations observed across the $z = 0$ galaxy population. In addition to shedding light on the nature of the above scaling relations, placing the satellite progenitors in the context of these scaling relations also serves as an important consistency check.

To calculate the halo virial mass at the time of accretion for each of the satellite progenitors, we use the semi-empirical determination of the average relation between the stellar mass and halo mass as a function of redshift from Moster, Naab & White (2013), which is constrained by the observed evolution of the galaxy stellar mass function. We use their analytical expression to predict the stellar mass as a function of halo mass and redshift, and then numerically invert it by cubic interpolation to calculate the halo mass over the halo mass range $M_{\rm h} = 10^{10}$–10^{12} M$_\odot$ and the redshift range $z = 0$–5. In addition, we obtain the number of GCs by simply counting the members (see Table 1), and the total mass of the GC system by adding up the individual (dynamical) GC masses from Baumgardt & Hilker (2018). Strictly speaking, this provides a lower limit, because there is no guarantee that the GC membership list is complete. The restriction of the sample to GCs for which all eight feature variables are known results in an underestimation of the number of GCs by 0.0–0.3 dex, or 0.12 dex on average (compare our Table 1 to Massari et al. 2019). Likewise, we necessarily omit any effects of GC disruption in the Milky Way halo – if we could assign any of the fossil streams shown in Fig. 2 to individual satellite progenitors, this would also increase the number of GCs and their total mass. None the less, it allows us to calculate (lower limits on) $T_{\rm N}$, $\eta_{\rm N}$, and $\eta_{\rm M}$ at the time of accretion for each of the five satellite progenitors and their GC membership permutations. Uncertainties on the resulting numbers are obtained by calculating the 16th and 84th percentiles across all 10 000 Monte Carlo realizations of the neural network.

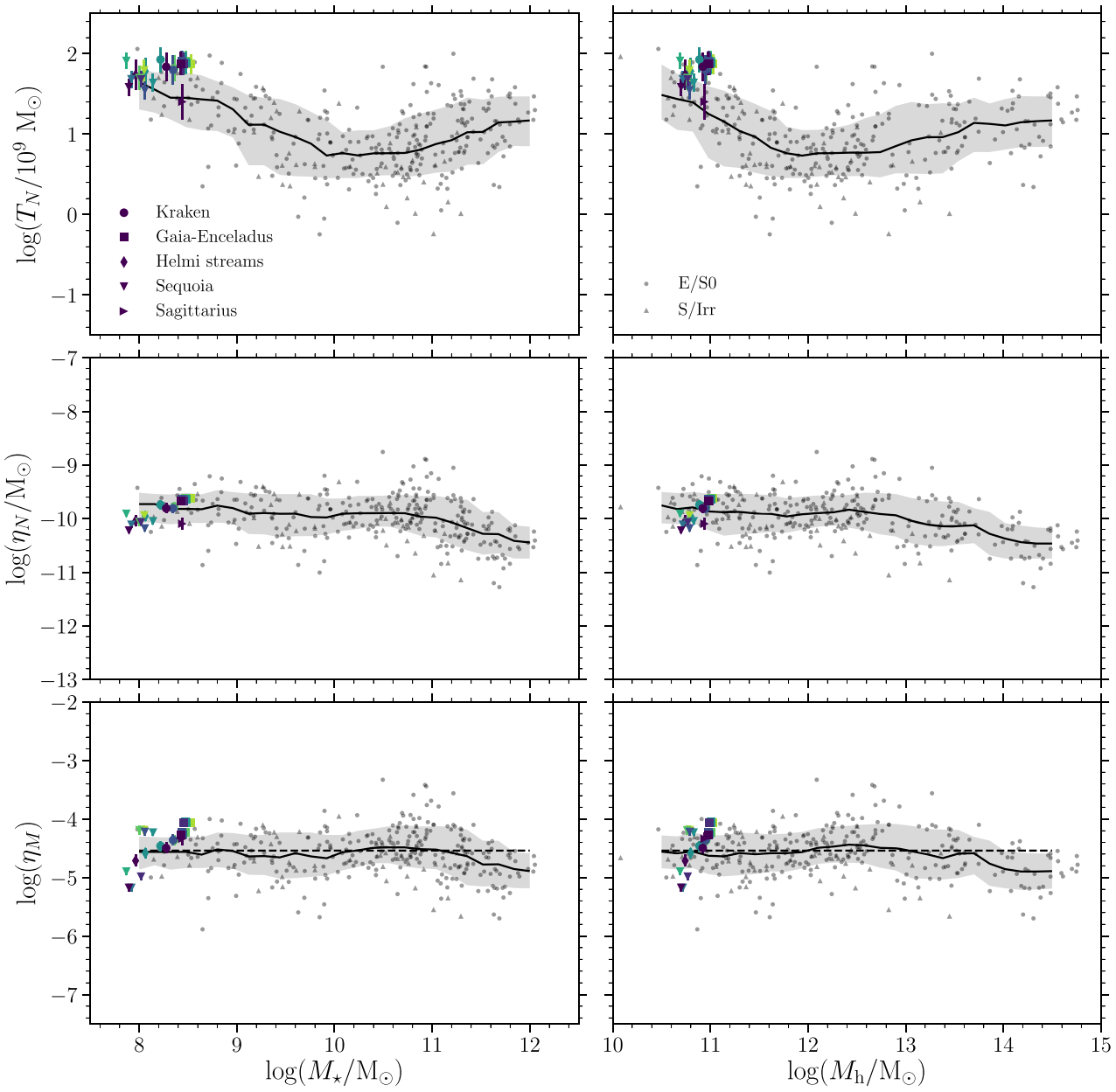

Figure 6. GC specific frequency (top row, $T_\mathrm{N} \equiv N_\mathrm{GC}/M_\star$), the number of GCs per unit halo virial mass (middle row, $\eta_\mathrm{N} \equiv N_\mathrm{GC}/M_\mathrm{h}$), and the ratio between the GC system mass and halo mass (bottom row, $\eta_\mathrm{M} \equiv M_\mathrm{GCS}/M_\mathrm{h}$), shown as a function of the galaxy stellar mass (left-hand column) and halo mass (right-hand column). The predictions for the satellite progenitors considered in this work are shown by the coloured symbols, with colours matching the GC membership permutations from Figs 4 and 5, and symbol shapes referring to the different progenitors as indicated by the legend. The observed galaxy population at $z = 0$ (Harris et al. 2017) is represented by the small grey symbols, showing early-type (E/S0) and late-type (S/Irr) galaxies. The solid line shows the running median across a 1 dex window, with 16th and 84th percentiles indicated by the grey-shaded band. In the bottom row, the horizontal dashed line indicates the roughly constant value of $\eta_\mathrm{M} = 2.9 \times 10^{-5}$ observed at $z = 0$ (Harris et al. 2017). The predictions for the satellite progenitors are lower limits by 0.0–0.3 dex (0.12 dex on average) due to sample incompleteness.

Fig. 6 shows how T_N, η_N, and η_M vary as a function of $M_\star$ and M_h, both for the satellite progenitors considered here and for the $z = 0$ galaxy sample from Harris et al. (2017), which includes 257 early-type galaxies (E/S0) and 46 late-type galaxies (S/Irr). In general, the satellites satisfy the scaling relations between the number or mass of GCs and the halo mass, exhibiting a similar scatter of η_N and η_M as the $z = 0$ galaxy population. However, for the specific frequency (T_N), they fall above the observed relation, suggesting that the relation between N_GC and $M_\star$ (or M_h) evolved since the accretion redshifts of the satellite progenitors. Interestingly, the satellite that accreted the most recently (Sagittarius) is the most consistent with the observations, whereas the satellite that accreted the earliest (Kraken) shows the strongest excess, together with *Gaia*-Enceladus, falling 0.5 dex above the observations at $z = 0$. This is no firm evidence, but at least suggests that the stellar mass of isolated galaxies with halo masses around $M_\mathrm{h} \sim 10^{11}\ \mathrm{M}_\odot$ grew by a factor of ∼3 in stellar

mass since $z_{\rm acc} \sim 1$. This is consistent with the abundance matching models of Moster et al. (2013), who predict a stellar mass growth by a factor of ∼5 over the same time interval, and of Behroozi, Wechsler & Conroy (2013), who predict a factor of ∼4. If true, this result implies that future observations of GC populations at $z > 1$ should find higher specific frequencies.[8]

Across all six relations shown in Fig. 6, there is no strong dependence on the GC membership permutation. As before, Sequoia shows the most pronounced variation, but this falls within the scatter of the relations observed at $z = 0$. The robustness of these results has two main implications. First, the inferred stellar masses and accretion redshifts of the satellite progenitors obtained here produce scaling relations that are in satisfactory agreement with observations at $z = 0$, lending some further credence to the analysis presented in this work. Secondly, we find some evidence that the specific frequency $T_{\rm N}$ changes with redshift, consistent with models suggesting that GC formation is fundamentally a baryonic process that results in relations with host galaxy properties that evolve in the context of hierarchical galaxy formation and evolution (e.g. Kruijssen 2015; Choksi et al. 2018; Pfeffer et al. 2018; El-Badry et al. 2019; Kruijssen et al. 2019a; Bastian et al. 2020).

