

Predictions for the detection of tidal streams with *Gaia* using great-circle methods

Cecilia Mateu,^{1,2★} Andrew P. Cooper,³ Andreea S. Font,⁴ Luis Aguilar,² Carlos Frenk,³ Shaun Cole,³ Wenting Wang³ and Ian G. McCarthy⁴

¹Centro de Investigaciones de Astronomía, AP 264, Mérida 5101-A, Venezuela

²Instituto de Astronomía, Universidad Nacional Autónoma de México, Apartado Postal 877, 22860 Ensenada, B.C., Mexico

³Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK

⁴Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK

Accepted 2017 April 6. Received 2017 April 5; in original form 2016 August 24

ABSTRACT

The *Gaia* astrometric mission may offer an unprecedented opportunity to discover new tidal streams in the Galactic halo. To test this, we apply nGC3, a great-circle-cell count method that combines position and proper motion data to identify streams, to 10 mock *Gaia* catalogues of K giants and RR Lyrae stars constructed from cosmological simulations of Milky Way analogues. We analyse two sets of simulations, one using a combination of *N*-body and semi-analytical methods, which has extremely high resolution, the other using hydrodynamical methods, which captures the dynamics of baryons, including the formation of an *in situ* halo. These 10 realizations of plausible Galactic merger histories allow us to assess the potential for the recovery of tidal streams in different Milky Way formation scenarios. We include the *Gaia* selection function and observational errors in these mock catalogues. We find that the nGC3 method has a well-defined detection boundary in the space of stream width and projected overdensity, which can be predicted based on direct observables alone. We predict that about 4–13 dwarf galaxy streams can be detected in a typical Milky Way-mass halo with *Gaia*+nGC3, with an estimated efficiency of >80 per cent inside the detection boundary. The progenitors of these streams are in the mass range of the classical dwarf galaxies and may have been accreted as early as redshift ~ 3 . Finally, we analyse how different possible extensions of the *Gaia* mission will improve the detection of tidal streams.

Key words: methods: data analysis – Galaxy: evolution – dark matter.

1 INTRODUCTION

The *Gaia* mission, whose first data release is now publicly available, is expected to revolutionize our knowledge of the formation of the Milky Way (MW), by mapping, for the first time, close to a billion stars in the disc, bulge and halo with exquisite astrometric precision (Perryman et al. 2001; de Bruijne 2012). It is anticipated that this detailed information will enable a breakthrough in understanding the formation history of the MW.

The stellar halo, in particular, holds a wealth of information about the merger history of the Galaxy, being a repository of most of the tidal debris from the past merger events. The number of tidal streams surviving at the present day in the halo, their morphologies, their total luminosities and their chemical abundance patterns, all encode important information from which the series of accretion events can be reconstructed (Helmi & White 1999; Bullock & Johnston 2005;

Johnston et al. 2008; Cooper et al. 2010; Helmi et al. 2011). Tidal streams can also be used to infer the gravitational potential of the MW (e.g. Price-Whelan & Johnston 2013; Sanderson, Helmi & Hogg 2015; Sanderson 2016). Increasing the number of stream detections can improve this measurement (Deg & Widrow 2014).

While it is expected that *Gaia* will uncover new tidal streams in the halo (Helmi & de Zeeuw 2000; Gómez & Helmi 2010), quantitative theoretical predictions for the likely number of such discoveries have not been made to date, mainly because of the uncertainties in modelling the physical processes associated with baryons in the framework of hierarchical structure formation, the computational resolution and re-sampling issues associated with producing adequate simulated catalogues at the level of individual stars, and the need to develop algorithms for making such detections by mining the *Gaia* data set. In this study, we aim to make progress by employing a series of state-of-the-art simulations of MW-mass haloes from which we construct mock *Gaia* star catalogues, which we search for tidal streams with a robust, quantifiable method.

★ E-mail: cmateu@cida.gob.ve

We use two suites of cosmological simulations to produce the mock *Gaia* catalogues: the Aquarius simulations, a set of high-resolution dark-matter-only simulations of MW-mass haloes (Springel et al. 2008a), combined with the GALFORM semi-analytic prescriptions (Cooper et al. 2010); and a second set, called HYDRO-zooms, which comprises several medium-resolution hydrodynamical simulations of MW-mass disc galaxies (Font et al., in preparation, hereafter F17), the initial conditions of which were extracted from the EAGLE simulation (Schaye et al. 2015). Aquarius allows us to study tidal streams from progenitors that span a wide range of masses and orbits, and hence to test our method on a realistic set of stream luminosities and morphologies. On the other hand, the HYDRO-zooms, although of lower resolution than Aquarius, have the benefit of modelling the hydrodynamical effects of baryons self-consistently. Baryonic effects, including modification of the density profiles of satellites by stellar feedback and interactions between satellites and the central stellar disc, may alter the morphology of tidal streams, and, together with the possible presence of an *in situ* halo, this may change (most likely decrease) the number of streams that can be detected. The objective of this paper is not to perform a detailed comparison between these two simulation techniques, but rather to estimate the detectability of the tidal streams they predict.

This work goes beyond earlier studies of tidal stream detection in several ways. For the first time, we make predictions based on fully cosmological simulations of MW-mass galaxies that we combine with the most up-to-date *Gaia* error estimates and selection function. The simulated tidal streams evolve within a realistic gravitational potential (non-axisymmetric and changing in time). Thus, the mock *Gaia* star catalogues constructed here complement existing *Gaia* mocks that do not include substructure in the stellar halo (e.g. Robin et al. 2012). Examining a number of MW-mass haloes with a variety of merger histories helps to make our predictions robust against our ignorance of the details of the Galaxy’s accretion history. This is a step forward towards comparing the models and observations on a level playing field. Also, with the HYDRO-zoom simulations, the effect of halo component formed *in situ* (Zolotov et al. 2010; Font et al. 2011a; McCarthy et al. 2012; Cooper et al. 2015; Pillepich, Madau & Mayer 2015) on the detectability of tidal streams can be taken into account. To our knowledge, the contaminating effect of combined *in situ* and accreted halo components has only been estimated for *Gaia* by Brown, Velazquez & Aguilar (2005) and Mateu et al. (2011), who embedded a set of stellar streams in a smooth Galactic background with a constrained luminosity normalization. However, these streams were evolved in a fixed axisymmetric potential and their progenitors selected ad hoc.