4 MERGER HISTORY OF THE MILKY WAY

4.1 Merger mass ratios

Having inferred the accretion redshifts and the stellar masses at the time of accretion of Kraken, *Gaia*-Enceladus, the progenitor of the Helmi streams, Sequoia, and Sagittarius, we place these results in the context of the formation and assembly history of the Milky Way. Because these accretion events took place at different points in the Milky Way's history, it is somewhat non-trivial to assess how major (or minor) these mergers were. To calculate the merger mass ratios, we combine the stellar masses and accretion redshifts of the five satellite progenitors with the stellar mass growth history of the Milky Way that we inferred in Kruijssen et al. (2019b, fig. 4). We use a Monte Carlo approach to account for the uncertainties on this mass growth history. The mass growth history results from comparing the age–metallicity distribution of *in-situ* GCs, formed in the Main progenitor of the Milky Way, to the age–metallicity–mass distribution of central galaxies in the EAGLE simulations. Both the mass growth history and its corresponding chemical enrichment history are consistent with independent constraints inferred from the chemical abundances of thick-disc field stars and the resulting star formation history (see Kruijssen et al. 2019b, for further discussion; Snaith et al. 2014, 2015). The merger mass ratio then follows as $M_\star/M_{\star,\rm MW}(z_{\rm acc})$, where $M_{\star,\rm MW}(z_{\rm acc})$ is the stellar mass of the Milky Way at the time of accretion. By following this procedure, we obtain a predicted mass ratio for each of the 10 000 Monte Carlo realizations of the neural network and for each GC membership permutation.

Fig. 7 shows the PDFs of the merger stellar mass ratio, i.e. the ratio between the stellar mass of the satellite progenitors at the time of accretion (see Section 2.1 for its definition) and the stellar mass of the Milky Way at that time. Because the stellar mass ratio is a derived quantity that combines several variables with their own uncertainties (satellite stellar mass, accretion redshift, and Milky Way stellar mass), the uncertainties on the mass ratios are considerable. None the less, Fig. 7 reveals a wide range of mass ratios, even if all accretion events represent minor mergers, i.e. with mass ratios of $r_{M_\star} < 1{:}4$. Kraken is very likely to have been the most major merger that the Milky Way ever experienced,[9] with a mass ratio of $r_{M_\star} = 1{:}31^{+34}_{-16}$. This mass ratio is well in excess of that of *Gaia*-Enceladus ($r_{M_\star} = 1{:}67^{+41}_{-27}$), because Kraken accreted much earlier, when the stellar mass of the Milky Way was a factor of ∼3 lower than that when *Gaia*-Enceladus accreted. Inspection of the PDF across all 10 000 Monte Carlo realizations shows that Kraken has a probability of 0.3 per cent to have been a major merger, with a mass ratio of $r_{M_\star} > 1{:}4$. For all other satellite progenitors, this probability is <0.01 per cent – we can only provide an upper limit, because among the 10 000 Monte Carlo realizations there are no cases of a mass ratio $r_{M_\star} > 1{:}4$. The other accretion events, i.e. of the progenitor of the Helmi streams, Sequoia, and Sagittarius, are truly minor, with mass ratios of $r_{M_\star} = 1{:}110^{+99}_{-54}$, $1{:}191^{+100}_{-67}$, and $1{:}104^{+70}_{-43}$, respectively. This implies that the Milky Way must have grown in mass mostly by gas accretion and *in-situ* star formation (also see Trujillo-Gomez et al. 2020).

For Kraken and *Gaia*-Enceladus, the choice of GC membership permutation affects the merger mass ratio by 0.1 dex or less,[10] highlighting that these mass ratios are robust. However, the progenitor of the Helmi streams and (especially) Sequoia have merger mass ratios that are quite sensitive to membership selection. If NGC 5634 and NGC 5904 (representing the 'H99/G-E' group in Table 1) are included, the merger mass ratio of the progenitor of the Helmi streams becomes $r_{M_\star} = 1{:}70^{+59}_{-33}$, due to the combination of a higher accretion redshift and a higher stellar mass. Likewise, adding any of the ambiguous GCs to Sequoia increases its merger mass ratio, which for most membership permutations also results from a combination of a higher accretion redshift and a higher stellar mass. In all cases including at least one group of ambiguous GCs, Sequoia's mass ratio increases to around $r_{M_\star} \approx 1{:}130$. While these represent genuine systematic uncertainties, the qualitative conclusion remains unchanged – both the progenitor of the Helmi streams and Sequoia represent truly minor mergers.

There are no estimates of the merger mass ratio available in the literature for any of the satellite progenitors considered in this work, except for *Gaia*-Enceladus. Helmi et al. (2018) estimate that the mass ratio of the *Gaia*-Enceladus accretion event was $r_{M_\star} \approx 1{:}17$, which is about a factor of 4 higher than that predicted by our analysis. Two ingredients contribute roughly equally to this difference. First,

[8]This prediction is not affected directly by the details of the GC formation and disruption model in E-MOSAICS, because we use the observed numbers of GCs rather than the numbers of GCs produced in the E-MOSAICS simulations. Of course, the inferred satellite progenitor masses (i.e. the denominators of $T_{\rm N}$, $\eta_{\rm N}$, and $\eta_{\rm M}$) do rely on the GC demographics from E-MOSAICS, but the medians and IQRs of their ages, metallicities, apocentre radii, and eccentricities are not as strongly affected by the details of GC formation and disruption as their absolute numbers are.

[9]Of course, the Milky Way may have experienced very early major mergers with galaxies hosting few or no GCs. In Section 4.2, we suggest that any such mergers must have had masses considerably smaller than $M_\star \sim 10^8$ M$_\odot$, suggesting that the Milky Way would have had a mass of $M_\star \lesssim 10^9$ M$_\odot$. For the stellar mass growth history of the Milky Way that we inferred in Kruijssen et al. (2019b), this would imply merger redshifts of $z \gtrsim 4$. These redshifts were characterized by such high merger rates (with major merger rates of $\sim$1 Gyr^{-1}; see e.g. Fakhouri, Ma & Boylan-Kolchin 2010) that it would be challenging to identify discrete events. We therefore maintain the statement made here and consider the accretion history of the Milky Way prior to $z \sim 4$ to have been 'a mess'.

[10]This spread is smaller than the corresponding spread of satellite stellar mass in Fig. 4, because the stellar mass and accretion redshift are covariant. For lower accretion redshifts (and a more massive Milky Way at the time of accretion), the model also predicts a higher satellite mass, which suppresses the variation of the merger mass ratio.

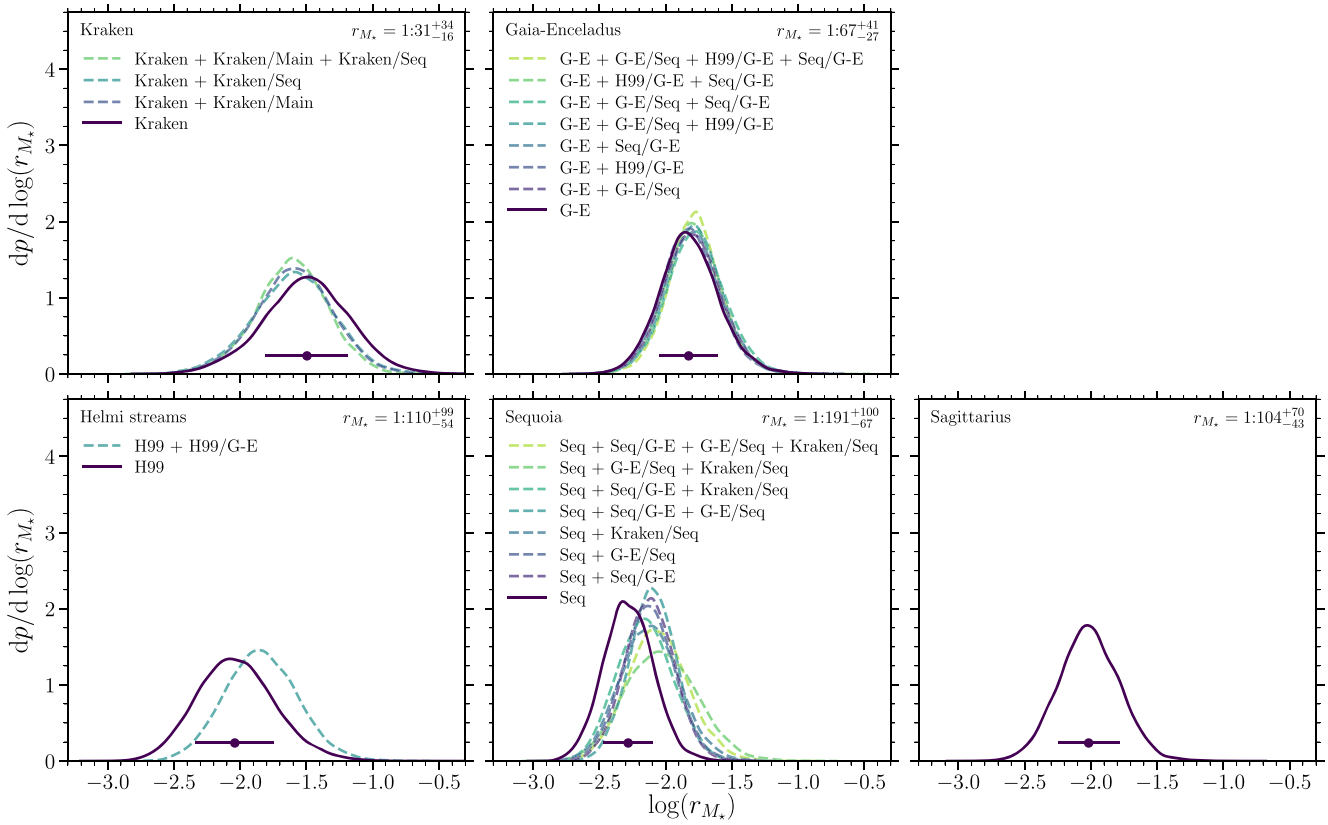

Figure 7. PDFs of the predicted merger stellar mass ratios of the satellite progenitor accretion events on to the Milky Way, inferred by applying a neural network trained on the E-MOSAICS simulations to their GC populations. Each PDF shows the distribution across all 10 000 Monte Carlo realizations of the neural network, applied to different GC membership permutations as indicated by the legend (also see Table 1). In each panel, the data point with error bar and the annotation in the top right indicate the median and 16–84th percentiles for the experiment using the unambiguous member GCs (corresponding to the solid line in that panel). This figure shows that the five accretion events are all minor mergers, spanning a wide range of mass ratios $r_{M_\star} = 1{:}(15–291)$. The accretion of Kraken likely represents the most major merger that the Milky Way ever experienced.

we predict a mass of *Gaia*-Enceladus itself that is about a factor of 2.2 lower than that quoted by Helmi et al. (2018). We discuss in Section 3.2.1 why we expect our estimate to be more accurate. Secondly, we predict that the mass of the Milky Way at the time of accretion was about a factor of 1.8 higher than the mass adopted by Helmi et al. (2018). Given the approximate nature of the numbers quoted by Helmi et al. (2018), we consider our merger mass ratio to be in rough agreement, but also expect it to be more accurate, because it is based on the self-consistent orbital and chemical evolution of satellite galaxies in hydrodynamical cosmological simulations.

To complement Fig. 7, Fig. 8 shows the PDFs of the merger halo mass ratio, i.e. the ratio between the halo mass of the satellite progenitors at the time of accretion (see Section 2.1 for its definition) and the halo mass of the Milky Way at that time (again obtained using the relation between stellar mass, halo mass, and redshift from Moster et al. 2013). Despite the fact that the halo mass ratio is a derived quantity that combines several variables with their own uncertainties, the relative uncertainties on the halo mass ratio are smaller than those on the stellar mass ratio, because the halo mass is a sublinear function of the stellar mass. As a result, Fig. 8 reveals a relatively narrow range of halo mass ratios, with $r_{M_\mathrm{h}} = 1{:}(5–24)$ and most halo mass ratios being around $r_{M_\mathrm{h}} \approx 1{:}10$. The halo mass ratio obtained for *Gaia*-Enceladus ($r_{M_\mathrm{h}} = 1{:}11^{+4}_{-3}$) is considerably smaller than that estimated by Helmi et al. (2018), who find $r_{M_\mathrm{h}} \approx 1{:}4$. The reason is the same as for the stellar mass ratios above – we obtain masses of both *Gaia*-Enceladus and the Milky Way at the time of accretion that are both somewhat lower. In part, the discrepancy arises because Helmi et al. (2018) assume that *Gaia*-Enceladus was similar in stellar mass to the Large Magellanic Cloud, which we find is an overestimation.

The halo masses and merger halo mass ratios obtained here may carry some systematic uncertainty, because the Moster et al. (2013) model predicts a halo mass of $M_\mathrm{h} = 2 \times 10^{12}$ M$_\odot$ for a galaxy with the stellar mass of the Milky Way. This is twice as large as recent estimates of the actual virial mass ($1.1 \pm 0.2 \times 10^{12}$ M$_\odot$; Cautun et al. 2020) and causes galaxies in the Moster et al. (2013) model with the stellar mass of the Milky Way to reach the currently observed virial mass already by $z = 1.4$. It is not clear if the empirical model of Moster et al. (2013) systematically overestimates all halo masses by a similar factor or if the Milky Way is simply an outlier. In the former case, our merger halo mass ratios are accurate, but the absolute halo masses of the satellite progenitors are overestimated by up to a factor of 2. In the latter case, our merger halo mass ratios are underestimated by up to a factor of 2, but the absolute halo masses of the satellite progenitors are accurate. Regardless, the estimated uncertainties of empirical halo abundance matching models are large enough to be consistent with the virial mass of the Milky Way (e.g. Behroozi et al. 2013, fig. 14).

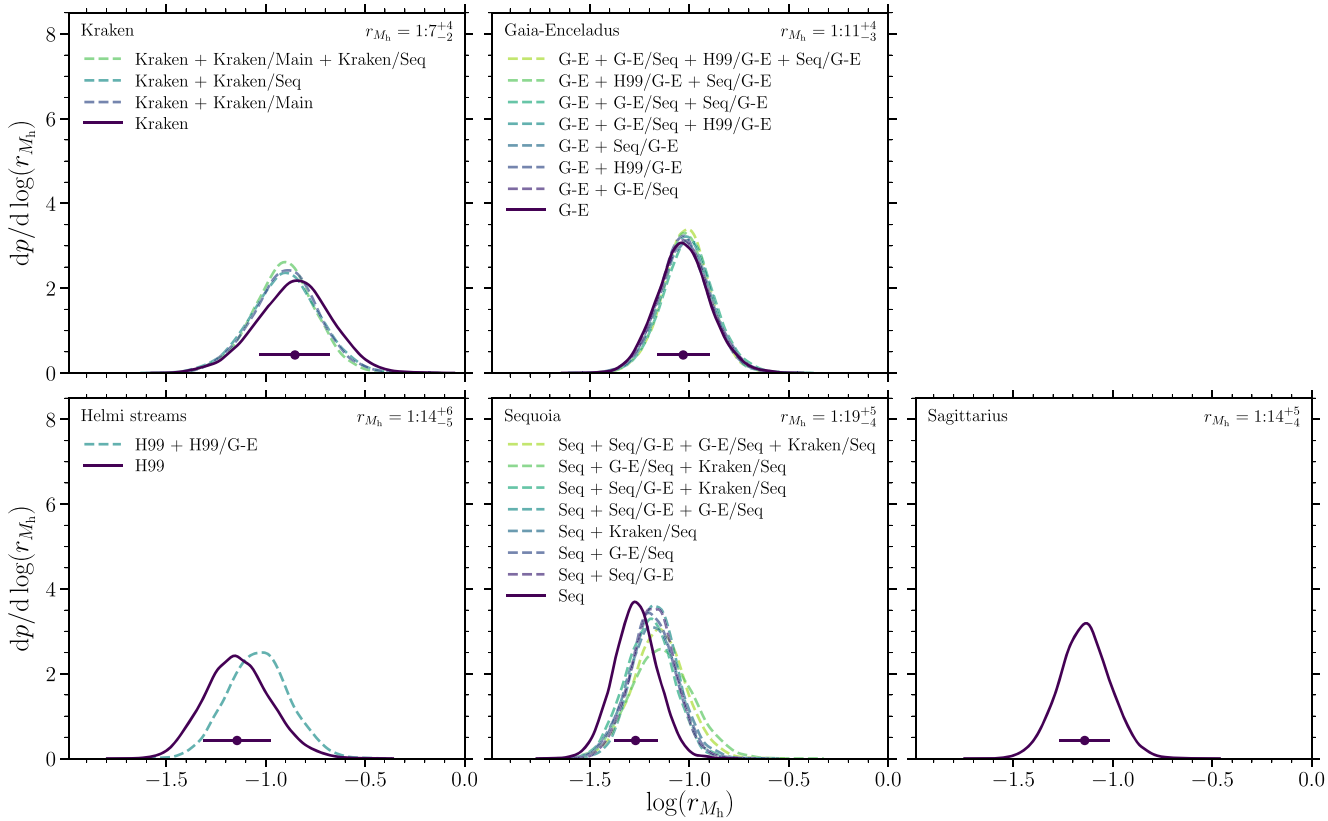

Figure 8. PDFs of the predicted merger halo mass ratios of the satellite progenitor accretion events on to the Milky Way, inferred by applying a neural network trained on the E-MOSAICS simulations to their GC populations and adopting the relation between stellar mass and halo mass at the accretion redshift from Moster et al. (2013) to obtain the halo masses. Each PDF shows the distribution across all 10 000 Monte Carlo realizations of the neural network, applied to different GC membership permutations as indicated by the legend (also see Table 1). In each panel, the data point with error bar and the annotation in the top right indicate the median and 16–84th percentiles for the experiment using the unambiguous member GCs (corresponding to the solid line in that panel). This figure shows that the five accretion events are all minor mergers, spanning a wide range of mass ratios $r_{M_\mathrm{h}} = 1{:}(5\text{–}24)$. The accretion of Kraken likely represents the most major merger that the Milky Way ever experienced.

4.2 Satellite accretion history and *in-situ* growth

Table 2 summarizes the properties of the five satellite progenitors that we have quantified in this work, including each possible GC membership permutation. We also include the numbers of GCs, but point out that these are necessarily lower limits, because there is no guarantee that any of the groups are complete. All quantities listed in the table are evaluated at the time of accretion, corresponding to the accretion redshift listed in each row.