Rather than starting from the information available in the simulations, in which every star particle and hence every stream can be identified unambiguously with a specific progenitor, we first apply an observational stream-finding algorithm based on the great-circle-cell counts (GC3) method. This method, described in detail below, uses combined positions and proper motions to assign stars to discrete groups with common orbital poles. GC3 methods are an efficient way to search for tidal streams in the Galactic halo. They exploit the fact that streams will be approximately confined to planes in potentials that are close to spherical, by searching for overdensities of stars along great circles (as seen from the Galactic Centre). The idea was initially proposed by Lynden-Bell & Lynden-Bell (1995) and Johnston, Hernquist & Bolte (1996) and later modified by Mateu et al. (2011, hereafter M11) to improve its efficacy by including kinematical information (mGC3), with the *Gaia*

mission in mind. Its main advantage is that, with the implementation proposed in M11, the GC3 family of methods works directly in observable space (positions, parallax, proper motion, radial velocity), rather than using physical parameters such as energy or angular momentum, greatly reducing the effect of the propagation of observational errors, which Brown et al. (2005) have shown can be quite substantial even for *Gaia*.

Finally, we assess the efficiency of our stream detection method by using our knowledge of the ‘true’ population of streams in the simulations to determine which progenitors are recovered and with what ‘purity’. This knowledge of the method’s efficiency and selection biases will be a key ingredient in the inverse process of inferring the Galactic accretion history.

The paper is structured as follows: Section 2 summarizes the simulations employed in this study. Section 3 describes the construction of the mock *Gaia* catalogues and the *Gaia* error simulation. Section 4 presents what *Gaia*-like surveys would ‘see’ in the simulated stellar haloes based on a selection of specific stellar tracers. Section 5 describes the great-circle method used to identify tidal streams. The appearance of observable tidal streams in the diagnostic space of the method, which we call pole-count maps, is explored in detail for a fiducial halo in Section 6. Section 7 summarizes the results of applying our algorithm to all the other haloes in our sample. In Section 8, we investigate the properties of progenitors of the streams that are detected in the mock *Gaia* surveys of our simulations. In Section 9, we analyse how the detectability of tidal streams changes under various scenarios for extending the lifetime of the *Gaia* mission. Finally, in Section 10, we discuss several ways in which this stream-finding method can be further improved and give a summary of our conclusion in Section 11.

2 COSMOLOGICAL SIMULATIONS

2.1 Aquarius simulations

Aquarius is a set of six collisionless cosmological ‘zoom’ simulations of individual dark matter haloes of mass $\sim 10^{12} M_{\odot}$ (Springel et al. 2008a,b; Navarro et al. 2010). The simulations assume a Λ cold dark matter (Λ CDM) cosmogony with parameters determined from the *Wilkinson Microwave Anisotropy Probe* 1-year results (Spergel et al. 2003) and the 2dF Galaxy Redshift Survey data (Colless et al. 2001): $\Omega_M = 0.25$, $\Omega_{\Lambda} = 0.75$, $n_S = 1$, $\sigma_8 = 0.9$ and Hubble parameter $h = 0.73$. The six haloes were selected randomly from a parent sample of isolated haloes of similar mass in a lower resolution $(100 h^{-1})^3 \text{ Mpc}^3$ cosmological volume simulation (Gao et al. 2008). Isolation was defined by the absence of any neighbours with more than half the mass of the target halo within $1 h^{-1} \text{ Mpc}$. A Lagrangian region several times larger than the $z = 0$ virial radius of each target halo was re-simulated with a much larger number of lower mass particles, coarsely sampling the surrounding large-scale structure with a smaller number of higher mass particles, subject to exactly the same spectrum of initial density perturbations.

The Aquarius simulations are labelled Aq-A to Aq-F; we do not use Aq-F in this paper because its recent merger history makes it highly unlikely to be representative of a system like the MW (Boylan-Kolchin et al. 2010; Cooper et al. 2010). We use the level 2 set of simulations, the highest resolution level at which all six haloes were simulated. The particle mass varies slightly between

the level 2 simulations in the range $0.6 < m_p (\times 10^4 M_\odot) < 1.4$. The Plummer-equivalent gravitational softening length is $\epsilon \sim 66$ pc.

The Aquarius simulations use a single high-resolution particle species to model the collisionless dynamics of both dark matter and baryons. To represent the stellar component, we use the ‘particle tagging’ models described by Cooper et al. (2010). This technique first uses a semi-analytic galaxy formation model to determine the star formation history of each dark matter halo in the simulation, and then applies dynamical criteria to select subsets of collisionless particles occupying regions in phase space associated with each distinct single-age stellar population at the time of its formation. The Cooper et al. (2010) technique improves on earlier tagging approaches (e.g. Bullock & Johnston 2005) in the use of a single, self-consistent cosmological simulation to treat the dynamics of the satellites and the host halo, and in the use of a galaxy formation model constrained by large cosmological data sets as well as the properties of MW and M31 satellites (Bower et al. 2006; Cooper et al. 2010; Font et al. 2011b). The five Aquarius simulations we use show considerable diversity in the properties of their stellar haloes, owing to their range of virial masses and, more significantly, to the intrinsically stochastic nature of dwarf galaxy accretion and disruption in Λ CDM.

The particle tagging technique involves a dynamical approximation with clear limitations, and unlike Bullock & Johnston (2005) the Cooper et al. (2010) simulations do not include the gravitational contribution of a massive stellar disc at the centre of the host potential. The presence or absence of a disc may accelerate the tidal disruption of some satellites. This is likely to affect predominantly those substructures with orbits passing through the inner ~ 20 kpc of the galaxy after $z > 2$, which nevertheless may include satellites and streams located far from the disc at $z = 0$. Errani et al. (2017) find that the total number of potentially luminous subhaloes disrupted in the inner region of the halo changes by a factor of ~ 2 when an idealized disc component is added to the potential in one of the Aquarius simulations. They demonstrate that the inner slope of the satellite mass density profile (which depends on the physics of galaxy formation) has an even larger effect on the number of surviving satellites (almost an order of magnitude; the conclusions of Errani et al. relate only to whether or not an identifiable self-bound core survives, rather than to the presence of tidal streams). Likewise, Garrison-Kimmel et al. (2017) find a factor of 2–5 depletion of massive subhaloes in a dark-matter-only simulation when they introduce a growing analytic disc potential based on a hydrodynamical realization from the same initial conditions. Our results here concern the disruption of well-resolved satellites with very high mass-to-light ratios, predominantly in the outer halo; for further discussion of related issues, we refer the reader to Cooper et al. (2010, 2013, 2016) and Le Bret et al. (2015). The number of these more distant satellite haloes surviving at $z = 0$ may therefore be considered uncertain by no more than a factor of ~ 2 as the result of neglecting the (still somewhat uncertain) influence of a disc potential. As we describe in the following subsection, we also analyse a suite of lower resolution gas dynamical simulations that account self-consistently for the gravitational effects of baryons neglected by the particle tagging approach. This allows us to check for large differences in the number of streams from bright satellites that could be due to the presence of a disc, albeit in the context of only one hydrodynamical model and in different dark matter haloes to Aquarius. If the MW disc has significantly depleted the number of luminous satellite subhaloes surviving to $z = 0$, our predictions based on Aquarius are likely to provide a lower limit to the total number of streams that *Gaia* will discover.

2.2 Gas dynamical simulations

For the gas dynamical simulations, we use a suite of ‘zoom’ simulations of MW-mass haloes using the high-resolution ‘Recal’ model from the recent EAGLE project (Crain et al. 2015; Schaye et al. 2015). The zoom simulations will be described in more detail in a future study (F17), so we provide only a brief description here.

We recall that the main aim of the EAGLE project was to simulate, at relatively high resolution (baryon particle mass $\approx 10^6 M_\odot$, softening length of 500 pc), the evolution of the main galaxy population. The stellar and AGN feedback parameters were adjusted so as to reproduce the observed galaxy stellar mass function and the size–mass relation of local galaxies. Unfortunately, the resolution of the main EAGLE box (L100N1504) is too low for our purposes, motivating our use of significantly higher resolution zoom simulations. Note that Schaye et al. (2015) have found that when the resolution is increased, some re-calibration of the stellar and AGN feedback is required to preserve a match to the galaxy stellar mass function. Using this re-calibrated model (called ‘Recal’), they have simulated a 25 Mpc volume with a factor of 8 (2) better mass (spatial) resolution (i.e. L025N0752). This simulation volume served as the parent volume from which several haloes were selected for re-simulation.

Specifically, F17 identified a volume-limited sample of 25 haloes that fall in the mass range $7 \times 10^{11} < M_{200}/M_\odot < 3 \times 10^{12}$ at $z = 0$ (M_{200} denotes the mass within the virial radius r_{200}). Inspection of the visual morphologies indicates that not all of these systems have significant stellar disc components. While such systems are interesting in their own right (and the intention is to eventually simulate all 25 haloes), priority was given to 10 systems that have the most disc-like morphology. F17 have carried out zoom simulations with a factor of 8 (2) better mass (spatial) resolution than the parent volume (i.e. baryon particle mass of $\approx 1.5 \times 10^4 M_\odot$, Plummer-equivalent softening length of 125 pc) using the Recal model.¹ For further details of the Recal model, including a description of the employed hydrodynamic solver and subgrid prescriptions for radiative cooling, star formation, stellar and chemical evolution, and feedback, we refer the reader to Schaye et al. (2015).

In the present study, we analyse a random subset of 5 of the 10 zoom simulations carried out by F17. At $z = 0$, this subset spans virial masses $7.14 \times 10^{11} < M_{200}/M_\odot < 1.93 \times 10^{12}$ and stellar masses $7.33 \times 10^9 < M_*(< 30 \text{ kpc})/M_\odot < 1.99 \times 10^{10}$, respectively, similar to the corresponding values of the five Aquarius haloes (Cooper et al. 2010). Apart from the fact that these galaxies resemble the MW in terms of total and stellar mass, the properties of their bound substructure also match the main properties of MW satellites, e.g. the luminosity function and the stellar mass–metallicity relation [see also Schaye et al. (2015) for the properties of low-mass galaxies in the Recal model]. A more detailed investigation of the properties of these galaxies will be presented in F17. We note, however, that, due to the limited numerical resolution, these gas dynamical zoom simulations can follow reliably only the properties of satellites in the classical dwarf galaxies regime ($M_* \geq 10^7 M_\odot$).

Following the methods described in Font et al. (2011a), we construct simple merger histories for each of the simulated galaxies, identifying which star particles were formed *in situ* (i.e. within the

¹ No additional re-calibration of the model was performed when increasing the resolution beyond that of the Recal-L025N0752 parent volume, but F17 have verified that the stellar masses of the zoomed haloes agree with those of the parent volume to typically better than 10 per cent.

main progenitor branch), which were brought in via mergers/tidal disruption of infalling satellites and which star particles still reside in orbiting satellites at the present day. For the star particles that were/are in satellites, we record the properties of the halo to which the particles belonged just prior to joining the main friends-of-friends group.

The HYDRO-zoom simulations have the benefit of treating various baryonic physical processes, such as gas infall, star formation and stellar feedback, self-consistently. Anticipating the results, we expect that the gas dynamical simulations will obtain a somewhat different number of tidal streams and different stream morphologies, than in the case of the particle tagging methodology. For example, the stellar feedback may change the internal spatial and kinematical distributions of stars in satellite galaxies and may transform cuspy density profiles into cored ones. This, in turn, can affect the rate at which material is tidally stripped from satellites that changes the time when tidal streams are formed and their morphological properties. The presence of a disc may influence the spatial distribution of satellites in the inner region of the galaxy, by inducing changes in the orientation of their angular momentum and by accelerating their tidal disruption. Additionally, the hydrodynamical gas dynamical simulations have been shown to produce stellar haloes with dual components: accreted and *in situ* (Zolotov et al. 2010; Font et al. 2011a; McCarthy et al. 2012; Cooper et al. 2015; Pillepich et al. 2015). We caution that the origin of the *in situ* component is still debated, current gas dynamical simulations suggesting different scenarios: stars being ejected from the disc by disc–satellite interactions, or formed in the wake of the gas stripped from infalling satellites, or formed in cold gas filaments. Understanding the origin of *in situ* halo stars is crucial for predicting the physical properties of this halo component and, implicitly, for modelling the environment in which tidal streams evolve. Strictly from the point of view of the detectability of tidal streams, the *in situ* component of the stellar halo is another source of foreground/background contamination, whose effect needs to be assessed.