In Kruijssen et al. (2019b), we estimated a total of 15 ± 3 accretion events with stellar masses $M_\star > 4.5 \times 10^6$ M$_\odot$. Even though we only discuss five such events in this paper, it is likely that these represent the most massive galaxies ever accreted by the Milky Way. Accordingly, the Milky Way never experienced a major merger since reaching a mass of $M_\star \sim 10^9$ M$_\odot$, which it had already attained at $z = 4$ (Snaith et al. 2014; Kruijssen et al. 2019b). This falls just within the Poisson noise of the expected number of major mergers. Milky Way-mass galaxies are expected to have experienced two major mergers since $z = 4$, or their most recent one at around $z = 2$ (Fakhouri et al. 2010). We cannot rule out the possibility that the Milky Way may have experienced a major merger before $z = 4$, but it is questionable how useful the concept of a major merger is when only up to a few per cent of the current Milky Way's stellar mass could have participated in the event.

Summing the stellar masses of the satellite progenitors at the time of their accretion yields a total accreted stellar mass of $\log(M_{\star,\,\mathrm{tot}}/\mathrm{M}_\odot) = 9.0 \pm 0.1$, which is similar in mass to the stellar halo [$(1.4 \pm 0.4) \times 10^9$ M$_\odot$; see Deason, Belokurov & Sanders 2019] and matches the total accreted stellar halo mass ($0.9^{+0.2}_{-0.1} \times 10^9$ M$_\odot$) estimated from APOGEE DR14 by Mackereth & Bovy (2020). The remaining 10 satellites are expected to have masses below $M_\star \sim 10^8$ M$_\odot$ (or otherwise they would have contributed $\gtrsim$5 GCs, which would plausibly have been identified), implying that the total accreted stellar mass is likely to be close to the sum of the five satellites discussed here [which is on the low side of, but consistent with the predictions for Milky Way-mass galaxies by Behroozi et al. (2019) and Moster, Naab & White (2019), who find (1–5) $\times 10^9$ M$_\odot$]. Therefore, the Galactic stellar halo can comfortably accommodate the total accreted satellite population, especially because some of the accreted mass is expected to contribute to the bulge and thick disc. In a broader context, this result means that the vast majority of the Milky Way's stellar mass formed *in situ*, with only a few per cent having been accreted in the form of stars from satellite galaxies. By contrast, the five satellite progenitors characterized here alone already contributed a total halo mass of $\log(M_{\mathrm{h,\,tot}}/\mathrm{M}_\odot) = 11.57 \pm 0.04$. This represents 25–45 per cent of the Milky Way's halo virial mass [$(1.1 \pm 0.2) \times 10^{12}$ M$_\odot$; see Cautun et al. 2020] and

Table 2. Summary of the inferred properties of the satellite progenitors, for all GC membership permutations (see Table 1 for the GC memberships). For each progenitor, the first row lists the results obtained using the GCs that are unambiguous members. From left to right, the columns list the progenitor (and GC membership permutation), the number of associated GCs, the accretion redshift, log stellar mass, log halo mass, specific frequency, log number of GCs per unit halo mass, log total GC system mass per unit halo mass, the stellar mass ratio of the merger with the Milky Way, and the halo mass ratio of the merger with the Milky Way (determined indirectly using the relation between stellar mass and halo mass from Moster et al. 2013; see the text). We remind the reader that the quoted error bars represent random uncertainties. All values carry an additional systematic uncertainty of 0.30 dex in $\log M_\star$ and 0.25 points in $z_{\rm acc}$.

Progenitor	$N_{\rm GC}$	$z_{\rm acc}$	$\log M_\star$ ($\rm M_\odot$)	$\log M_{\rm h}$ ($\rm M_\odot$)	$T_{\rm N}$ ($10^{-9}\,\rm M_\odot^{-1}$)	$\log \eta_{\rm N}$ ($\rm M_\odot^{-1}$)	$\log \eta_{\rm M}$	$r_{M_\star}$	$r_{M_{\rm h}}$
Kraken	13	$2.26^{+0.39}_{-0.45}$	$8.28^{+0.18}_{-0.17}$	$10.92^{+0.10}_{-0.10}$	69^{+33}_{-23}	$-9.81^{+0.10}_{-0.10}$	$-4.50^{+0.10}_{-0.10}$	$1{:}31^{+34}_{-16}$	$1{:}7^{+4}_{-2}$
Kraken + Kraken/Main	14	$1.88^{+0.37}_{-0.37}$	$8.35^{+0.18}_{-0.19}$	$10.95^{+0.10}_{-0.10}$	63^{+34}_{-21}	$-9.81^{+0.10}_{-0.10}$	$-4.36^{+0.10}_{-0.10}$	$1{:}39^{+36}_{-18}$	$1{:}8^{+4}_{-2}$
Kraken + Kraken/Seq	14	$2.17^{+0.34}_{-0.40}$	$8.22^{+0.16}_{-0.16}$	$10.89^{+0.09}_{-0.09}$	85^{+37}_{-26}	$-9.74^{+0.09}_{-0.09}$	$-4.46^{+0.09}_{-0.09}$	$1{:}39^{+40}_{-19}$	$1{:}8^{+4}_{-3}$
Kraken + Kraken/Main + Kraken/Seq	15	$1.82^{+0.32}_{-0.33}$	$8.36^{+0.16}_{-0.17}$	$10.96^{+0.09}_{-0.09}$	66^{+31}_{-21}	$-9.78^{+0.09}_{-0.09}$	$-4.36^{+0.09}_{-0.09}$	$1{:}40^{+34}_{-18}$	$1{:}8^{+4}_{-2}$
G-E	20	$1.35^{+0.26}_{-0.23}$	$8.43^{+0.15}_{-0.16}$	$10.98^{+0.07}_{-0.08}$	75^{+32}_{-21}	$-9.68^{+0.08}_{-0.07}$	$-4.27^{+0.08}_{-0.07}$	$1{:}67^{+41}_{-27}$	$1{:}11^{+4}_{-3}$
G-E + G-E/Seq	21	$1.34^{+0.25}_{-0.22}$	$8.45^{+0.14}_{-0.15}$	$10.99^{+0.07}_{-0.08}$	74^{+31}_{-21}	$-9.67^{+0.08}_{-0.07}$	$-4.06^{+0.08}_{-0.07}$	$1{:}63^{+39}_{-25}$	$1{:}11^{+4}_{-3}$
G-E + H99/G-E	22	$1.35^{+0.24}_{-0.21}$	$8.44^{+0.13}_{-0.14}$	$10.98^{+0.07}_{-0.07}$	81^{+32}_{-21}	$-9.64^{+0.07}_{-0.07}$	$-4.23^{+0.07}_{-0.07}$	$1{:}66^{+39}_{-26}$	$1{:}11^{+3}_{-3}$
G-E + Seq/G-E	22	$1.36^{+0.25}_{-0.22}$	$8.44^{+0.14}_{-0.15}$	$10.99^{+0.07}_{-0.08}$	79^{+32}_{-22}	$-9.64^{+0.08}_{-0.07}$	$-4.26^{+0.08}_{-0.07}$	$1{:}63^{+40}_{-25}$	$1{:}10^{+4}_{-3}$
G-E + G-E/Seq + H99/G-E	23	$1.34^{+0.23}_{-0.20}$	$8.48^{+0.13}_{-0.14}$	$11.00^{+0.07}_{-0.07}$	77^{+29}_{-19}	$-9.64^{+0.07}_{-0.07}$	$-4.05^{+0.07}_{-0.07}$	$1{:}61^{+37}_{-23}$	$1{:}10^{+3}_{-3}$
G-E + G-E/Seq + Seq/G-E	23	$1.34^{+0.25}_{-0.22}$	$8.48^{+0.14}_{-0.15}$	$11.01^{+0.07}_{-0.08}$	76^{+31}_{-21}	$-9.64^{+0.08}_{-0.07}$	$-4.07^{+0.08}_{-0.07}$	$1{:}59^{+37}_{-23}$	$1{:}10^{+3}_{-3}$
G-E + H99/G-E + Seq/G-E	24	$1.31^{+0.23}_{-0.20}$	$8.48^{+0.13}_{-0.14}$	$11.00^{+0.07}_{-0.07}$	80^{+31}_{-21}	$-9.62^{+0.07}_{-0.07}$	$-4.23^{+0.07}_{-0.07}$	$1{:}63^{+37}_{-24}$	$1{:}10^{+3}_{-3}$
G-E + G-E/Seq + H99/G-E + Seq/G-E	25	$1.26^{+0.21}_{-0.19}$	$8.53^{+0.12}_{-0.13}$	$11.03^{+0.06}_{-0.07}$	74^{+27}_{-18}	$-9.63^{+0.07}_{-0.06}$	$-4.06^{+0.07}_{-0.06}$	$1{:}60^{+33}_{-22}$	$1{:}10^{+3}_{-2}$
H99	5	$1.75^{+0.42}_{-0.37}$	$7.96^{+0.19}_{-0.18}$	$10.74^{+0.10}_{-0.10}$	54^{+29}_{-19}	$-10.04^{+0.10}_{-0.10}$	$-4.71^{+0.10}_{-0.10}$	$1{:}110^{+99}_{-54}$	$1{:}14^{+6}_{-5}$
H99 + H99/G-E	7	$1.93^{+0.35}_{-0.35}$	$8.06^{+0.16}_{-0.16}$	$10.80^{+0.09}_{-0.09}$	61^{+27}_{-19}	$-9.95^{+0.09}_{-0.09}$	$-4.58^{+0.09}_{-0.09}$	$1{:}70^{+59}_{-33}$	$1{:}11^{+5}_{-3}$
Seq	3	$1.46^{+0.17}_{-0.17}$	$7.90^{+0.11}_{-0.11}$	$10.70^{+0.06}_{-0.06}$	38^{+11}_{-8}	$-10.22^{+0.06}_{-0.06}$	$-5.18^{+0.06}_{-0.06}$	$1{:}191^{+100}_{-67}$	$1{:}19^{+5}_{-4}$
Seq + Seq/G-E	5	$1.52^{+0.17}_{-0.17}$	$8.02^{+0.10}_{-0.10}$	$10.77^{+0.05}_{-0.05}$	48^{+12}_{-10}	$-10.07^{+0.05}_{-0.05}$	$-4.99^{+0.05}_{-0.05}$	$1{:}132^{+69}_{-45}$	$1{:}15^{+4}_{-3}$
Seq + G-E/Seq	4	$1.46^{+0.20}_{-0.19}$	$8.06^{+0.11}_{-0.11}$	$10.79^{+0.06}_{-0.06}$	35^{+10}_{-8}	$-10.18^{+0.06}_{-0.06}$	$-4.22^{+0.06}_{-0.06}$	$1{:}134^{+76}_{-48}$	$1{:}15^{+5}_{-4}$
Seq + Kraken/Seq	4	$1.69^{+0.23}_{-0.24}$	$7.92^{+0.11}_{-0.11}$	$10.72^{+0.06}_{-0.06}$	48^{+14}_{-11}	$-10.12^{+0.06}_{-0.06}$	$-5.18^{+0.06}_{-0.06}$	$1{:}131^{+87}_{-52}$	$1{:}15^{+5}_{-4}$
Seq + Seq/G-E + G-E/Seq	6	$1.37^{+0.17}_{-0.15}$	$8.14^{+0.11}_{-0.11}$	$10.83^{+0.06}_{-0.06}$	43^{+12}_{-10}	$-10.05^{+0.06}_{-0.06}$	$-4.23^{+0.06}_{-0.06}$	$1{:}126^{+61}_{-41}$	$1{:}15^{+4}_{-3}$
Seq + Seq/G-E + Kraken/Seq	6	$1.71^{+0.20}_{-0.21}$	$7.87^{+0.10}_{-0.10}$	$10.69^{+0.05}_{-0.05}$	81^{+20}_{-17}	$-9.91^{+0.05}_{-0.05}$	$-4.90^{+0.05}_{-0.05}$	$1{:}143^{+88}_{-56}$	$1{:}16^{+5}_{-4}$
Seq + G-E/Seq + Kraken/Seq	5	$1.69^{+0.33}_{-0.32}$	$8.00^{+0.15}_{-0.15}$	$10.76^{+0.08}_{-0.08}$	50^{+20}_{-14}	$-10.06^{+0.08}_{-0.08}$	$-4.19^{+0.08}_{-0.08}$	$1{:}109^{+90}_{-51}$	$1{:}14^{+6}_{-4}$
Seq + Seq/G-E + G-E/Seq + Kraken/Seq	7	$1.56^{+0.25}_{-0.23}$	$8.05^{+0.13}_{-0.13}$	$10.79^{+0.07}_{-0.07}$	62^{+22}_{-16}	$-9.94^{+0.07}_{-0.07}$	$-4.19^{+0.07}_{-0.07}$	$1{:}115^{+77}_{-48}$	$1{:}14^{+5}_{-4}$
Sagittarius	7	$0.76^{+0.22}_{-0.19}$	$8.44^{+0.22}_{-0.21}$	$10.94^{+0.11}_{-0.10}$	25^{+16}_{-10}	$-10.10^{+0.10}_{-0.11}$	$-4.34^{+0.10}_{-0.11}$	$1{:}104^{+70}_{-43}$	$1{:}14^{+5}_{-4}$