Overall, the additional effects present in gas dynamical simulations are expected to diminish the number of tidal streams that are dynamically cold at present day and therefore, those that are most likely to be detected. We note, however, that this discussion is mainly qualitative at this point, and a more rigorous assessment of the significance of the various baryonic processes will require an in-depth quantitative investigation. This is, however, beyond the scope of this present paper since the differences in the initial conditions and numerical resolution between the two types of simulations presented here do not allow for a fair comparison. In the case of these two types of simulations, we estimate that the main differences in the number of tidal streams are most likely due to the differences in the numerical resolution. We note, however, that in the range in which the HYDRO-zoom simulations are able to resolve the halo substructure – roughly, the domain of the classical dwarf galaxies –, the two types of simulations predict similar number of surviving satellites and of tidal streams.

3 MOCK GAIA CATALOGUES

3.1 Re-sampling the simulations

3.1.1 Phase space expansion of tracer particles

We use the method described by Lowing et al. (2015) to convert the massive ‘star particles’ in our simulations into mock catalogues of individual stars. Briefly, the steps are as follows. Star particles are

partitioned into disjoint sets according to the progenitor subhalo to which they were bound at the time of infall into the MW analogue halo (for this purpose, a small number of particles not bound to any halo at the time of infall and stars formed *in situ* are classified as a single set). The ENBID code (Sharma & Steinmetz 2006) is run separately on each of these sets to estimate the six-dimensional phase space volume associated with every star particle (the separation into sets avoids cross-talk between different streams in this estimate). The volume identified by ENBID is translated to an equivalent 6D Gaussian kernel. A sample of mock stars is generated from an isochrone appropriate to the stellar population represented by the parent star particle (we use the PARSEC isochrones from Bressan et al. 2012) and positions and velocities assigned to each of these by randomly sampling from the kernel. The advantage of using a 6D smoothing kernel is that ‘thin’ structures in configuration and velocity space are preserved – mock stars are distributed preferentially ‘along’ the streams defined by their parent particles, rather than orthogonal to them, as would be the case for an isotropic kernel.

As described in Lowing et al. (2015), mock catalogues of stars in the Aquarius simulations [based on a slightly updated version of the Cooper et al. (2010) galaxy formation model] are publicly available as online data bases.² Our Aquarius simulation catalogues were drawn from these data bases according to the criteria described in the following section. Analogous catalogues for the HYDRO-zoom simulations were generated by applying the Lowing et al. (2015) procedure in the same way as for the particle tagging models.

3.1.2 Stellar tracers

To produce *Gaia* mock catalogues, we generate samples of K giants and RR Lyrae stars (RRLs), two bright stellar tracers that can be observed by *Gaia* to large distances with reasonably small proper motion errors (see Section 3.2). Both tracers have been widely used in Galactic halo surveys (e.g. Morrison et al. 2000; Vivas et al. 2004; Sesar et al. 2009, 2013; Starkenburg et al. 2009; Xue et al. 2014, and references therein). K giants are found in any stellar population older than a few giga years, of any metallicity, and they are bright ($1 < M_r < -3$) and relatively numerous (e.g. Xue et al. 2014). RRLs are pulsating horizontal branch (HB) giants that trace old (> 10 Gyr) and metal-poor populations ($[\text{Fe}/\text{H}] < -0.5$; e.g. Smith 1995); although they are sparser and not as luminous ($M_V \sim 0.55$) as the brightest K giants, RRLs are well known for being excellent standard candles.

We select K giants using the colour and M_g cuts described in Lowing et al. (2015, section 3.4.3), defined by Xue et al. (2014). These cuts select all simulated K giant stars brighter than the HB and filter out any red clump or red dwarf contaminants. To select RRLs, we use the effective temperature and surface gravity cuts suggested by Baker & Willman (2015, i.e. $6100 < T_{\text{eff}}(\text{K}) < 7400$, $2.5 < \log g < 3.0$).

When dealing with real data, the actual samples of K giants and RRLs will be prone to some degree of contamination. For K giant samples, contamination can come from foreground main-sequence dwarfs, which should be effectively filtered out as these will be nearby stars with very precise *Gaia* parallaxes (see Section 3.2). RRLs can be very reliably identified based on their photometric variability and well-known light-curve shapes, so little contamination from other types of stars is expected (see e.g. Vivas et al. 2004;

² <http://virgodb.dur.ac.uk:8080/StellarHalo>

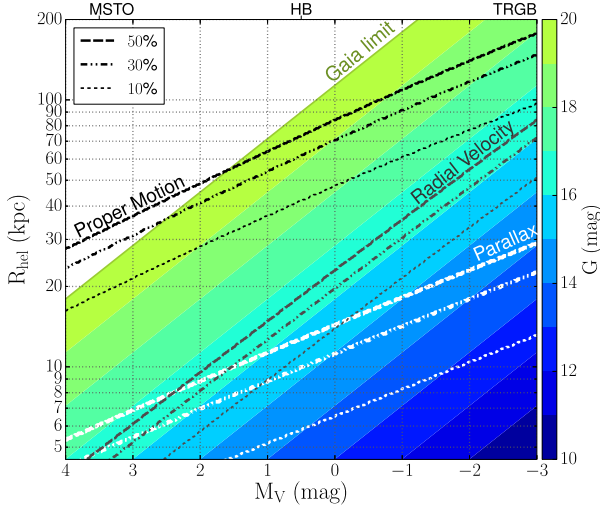


Figure 1. *Gaia* observational error horizons in the heliocentric distance R_{hel} versus absolute magnitude M_V plane. The colour scale is proportional to the apparent G magnitude and goes up to the *Gaia* magnitude limit ($G \leq 20$), assuming a fixed $V - I = 1$ colour and $A_V = 0$. Only stars brighter than $G = 16$ will have radial velocity measurements (light to dark blue areas). Dashed, dash-dotted and dotted lines, respectively, represent loci of 10, 30 and 50 per cent relative errors in parallax (white), radial velocity (grey) and proper motion (black). The absolute magnitudes of main-sequence turn-off (MSTO), HB and tip of the red giant branch (TRGB) stars are shown for reference on the top axis. The error estimates, as well as the limiting magnitudes cited, already take into account the stray light effect (de Bruijne et al. 2014).

Mateu et al. 2012). Therefore, in both cases, we expect the effect of contamination to be small and manageable.

3.2 The *Gaia* errors simulation

We simulate observational errors using the *GAIA-ERRORS* software from Romero-Gómez et al. (2015),³ which implements the latest post-launch end-of-mission prescriptions provided by the Data Processing and Analysis Consortium (DPAC; Mignard et al. 2008) and described in Rygl et al. (2014).