shows that the satellite progenitors are important contributors to the Milky Way's virial mass.

A similar balance can be made for the Galactic GC population. We find 55 GCs that were likely accreted as part of the five satellite progenitors. While this number is inevitably a lower limit on the true number of *ex-situ* GCs, we do not expect the sample to be hugely incomplete. If the scaling relations between the number of GCs and the host galaxy mass discussed in Section 3.2.3 apply also at the accretion redshifts of the satellite progenitors (which can be disputed), we do not expect the current sample to be incomplete by more than 40 per cent, such that the total might be as high as 90 accreted GCs. A comparison to the GC membership assignments in Massari et al. (2019) confirms this interpretation. Given a total number of 157 GCs in the catalogue of Harris (1996, 2010 edition), this implies that 35–60 per cent of the Galactic GCs formed *ex situ*, with the complementing 40–65 per cent having formed *in situ*. This is consistent with the estimate that 40 per cent of GCs formed *ex situ*, which we arrived at using the GC age–metallicity distribution only in Kruijssen et al. (2019b) and also using the kinematics of the GC population in Trujillo-Gomez et al. (2020).

Many of the results of this paper are summarized visually in Fig. 9, which provides an extensive update to the merger tree presented in Kruijssen et al. (2019b, fig. 6). Thanks to the discovery and characterization of several new satellites with *Gaia* DR2 (e.g. Koppelman et al. 2019a; Massari et al. 2019; Myeong et al. 2019) and the interpretative framework provided by the E-MOSAICS simulations (Pfeffer et al. 2018; Kruijssen et al. 2019a), we now have a reconstruction of the Milky Way's merger tree that specifically includes five out of the 15 ± 3 expected satellites with stellar masses of $M_\star > 4.5 \times 10^6$ $\rm M_\odot$. This merger tree includes the stellar mass growth history of the Milky Way's Main progenitor, the stellar masses and accretion redshifts of five satellite galaxies with stellar masses of $M_\star \gtrsim 5 \times 10^7$ $\rm M_\odot$, and the lower limits on the numbers of GCs contributed by each of these accretion events.

5 CONCLUSIONS

In this paper, we have used the E-MOSAICS simulations to quantify the stellar masses and accretion redshifts of five satellite galaxies that have been accreted by the Milky Way, as well as several other of their properties. These satellites are Kraken (previously referred to as the progenitor of the 'low-energy GCs' by Massari et al. 2019), *Gaia*-Enceladus, the progenitor of the Helmi streams, Sequoia, and Sagittarius. They likely represent the most massive objects that the Milky Way accreted since $z = 4$. We predict their properties by training an artificial neural network on the E-MOSAICS simulations, which follow the co-formation and co-evolution of galaxies and their GC populations. This network relates the ages, metallicities, and orbital properties of a group of GCs that formed in a common satellite