The *GAIA-ERRORS* code simulates Gaussian errors for the positions, parallaxes, proper motions and radial velocities, with a standard deviation that depends on the apparent magnitude and colour of each star, accounting for the ecliptic latitude dependence introduced by the *Gaia* scanning law (de Bruijne 2012). Reddening is simulated based on the 3D extinction maps from Drimmel, Cabrera-Lavers & López-Corredoira (2003) while the *Gaia* selection function is assumed to have 100 per cent completeness down to $G = 20$ and 16, respectively, for the astrometric observables (position, parallax and proper motion) and for radial velocities (de Bruijne, Rygl & Antoja 2014). We simulate end-of-mission combined errors for the nominal lifetime of 5 yr for the *Gaia* mission (de Bruijne et al. 2014), although the *GAIA-ERRORS* code allows for the simulation of errors at an arbitrary mission operation time. In Section 9, we discuss the effect of possible extensions for the mission lifetime.

In Fig. 1, we illustrate the volume that can be probed with *Gaia* at a fixed relative precision, for stars of different brightness. The plot shows proper motion, radial velocity and parallax relative error

horizons in the heliocentric distance R_{hel} versus absolute magnitude M_V plane, and the colour scale is proportional to the apparent G magnitude. Dashed, dash-dotted and dotted lines, respectively, represent loci of 10, 30 and 50 per cent relative errors in parallax (white), proper motion (black) and radial velocity (grey). To be able to plot relative rather than absolute errors in proper motion and radial velocity, we have assumed that the stars are moving at half the escape velocity v_e for the corresponding distance.⁴ Therefore, for the proper motion and radial velocity, the different lines represent the *best* relative precision achievable at a given distance.

For bright stars ($M_V < 0$), Fig. 1 shows how *Gaia* can achieve remarkable relative proper motion precision of ~ 30 per cent for HB stars at large distances as ~ 60 kpc, and better than 30 per cent for brighter K giants ($-1.5 \lesssim M_V$) beyond 100 kpc. For main-sequence turn-off (MSTO) stars, the relative proper motion precision will be $\lesssim 10$ per cent, but these stars are only bright enough to be observable up to ~ 20 kpc.

Radial velocities will be available for the brightest stars with $G \leq 16$ (light to dark blue areas in the figure), all of which will have radial velocity relative errors smaller than 30 per cent. The maximum distance for HB and red giant branch stars with radial velocities will be ~ 15 kpc and ~ 30 – 60 kpc, respectively. The limiting magnitude of $G = 16$ for *Gaia* radial velocities is ~ 1 mag brighter than what was originally expected for the mission, due to the increased background caused by stray light (de Bruijne et al. 2014), which is also already taken into account in the errors shown in Fig. 1.

Gaia parallaxes will be useful only up to ~ 10 kpc for most stars, and even for the very brightest giants ($M_V \gtrsim -2$), reasonable errors (< 50 per cent) will be achieved only up to ~ 20 kpc. Hence, photometric distance measurements for standard candle tracers will be crucial to probe the outer halo.

3.2.1 Photometric distances

Fig. 1 illustrates that *Gaia* parallax errors will be prohibitively large beyond a few kpc. However, photometric distances can be estimated for the stellar tracers we have selected, with much better precision.

RRLSs are well-known standard candles, for which relative errors in distance are as low as ~ 7 per cent, or even 5 per cent, if there is a relatively small (0.3 dex) uncertainty in metallicity (Vivas & Zinn 2006; Mateu et al. 2012; Sesar et al. 2013). For K giants, the dependence of M_V on colour and metallicity makes photometric distance determinations more challenging. Xue et al. (2014) find that these can be estimated with a 16 per cent median error, based on gr photometry and spectroscopic metallicities from Sloan Extension for Galactic Understanding and Exploration, using their probabilistic algorithm. Liu et al. (2014) use 2MASS photometry and Large Sky Area Multi-Object Fibre Spectroscopic Telescope (LAMOST) spectra and get a mean distance error of ~ 30 per cent, which they attribute to the shallower photometry used in their procedure.

In what follows, we simulate photometric distances in our mock catalogues, with a constant (Gaussian) error of 7 per cent for RRLSs, and an intermediate value of 20 per cent for K giants.

⁴ We assume that the radial velocity is on average $v_r \sim v/\sqrt{3}$ and the total proper motion $\mu \sim \sqrt{2/3}v$, where we assume $v \sim v_e/2$ and approximate the escape velocity as $v_e = v_c \sqrt{2(1 - \ln(R_{\text{gal}}/r_t))}$, with $v_c = 200 \text{ km s}^{-1}$, $r_t = 200 \text{ kpc}$ and R_{gal} the Galactocentric distance.

³ The code is publicly available at <https://github.com/mromerog/Gaia-errors>

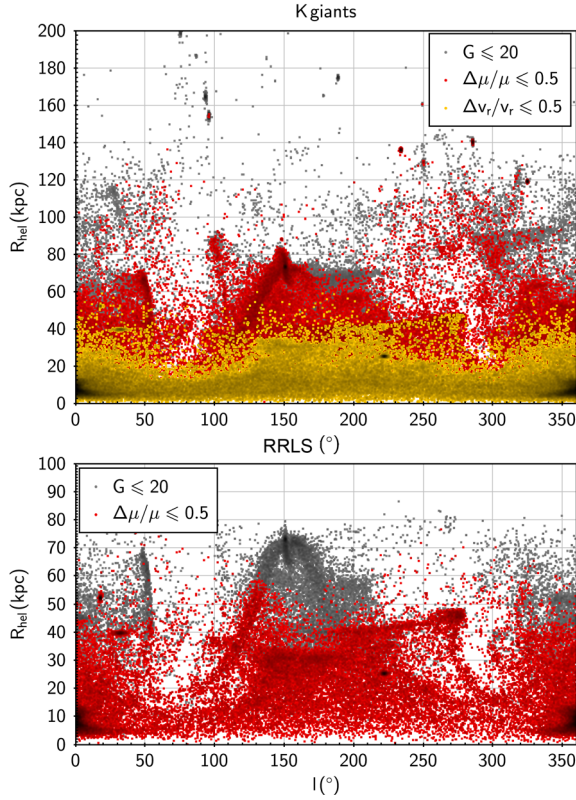


Figure 2. Heliocentric distance R_{hel} versus Galactic longitude l for K giants (top) and RRLSs (bottom) in the Aquarius A2 mock catalogue. *Gaia* observable stars are shown in grey, and stars with relative proper motion and radial velocity errors <50 per cent in red and ochre, respectively.

4 WHAT GAIA CAN SEE

What *Gaia* will be able to ‘see’ will ultimately be determined by the combination of different factors: the selected tracer, the effect of extinction that will depend on the line of sight and, for a chosen proper motion precision, the actual velocity distribution of the stars in the different tidal streams. Hence, the previous section and Fig. 1 provide a simplified description.

To illustrate this in a more realistic case, Fig. 2 shows two mock *Gaia* catalogues of the Aquarius A2 halo produced as we have described in Section 3.2: the upper panel for K giant stars, the bottom panel for RRLSs. The plot shows heliocentric distance R_{hel} versus Galactic longitude l for all stars observable by *Gaia* (grey) and stars for which *Gaia* proper motions (red) and radial velocities (ochre) have relative errors better than 50 per cent. In this case, we have used each star’s own individual proper motion and radial velocity to compute the relative errors.

In the upper panel, we can see that most structures up to about 100 kpc are traced by K giants with good proper motions. Beyond this, there are a few K giant stars with good proper motions observable even as far as ~ 150 kpc, in the denser structures that are more likely to host more of the intrinsically brightest K giants. On the other hand, note also that although all structures are very well traced by K giants below 80 kpc, there is a severe lack of observable stars with good proper motions in the tidal arm at $(l, R_{\text{hel}}) \sim (200^\circ, 70 \text{ kpc})$. This is a case of a stream that happens to have most of its velocity along the line of sight (not shown), and so for the typical proper motion precision attainable at this distance, the fractional proper motion error is larger than the imposed cut of

50 per cent. The volume that can be probed with K giants including radial velocities reaches out to ~ 40 kpc on average.

The volume that can be probed with RRLSs with good proper motions reaches out to ~ 40 kpc. Note that, incidentally, this is roughly the same volume inside which K giants will have full 6D information with good precision. So structures in this volume can in principle be tracked down with both tracers. We do not show the analogous coverage for RRLS radial velocities as the *Gaia* standard errors are end-of-mission error prescriptions for the combined spectra, and so do not apply for pulsating stars. For RRLSs and other pulsating variables, single-epoch spectra must be used in order to account for the pulsation component in the radial velocity; however, no error prescriptions are yet provided by DPAC for radial velocities from single-epoch spectra.

5 GREAT-CIRCLE METHODS

The GC3 method was proposed by Johnston et al. (1996) to search for tidal streams in the Galactic halo, by using the fact that stars that belong to a tidal stream produced in a spherical potential orbit in a fixed plane as the total angular momentum is conserved. This means that a tidal stream will lie approximately in a great-circle band, which is the projection of the orbital plane on to the Galactocentric celestial sphere. In fact, the idea of searching for great-circle alignments had already been introduced by Lynden-Bell & Lynden-Bell (1995), who looked at the intersection of great circles of orbital poles to search for alignments of dwarf galaxy satellites and globular clusters along great circles. These authors even proposed a way to include kinematic information, by assuming energy and angular momentum conservation.

In light of the (then) upcoming *Gaia* astrometric mission, starting with M11 we extended GC3 into a family of great-circle-cell methods that includes kinematic information: adding a total velocity criterion in M11 (the mGC3 method) and a proper-motion-only version in Abedi et al. (2014, the nGC3 method).

5.1 GC3, mGC3 and nGC3

In general, the GC3 family of great-circle-cell methods is defined by the use of a geometric selection criterion to choose stars along a great-circle band in the sky, orthogonal to a particular direction marked with what is called its *pole* vector $\hat{\mathbf{L}}$. The number of stars that satisfy this criterion are counted and assigned to this particular pole vector. The pole vector is then changed in direction following the nodes of a spherical coordinate mesh in the sky, and the whole operation is repeated at each node. This produces the so-called pole-count maps (hereafter PCMs) where the number of stars at each mesh node is indicated. Maxima in this map indicate the presence of substructure. To decide whether a star is associated with a given pole $\hat{\mathbf{L}}$, we use the following position and velocity criteria (equation 6 in M11):

$$|\hat{\mathbf{L}} \cdot \mathbf{r}'_{\text{gal}}| \leq \|\mathbf{r}'_{\text{gal}}\| \delta_c \quad (1)$$

$$|\hat{\mathbf{L}} \cdot \mathbf{v}'_{\text{gal}}| \leq \|\mathbf{v}'_{\text{gal}}\| \delta_c \quad (2)$$

where $\delta_c = \sin \delta\theta$ is sine of the tolerance that allows for the width $\delta\theta$ of each great-circle band and \mathbf{r}'_{gal} and \mathbf{v}'_{gal} are simply the Galactocentric position and velocity vectors \mathbf{r}_{gal} and \mathbf{v}_{gal} , multiplied by

the parallax,⁵ which in terms of the heliocentric observables (l , b , ϖ , μ_l , μ_b , v_r) are given by

$$\begin{aligned} \mathbf{r}'_{\text{gal}} &= \varpi \mathbf{r}_{\odot} + A_p (\cos l \cos b) \hat{\mathbf{x}} + (\sin l \cos b) \hat{\mathbf{y}} + (\sin b) \hat{\mathbf{z}} \\ \mathbf{v}'_{\text{gal}} &= \varpi \mathbf{v}_{\odot} + \varpi v_r \hat{\mathbf{r}} + (A_p \mu_l \cos b) \hat{\mathbf{l}} + (A_p \mu_b) \hat{\mathbf{b}}, \end{aligned} \quad (3)$$

where $A_p = 10^3$ mas pc, $A_v = 4.74047$ yr km s⁻¹; $\{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$ are the unit vectors in the Cartesian Galactocentric reference frame and $\{\hat{\mathbf{r}}, \hat{\mathbf{l}}, \hat{\mathbf{b}}\}$ are the unit vectors in a spherical heliocentric reference frame (see appendix A in M11).

A significant advantage of the GC3 family of methods is that the computation of PCMs works directly in observable space (see M11 for a detailed description), instead of working with physical quantities like velocity, energy or angular momentum, for which errors are propagated in complicated ways because of the non-linearity of the transformations involved.