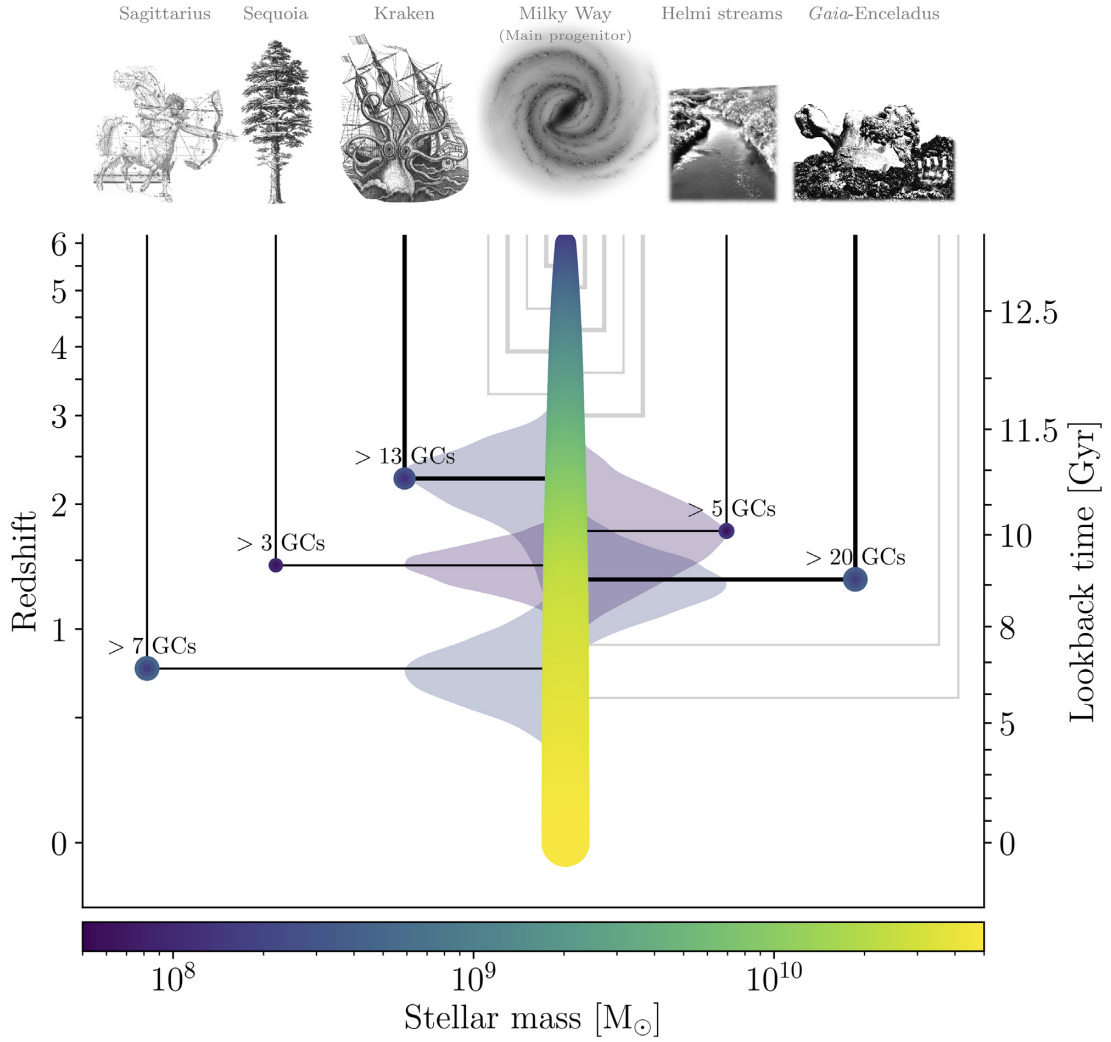

Figure 9. Galaxy merger tree of the Milky Way, inferred by applying the insights gained from the E-MOSAICS simulations to the Galactic GC population. This figure summarizes many of the results presented in this paper. The main progenitor is denoted by the trunk of the tree, coloured by stellar mass (based on the reconstruction from Kruijssen et al. 2019b; see the colour bar). Black lines indicate the five identified (and likely most massive) satellites, with the shaded areas visualizing the PDFs of the accretion redshifts from Fig. 5. The coloured circles indicate the stellar masses at the time of accretion (both with their colours and sizes) and the subtly different colours of the outer rings and central dots visualize the uncertainties on the stellar masses. The annotations list the minimum number of GCs brought in by each satellite. Light grey lines illustrate the global statistics of the Milky Way's merger tree inferred by Kruijssen et al. (2019b), but have no absolute meaning. Thin lines mark tiny mergers (with a mass ratio of $r_{M_\star} < 1{:}100$) and thick lines denote minor (or possibly major) mergers (with a mass ratio of $r_{M_\star} > 1{:}100$). This merger tree is consistent with the stellar mass growth history of the Milky Way, the total number of mergers ($N_{\rm br}$), the number of high-redshift mergers ($N_{{\rm br},z>2}$), and the numbers of tiny and minor mergers ($N_{<1:100}$ and $N_{1:100–1:4}$) from Kruijssen et al. (2019b), as well as with the five identified satellite progenitors discussed in this work, including their stellar masses, accretion redshifts, and GC populations. Note that only progenitors with masses of $M_\star > 4.5 \times 10^6$ M$_\odot$ are included. From left to right, the six images along the top of the figure indicate the identified progenitors, i.e. Sagittarius, Sequoia, Kraken, the Milky Way's Main progenitor, the progenitor of the Helmi streams, and *Gaia*-Enceladus.

progenitor to the properties of that progenitor. The conclusions of this work are as follows:

(i) The neural network is capable of predicting the stellar masses and accretion redshifts of the satellite progenitors to high precision, with a validation score of $0.89^{+0.06}_{-0.06}$ and a scatter of $\sigma[\log_{10}(M_{\star,\rm pred}/M_{\star,\rm true})] = 0.41^{+0.05}_{-0.04}$ and $\sigma[\log_{10}(1+z_{\rm acc,pred}/1+z_{\rm acc,true})] = 0.13^{+0.01}_{-0.01}$ around the true values, respectively. Omitting the GC age–metallicity information leads to considerably worse constraints on both quantities. Additionally, the accretion redshifts strongly rely on the GC orbital information (Fig. 1).

(ii) Modulo a small number of changes, we adopt the proposed GC membership of the satellite progenitors from Massari et al. (2019, see our Table 1). The five groups of GCs associated with the different satellite progenitors show clear differences in apocentre–eccentricity space and age–metallicity space. By comparing the orbital properties of GCs to those of fossil streams from Bonaca & Kruijssen (2020), we find that the streams have similar orbital properties as the GCs associated with Kraken, *Gaia*-Enceladus, the progenitor of the Helmi streams, and Sagittarius. The Fimbulthul stream falls in between the orbital properties spanned by the GCs associated with Kraken and *Gaia*-Enceladus, suggesting that it may be the relic of a GC that formed in either of these two satellites. In age–metallicity space, the Kraken GCs have the highest metallicities at a given age, closely followed by the *Gaia*-Enceladus GCs, suggesting that these two galaxies were the most massive at any

moment in time, with Kraken having a slightly higher mass (Figs 2 and 3).

(iii) By applying the neural network to the Galactic GCs associated with the five satellite progenitors, we find that the satellites span a relatively narrow stellar mass range at the time of accretion, of $M_\star = (0.6–4.6) \times 10^8$ M$_\odot$. The top end of this range is occupied by Kraken, *Gaia*-Enceladus, and Sagittarius, with the latter two being the most massive. This differs from the stellar masses at any given moment in time, because the accretion redshifts of these satellites differ. By accreting later, Sagittarius was able to attain a higher mass, despite initially being considerably less massive than Kraken and *Gaia*-Enceladus. The predicted stellar masses are consistent with previous literature estimates given their uncertainties,[11] but considerably more precise (Fig. 4).

(iv) The accretion redshifts of the five satellites span a wide range, of $z_{\rm acc} = 0.57–2.65$ (corresponding to lookback times of $t_{\rm acc} = 5.7–11.3$ Gyr). Kraken was the first galaxy to be accreted, followed by the progenitor of the Helmi streams, Sequoia and *Gaia*-Enceladus, and finally Sagittarius. This order explains why the stellar mass of Sagittarius at the time of accretion exceeded that of Kraken, despite having a lower mass at any given age prior to the accretion of both satellites. The predicted accretion redshifts are consistent with the rough ranges available from other works in the literature. The only possible tension with previous results is the progenitor of the Helmi streams, for which dynamical models predict a much more recent accretion ($t_{\rm acc} = 5–8$ Gyr) than that obtained in this work ($t_{\rm acc} = 10.1^{+0.7}_{-0.9}$ Gyr), whereas the stellar age range of $\tau = 11–13$ Gyr is consistent with our prediction. Reconciling these predictions is an important point of attention for future work (Fig. 5).

(v) The presented results are generally not sensitive to the details of the GC membership. By including and excluding GCs with ambiguous memberships, we find that the stellar masses and accretion redshift of the satellite progenitors exhibit differences smaller than the uncertainties estimated from the neural network. Only when including GCs that are highly unlikely to have been a member of a satellite progenitor do the results change by more than the uncertainties (Figs 4 and 5 and Table 2).

(vi) Taking together the above results, the 'low-energy' group of GCs identified by Massari et al. (2019) has the properties predicted for Kraken by Kruijssen et al. (2019b), i.e. having a high mass and contributing a large number of GCs, similar to that of *Gaia*-Enceladus. This high mass is required by the position of the Kraken GCs in age–metallicity space, where they occupy the high-metallicity side of the 'satellite branch'. In addition, the small apocentre radii of the Kraken GCs ($R_{\rm a} < 7$ kpc) require either a high progenitor mass or a very high accretion redshift (Pfeffer et al. 2020). The fact that the ages of the Kraken GCs reach down to $z \sim 2$ rules out the latter option. The only alternative that remains is a satellite progenitor that at any given time had a mass similar to (or higher than) *Gaia*-Enceladus, as originally predicted. Kraken may represent the tip of the iceberg of several further accretion events hidden towards the Galactic Centre (Helmi 2020), where the identification of substructure is complicated by high stellar densities and phase–space mixing on short orbital time-scales. Alternatively, in view of its high mass, it is possible that Kraken represents the main accretion event contributing to the formation of the bulge. This interpretation seems to be favoured by the recent chemical abundance analysis of Horta et al. (2020a), who find that Kraken contributed a significant fraction of the bulge stars with an *ex-situ* origin.

(vii) We calculate the number and total mass of GCs per unit halo mass in each satellite progenitor at the time of its accretion, finding that these are consistent with the typical numbers found for the GC populations of galaxies at $z = 0$. However, the specific frequencies of the Kraken and *Gaia*-Enceladus exceed those of galaxies of a similar mass at $z = 0$, suggesting that the scaling relations between the number of GCs and the host galaxy mass evolve with redshift (e.g. Kruijssen 2015; Choksi & Gnedin 2019; Bastian et al. 2020). We propose that the excess can be explained by the continued stellar mass growth of galaxies after $z_{\rm acc}$. As a result, we predict that future observations of GC populations at $z > 1$ should find higher specific frequencies than at $z = 0$ (Fig. 6).