Each of the variants in the GC3 method family⁶ associates stars with poles with different combinations of the criteria in equations (1) and (2):

- (i) GC3: 3D positional information only (equation 1);
- (ii) mGC3: 3D position and 3D velocity (equations 1 and 2);
- (iii) nGC3: 3D position and proper motion (equations 1 and 2, without the $\varpi v_r \hat{\mathbf{r}}$ term in equation 3).

As we have shown in Sections 3.2 and 4, demanding radial velocities will severely restrict the volume we can probe with *Gaia*. In what follows, we will use the nGC3 method, which only requires proper motions, and we will also limit our samples to stars with proper motion relative errors less than 50 per cent, as we will discuss in more detail in Section 6.1.

Using proper motions alone and disregarding radial velocities has the advantage of allowing us to probe a much larger volume of the halo, as we have shown in Fig. 2. Also, for distant streams ignoring v_r makes no difference; as the Sun–Galactic Centre (GC) distance becomes negligible, the radial component of the velocity is approximately contained in the orbital plane by construction, so its contribution to the dot product in equation (2) will tend to zero. On the other hand, ignoring v_r will affect the contribution of contaminants to the PCMs. For planes roughly perpendicular to the Sun–GC direction ($\phi_{\text{pole}} \sim 0^\circ, 180^\circ$), the line of sight is off the plane, so in these directions fewer foreground/background contaminants will be filtered, and the PCM background level will be higher than for planes going through the Sun–GC line ($\phi_{\text{pole}} \sim 90^\circ, 270^\circ$).

5.2 Signatures of individual streams in PCMs

In this section, we will illustrate how streams produced in cosmological simulations do in fact produce recognizable peaks in the PCMs. This is a crucial test, as in all previous applications of the GC3 methods, we have used N -body simulations in a fixed axisymmetric potential, where substructure may be unrealistically enhanced against a smooth background. In cosmological simulations, the haloes are gradually assembled through time and so the potential is neither fixed nor even necessarily axisymmetric, so it is not obvious that the GC3 methods can still be applied in this case (note that this is also the case for the real MW).

⁵ This is done in order to avoid using the reciprocal of the parallax that would introduce a bias. See M11.

⁶ The PYTHON toolkit PYMGC3 provides an implementation of the GC3 family of methods and is publicly available at <https://github.com/cmateur/PYMGC3>

The morphology of PCMs for different progenitors is illustrated in Fig. 3. Five progenitors at different stages of disruption were chosen from the Aquarius and HYDRO-zoom haloes. In the figure, each column corresponds to one progenitor; for each, the top row shows the corresponding nGC3 PCM, the middle and bottom rows show two orthogonal projections: Y versus X and Z versus X , respectively. The nGC3 PCMs are plotted in a north-polar azimuthal equatorial projection showing the north pole at the centre; the concentric circles are parallels drawn at 20° intervals, and meridians are drawn at 30° intervals in longitude with $\phi = 90^\circ, 180^\circ$ in the right horizontal and top vertical axes at the centre of the plot. The PCMs were computed with a tolerance $\delta\theta = 1:5$ on a uniform grid with 1° spacing.

The first column (A2-117) shows the nGC3 PCM for a completely bound progenitor. The signature is a very localized peak in pole counts around the orbital plane's pole, thanks to the use of kinematical information (proper motions in this case). Not all bound progenitors will necessarily produce such a well-localized peak in an nGC3 or mGC3 PCM. The maxima in the pole counts will tend to stretch more and more along a great circle for more radial orbits, simply due to the geometric effect of the orbit collapsing into a line for the case of a perfectly radial orbit. The second column (HYDRO006-032) shows a tidal stream that has been largely disrupted but still produces a strong main peak with a second, less prominent lobe. The third column (C2-200) shows a more complex morphology where there is u-shaped maximum. Here, two effects are in play: orbital precession causes the u-shape and the low angular momentum of the orbit causes its stretching, as explained before. In the fourth column example (A2-096), the spatial morphology is more shell-like and the corresponding signature in the PCM more intricate, although there is still a recognizable maximum. Finally, the last column (B2-022) shows a completely phase-mixed event that produces no discernible or significant maximum. These events will end up contributing to the PCM background.

6 FULL HALO PCMS: A FIDUCIAL EXAMPLE

In this section, we will show in detail how the morphology of the PCMs is affected by the *Gaia* selection function and observational errors (Section 6.1), how the peaks are detected (Section 6.2) as well as the effect of the choice of tracer (Section 6.3). For clarity, we consider a single fiducial halo in this section as an illustrative example of these different aspects of the method. In Section 7, we will describe the results obtained for all the Aquarius and HYDRO-zoom haloes.

6.1 Selection function and errors

The nGC3 PCM for all *Gaia* observable stars ($G \leq 20$) in the Aquarius A2 halo is shown in the left-hand panel of Fig. 4. Some peaks have been labelled in this plot to facilitate discussion. The middle panel shows the effect of adding the *Gaia* selection function and observational errors. The right-hand panel shows the PCM after a cut in the proper motion errors has been imposed ($\Delta\mu/\mu \leq 0.5$). The colour scale is proportional to the logarithm of the star counts in each cell. The tolerance used to compute all nGC3 PCMs hereafter is $\Delta\theta = 1:5$.

Several well-defined, localized peaks are very noticeable, as well as some other more extended features corresponding to streams that have undergone more significant phase mixing. The two examples from the Aquarius A2 halo shown in Fig. 3 (first and fourth columns) are easily recognizable here as peaks d and f.