(viii) By combining the stellar masses and accretion redshifts of the satellite progenitors with the inferred stellar mass growth history of the Milky Way's Main progenitor, we infer the stellar mass ratios of the five accretion events. We find that all accretion events are minor mergers (i.e. with stellar mass ratios of $r_{M_\star} < 1{:}4$). Thanks to its high mass and accretion redshift, Kraken is the most major merger (i.e. the merger with the highest mass ratio) out of the five satellites considered here, with a mass ratio of $r_{M_\star} = 1{:}31^{+34}_{-16}$, making it very likely to be the most significant merger ever experienced by the Milky Way. Based on the full PDF obtained from the neural network, the accretion of Kraken by the Milky Way has a chance of 0.3 per cent to have been a major merger event (for which $r_{M_\star} > 1{:}4$). For all other satellite progenitors, the merger mass ratios are smaller than that of Kraken by factors of 2–6. This implies that the Milky Way must have grown its stellar mass mostly by gas accretion and *in-situ* star formation (Fig. 7).

(ix) Likewise, we find low merger halo mass ratios, in the range of $r_{M_{\rm h}} = 1{:}(5–24)$, again indicating that all accretion events were minor mergers. With a halo mass ratio of $r_{M_{\rm h}} = 1{:}7^{+4}_{-2}$, Kraken may have been the last merger to have significantly disrupted the Galactic disc (at $z_{\rm acc} = 2.26^{+0.39}_{-0.45}$ or $t_{\rm acc} = 10.9^{+0.4}_{-0.7}$ Gyr), as all subsequent mergers (including *Gaia*-Enceladus) had halo mass ratios of $r_{M_{\rm h}} < 1{:}10$ (Fig. 8).

(x) We tabulate all quantities inferred for the satellite progenitors in Table 2. The sum of their stellar masses is $\log(M_{\star,\,\rm tot}/{\rm M}_\odot) = 9.0 \pm 0.1$, similar to the mass of the Galactic stellar halo [$(1.4 \pm 0.4) \times 10^9$ M$_\odot$; see Deason et al. 2019] and in agreement with the total accreted stellar halo mass ($0.9^{+0.2}_{-0.1} \times 10^9$ M$_\odot$) estimated from APOGEE DR14 by Mackereth & Bovy (2020). We find that the stellar halo can accommodate the total accreted satellite population, and that only a few per cent of the Milky Way's stellar mass was accreted in the form of dwarf galaxies. The same applies for GCs, for which we estimate that 35–50 per cent has an *ex-situ* origin. In contrast with the stellar mass balance, the five satellite progenitors did contribute a significant total halo mass of $\log(M_{\rm h,\,tot}/{\rm M}_\odot) = 11.57 \pm 0.04$, which is 25–45 per cent of the Galactic virial mass [$(1.1 \pm 0.2) \times 10^{12}$ M$_\odot$; see Cautun et al. 2020].

(xi) We combine the results of our analysis with those from Kruijssen et al. (2019b) to present the most detailed reconstruction to date of the Milky Way's merger tree. This merger tree includes a total 15 ± 3 expected satellites with stellar masses of $M_\star > 4.5 \times 10^6$ M$_\odot$, as well as the stellar mass growth history of the Milky Way's Main progenitor, the stellar masses and accretion redshifts of the five satellite galaxies considered here, and the lower limits on the numbers of GCs contributed by these accretion events (Fig. 9).

The above results represent an example of how the phase–space clustering of stellar populations in the revolutionary *Gaia* data can

[11] Differences do exist relative to previous studies that estimate projected stellar masses at $z = 0$ rather than considering these at the time of accretion. See Section 3.2.1 for details.

unlock a wealth of information when interpreted using state-of-the-art numerical simulations of galaxy formation and evolution. The machine-learning approach employed here provides a very promising way of achieving this. By applying our formalism to the complete E-MOSAICS volume including all galaxies in a (34 comoving Mpc)^3 region, we aim to further boost its precision and diagnostic power through greatly improved statistics and a correspondingly larger training data set. This will be a critical step, because increasing the size of the training set allows the use of observables that are less information rich. Therefore, it is to be expected that the expansion of our formalism to a larger suite of simulated galaxies will enable reconstructing the assembly histories of galaxies beyond the Milky Way, for which less comprehensive diagnostics are available. While this paper provides an application to only a single galaxy, it is clear that the field has entered an era in which GCs are now firmly established, quantitative tracers of galaxy formation and assembly.

Focusing on the assembly history of the Milky Way, our results add to a growing body of evidence that the Milky Way experienced an unusual path to adolescence. Not only did it assemble unusually quickly for its mass, but it also experienced a striking paucity of major accretion events, with only a handful of minor mergers shaping the Galactic stellar halo. The fact that it grew most of its stellar mass through secular processes and *in-situ* star formation implies that it may not be the most representative example for understanding the evolution and assembly of the galaxy population, but is a correspondingly more pleasant environment to live in.

ACKNOWLEDGEMENTS

We thank the referee Davide Massari, as well as Vasily Belokurov, Charlie Conroy, Alis Deason, Benjamin Keller, and Rohan Naidu, for helpful discussions that improved the paper. We thank Bill Harris for providing the data from Harris et al. (2017) in an electronic form. JMDK gratefully acknowledges the hospitality of the Institute for Theory and Computation at Harvard University, where part of this work was carried out. JMDK and MC gratefully acknowledge funding from the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) through an Emmy Noether Research Group (grant number KR4801/1-1) and the DFG Sachbeihilfe (grant number KR4801/2-1). JMDK, STG, and MRC gratefully acknowledge funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme via the ERC Starting Grant MUSTANG (grant agreement number 714907). JMDK gratefully acknowledges funding from the DFG via the SFB 881 'The Milky Way System' (subproject B2). JP and NB gratefully acknowledge funding from the ERC under the European Union's Horizon 2020 research and innovation programme via the ERC Consolidator Grant Multi-Pop (grant agreement number 646928). MRC is supported by a Fellowship from the International Max Planck Research School for Astronomy and Cosmic Physics at the University of Heidelberg (IMPRS-HD). NB and RAC are Royal Society University Research Fellows. This work used the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). This equipment was funded by BIS National E-infrastructure capital grant ST/K00042X/1, STFC capital grants ST/H008519/1 and ST/K00087X/1, STFC DiRAC Operations grant ST/K003267/1, and Durham University. DiRAC is part of the National E-Infrastructure. This study also made use of high-performance computing facilities at Liverpool John Moores University, partly funded by the Royal Society and LJMU's Faculty of Engineering and Technology. This work has made use of MATPLOTLIB (Hunter 2007), NUMPY (van der Walt, Colbert & Varoquaux 2011), and ASTROPY (Astropy Collaboration 2013).

Credits of the images in Fig. 9 are as follows: The image of Sagittarius is adapted from Alexander Jamieson's *A Celestial Atlas*, 1822, plate 25. The image of Sequoia is adapted from a line engraving of a Sequoiadendron giganteum in California (1872–1873). The image of Kraken is adapted from an illustration in Pierre Denys de Montfort's *Histoire Naturelle, Générale et Particuliere des Mollusques*, Vol. 2, 1801, pp. 256–257. The image of the Milky Way is adapted from the illustration by R. Hurt, with credit NASA/JPL-Caltech/R. Hurt (SSC/Caltech). The image of the progenitor of the Helmi Streams is an adapted picture of the Arroyo Napostá river near Bahía Blanca, Argentina. The image of *Gaia*-Enceladus is an adapted picture of Gaspard Marsy's *Fontaine de l'Encelade* (1675–1676) at the Musée national des châteaux de Versailles et de Trianon in Versailles, France.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Astropy Collaboration, 2013, A&A, 558, A33
Balbinot E. et al., 2016, ApJ, 820, 58
Bastian N., Pfeffer J., Kruijssen J. M. D., Crain R. A., Trujillo-Gomez S., Reina-Campos M., 2020, MNRAS, preprint (arXiv:2005.05991)
Baumgardt H., Hilker M., 2018, MNRAS, 478, 1520
Baumgardt H., Hilker M., Sollima A., Bellini A., 2019, MNRAS, 482, 5138
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, MNRAS, 488, 3143
Bell E. F. et al., 2008, ApJ, 680, 295
Belokurov V. et al., 2006, ApJ, 642, L137
Belokurov V., Erkal D., Evans N. W., Koposov S. E., Deason A. J., 2018, MNRAS, 478, 611
Bonaca A., Kruijssen J. M. D., 2020, ApJ, submitted
Boylan-Kolchin M., 2017, MNRAS, 472, 3120
Buitinck L. et al., 2013, in ECML PKDD Workshop: Languages for Data Mining and Machine Learning. p. 108
Bullock J. S., Johnston K. V., 2005, ApJ, 635, 931
Burkert A., Forbes D. A., 2020, AJ, 159, 56
Cautun M. et al., 2020, MNRAS, 494, 4291
Choksi N., Gnedin O. Y., 2019, MNRAS, 488, 5409
Choksi N., Gnedin O. Y., Li H., 2018, MNRAS, 480, 2343
Chollet F. et al., 2015, Keras. available at: https://keras.io
Cooper A. P. et al., 2010, MNRAS, 406, 744
Crain R. A. et al., 2015, MNRAS, 450, 1937
de Boer T. J. L., Belokurov V., Koposov S. E., 2018, MNRAS, 473, 647
Deason A. J., Belokurov V., Evans N. W., Johnston K. V., 2013, ApJ, 763, 113
Deason A. J., Belokurov V., Weisz D. R., 2015, MNRAS, 448, L77
Deason A. J., Mao Y.-Y., Wechsler R. H., 2016, ApJ, 821, 5
Deason A. J., Belokurov V., Koposov S. E., 2018, MNRAS, 473, 2428
Deason A. J., Belokurov V., Sanders J. L., 2019, MNRAS, 490, 3426
Dolag K., Borgani S., Murante G., Springel V., 2009, MNRAS, 399, 497
Dotter A. et al., 2010, ApJ, 708, 698
Dotter A., Sarajedini A., Anderson J., 2011, ApJ, 738, 74
Durrell P. R. et al., 2014, ApJ, 794, 103
Eggen O. J., Lynden-Bell D., Sandage A. R., 1962, ApJ, 136, 748
El-Badry K., Quataert E., Weisz D. R., Choksi N., Boylan-Kolchin M., 2019, MNRAS, 482, 4528
Fakhouri O., Ma C.-P., Boylan-Kolchin M., 2010, MNRAS, 406, 2267

Fattahi A. et al., 2020, MNRAS, 497, 4459
Fernández-Alvar E. et al., 2018, ApJ, 852, 50
Forbes D. A., 2020, MNRAS, 493, 847
Forbes D. A., Bridges T., 2010, MNRAS, 404, 1203
Forbes D. A. et al., 2018a, Proc. R. Soc. A, 474, 20170616
Forbes D. A., Read J. I., Gieles M., Collins M. L. M., 2018b, MNRAS, 481, 5592
Gaia Collaboration, 2018, A&A, 616, A12
Harris W. E., 1996, AJ, 112, 1487
Harris W. E., Harris G. L. H., Alessi M., 2013, ApJ, 772, 82
Harris W. E., Blakeslee J. P., Harris G. L. H., 2017, ApJ, 836, 67
Haywood M., Di Matteo P., Lehnert M. D., Katz D., Gómez A., 2013, A&A, 560, A109
Haywood M., Di Matteo P., Lehnert M. D., Snaith O., Khoperskov S., Gómez A., 2018, ApJ, 863, 113
Helmi A., 2020, ARA&A, 58, in press
Helmi A., White S. D. M., de Zeeuw P. T., Zhao H., 1999, Nature, 402, 53
Helmi A., Babusiaux C., Koppelman H. H., Massari D., Veljanoski J., Brown A. G. A., 2018, Nature, 563, 85
Horta D. et al., 2020a, MNRAS, preprint (arXiv:2007.10374)
Horta D. et al., 2020b, MNRAS, 493, 3363
Hudson M. J., Harris G. L., Harris W. E., 2014, ApJ, 787, L5
Hughes M. E., Pfeffer J., Martig M., Bastian N., Crain R. A., Kruijssen J. M. D., Reina-Campos M., 2019, MNRAS, 482, 2795
Hughes M. E., Pfeffer J. L., Martig M., Reina-Campos M., Bastian N., Crain R. A., Kruijssen J. M. D., 2020, MNRAS, 491, 4012
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Ibata R. A., Gilmore G., Irwin M. J., 1994, Nature, 370, 194
Ibata R. A., Malhan K., Martin N. F., 2019, ApJ, 872, 152
Johnston K. V., Bullock J. S., Sharma S., Font A., Robertson B. E., Leitner S. N., 2008, ApJ, 689, 936
Keller B. W., Kruijssen J. M. D., Pfeffer J., Reina-Campos M., Bastian N., Trujillo-Gomez S., Hughes M. E., Crain R. A., 2020, MNRAS, 495, 4248
Koppelman H. H., Helmi A., Massari D., Roelenga S., Bastian U., 2019a, A&A, 625, A5
Koppelman H. H., Helmi A., Massari D., Price-Whelan A. M., Starkenburg T. K., 2019b, A&A, 631, L9
Kruijssen J. M. D., 2015, MNRAS, 454, 1658
Kruijssen J. M. D., 2019, MNRAS, 486, L20
Kruijssen J. M. D., Pelupessy F. I., Lamers H. J. G. L. M., Portegies Zwart S. F., Icke V., 2011, MNRAS, 414, 1339
Kruijssen J. M. D., Pelupessy F. I., Lamers H. J. G. L. M., Portegies Zwart S. F., Bastian N., Icke V., 2012, MNRAS, 421, 1927
Kruijssen J. M. D., Pfeffer J. L., Crain R. A., Bastian N., 2019a, MNRAS, 486, 3134
Kruijssen J. M. D., Pfeffer J. L., Reina-Campos M., Crain R. A., Bastian N., 2019b, MNRAS, 486, 3180
Law D. R., Majewski S. R., 2010, ApJ, 718, 1128
Leaman R., VandenBerg D. A., Mendel J. T., 2013, MNRAS, 436, 122
Li H., Gnedin O. Y., 2014, ApJ, 796, 10
McConnachie A. W. et al., 2009, Nature, 461, 66
Mackereth J. T., Bovy J., 2020, MNRAS, 492, 3631
Mackereth J. T. et al., 2019, MNRAS, 482, 3426
Marín-Franch A. et al., 2009, ApJ, 694, 1498
Martin N. F., Ibata R. A., Bellazzini M., Irwin M. J., Lewis G. F., Dehnen W., 2004, MNRAS, 348, 12
Massari D., Koppelman H. H., Helmi A., 2019, A&A, 630, L4
Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121
Moster B. P., Naab T., White S. D. M., 2019, preprint (arXiv:1910.09552)
Myeong G. C., Evans N. W., Belokurov V., Sanders J. L., Koposov S. E., 2018, ApJ, 863, L28
Myeong G. C., Vasiliev E., Iorio G., Evans N. W., Belokurov V., 2019, MNRAS, 488, 1235
Niederste-Ostholt M., Belokurov V., Evans N. W., Peñarrubia J., 2010, ApJ, 712, 516
Niederste-Ostholt M., Belokurov V., Evans N. W., 2012, MNRAS, 422, 207
Papovich C. et al., 2015, ApJ, 803, 26
Pedregosa F. et al., 2011, J. Mach. Learn. Res., 12, 2825
Peñarrubia J. et al., 2005, ApJ, 626, 128
Pfeffer J., Kruijssen J. M. D., Crain R. A., Bastian N., 2018, MNRAS, 475, 4309
Pfeffer J., Bastian N., Crain R. A., Kruijssen J. M. D., Hughes M. E., Reina-Campos M., 2019a, MNRAS, 487, 4550
Pfeffer J., Bastian N., Kruijssen J. M. D., Reina-Campos M., Crain R. A., Usher C., 2019b, MNRAS, 490, 1714
Pfeffer J. L., Trujillo-Gomez S., Kruijssen J. M. D., Crain R. A., Hughes M. E., Reina-Campos M., Bastian N., 2020, MNRAS, preprint (arXiv:2003.00076)
Pillepich A. et al., 2014, MNRAS, 444, 237
Planck Collaboration I, 2014, A&A, 571, A1
Qu Y. et al., 2017, MNRAS, 464, 1659
Reina-Campos M., Kruijssen J. M. D., Pfeffer J., Bastian N., Crain R. A., 2018, MNRAS, 481, 2851
Reina-Campos M., Kruijssen J. M. D., Pfeffer J. L., Bastian N., Crain R. A., 2019, MNRAS, 486, 5838
Reina-Campos M., Hughes M. E., Kruijssen J. M. D., Pfeffer J. L., Bastian N., Crain R. A., Koch A., Grebel E. K., 2020, MNRAS, 493, 3422
Schaye J. et al., 2015, MNRAS, 446, 521
Searle L., Zinn R., 1978, ApJ, 225, 357
Shipp N. et al., 2018, ApJ, 862, 114
Snaith O. N., Haywood M., Di Matteo P., Lehnert M. D., Combes F., Katz D., Gómez A., 2014, ApJ, 781, L31
Snaith O., Haywood M., Di Matteo P., Lehnert M. D., Combes F., Katz D., Gómez A., 2015, A&A, 578, A87
Spitler L. R., Forbes D. A., 2009, MNRAS, 392, L1
Springel V., Yoshida N., White S. D. M., 2001, New Astron., 6, 79
Trujillo-Gomez S., Kruijssen J. M. D., Reina-Campos M., Pfeffer J. L., Keller B. W., Crain R. A., Bastian N., Hughes M. E., 2020, MNRAS, preprint (arXiv:2005.02401)
Usher C., Pfeffer J., Bastian N., Kruijssen J. M. D., Crain R. A., Reina-Campos M., 2018, MNRAS, 480, 3279
VandenBerg D. A., Brogaard K., Leaman R., Casagrande L., 2013, ApJ, 775, 134
van der Walt S., Colbert S. C., Varoquaux G., 2011, Comput. Sci. Eng., 13, 22
van Dokkum P. et al., 2018, ApJ, 856, L30
Vasiliev E., 2019, MNRAS, 484, 2832
Waskom M. et al. , 2020, mwaskom/seaborn: v0.10.0 (January 2020). available at: https://zenodo.org/record/3629446#.X0KQ08gzbIU

This paper has been typeset from a TEX/LATEX file prepared by the author.