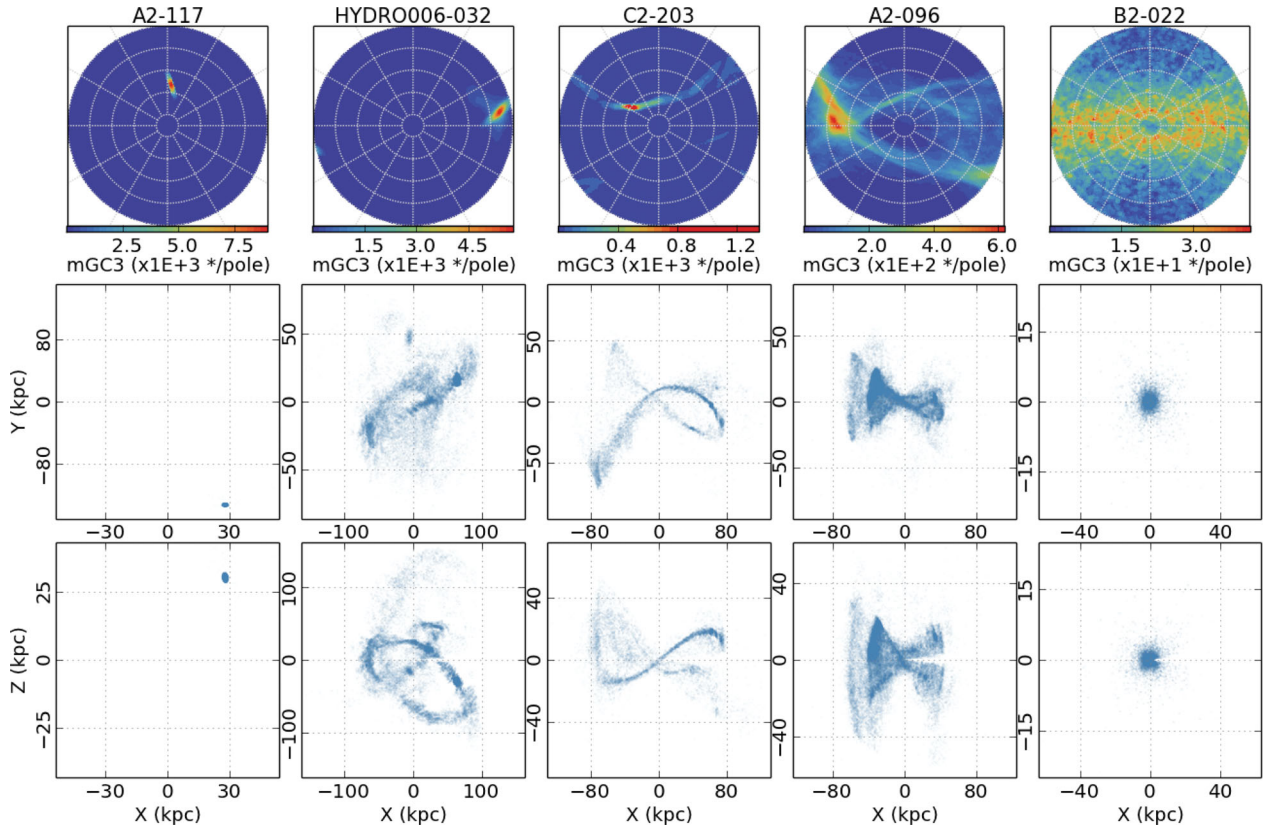


Figure 3. PCMs and spatial distributions for different progenitors in the Aquarius and HYDRO-zoom K giant mock catalogues. The five progenitors shown go from completely bound to completely disrupted from left to right. The top, middle and bottom rows show, respectively, the nGC3 PCMs, with a colour scale proportional to the number of stars per pole, the face-on Y versus X projection and the edge-on Z versus X projection, where the colour shade is proportional to density. Only *Gaia* observable stars are shown in these plots. The Sun is located at $X = -8.5$ kpc and $Z = 0$ is the plane of the disc (in the Aquarius simulations the X and Y axes correspond to the major and intermediate axes of the potential). PCMs are shown in a north-polar azimuthal projection.

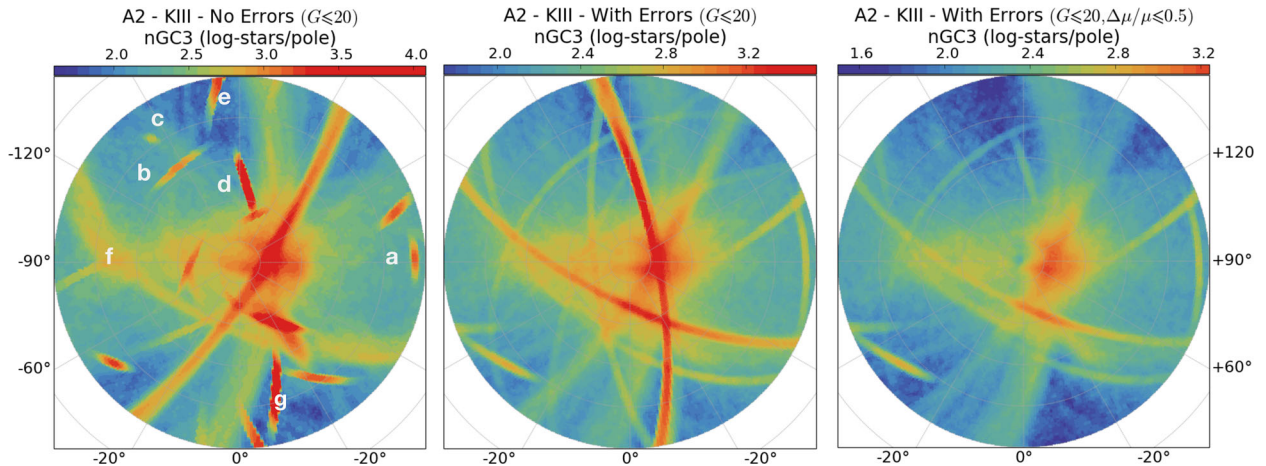


Figure 4. nGC3 PCMs for *Gaia* observable stars in the Aquarius A2 halo. Left: without errors ($G \leq 20$). Middle: with errors ($G \leq 20$). Right: with errors, after proper motion error cut ($G \leq 20$, $\Delta\mu/\mu \leq 0.5$). The three panels have the same colour scale limits. Some peaks have been labelled in the left-hand panel for reference (see the discussion in the text). PCMs are shown in a north-polar azimuthal projection. Meridians and parallels (light grey) are drawn at uniform 30° and 20° intervals, and their corresponding labels are shown where they intersect the vertical and horizontal axes, respectively.

When the observational errors are added, as illustrated in the middle panel of this figure, some peaks that were well defined are now stretched along great circles to different degrees (e.g. a, b) and some appear a bit more fuzzy (e.g. c). The stretching is mostly due to the degradation of the kinematic information, particularly in the cases where the progenitor is either mostly bound (as peak d,

corresponding to A2-117 in Fig. 3) or when few stars in the tidal stream can be detected. The use of bad proper motion data causes a given peak to be stretched out along the great circle defined by the poles of all possible planes that go through the clump and the Galactic Centre. Another way to look at this is to think of nGC3 PCMs as tending towards their GC3 counterparts as the precision

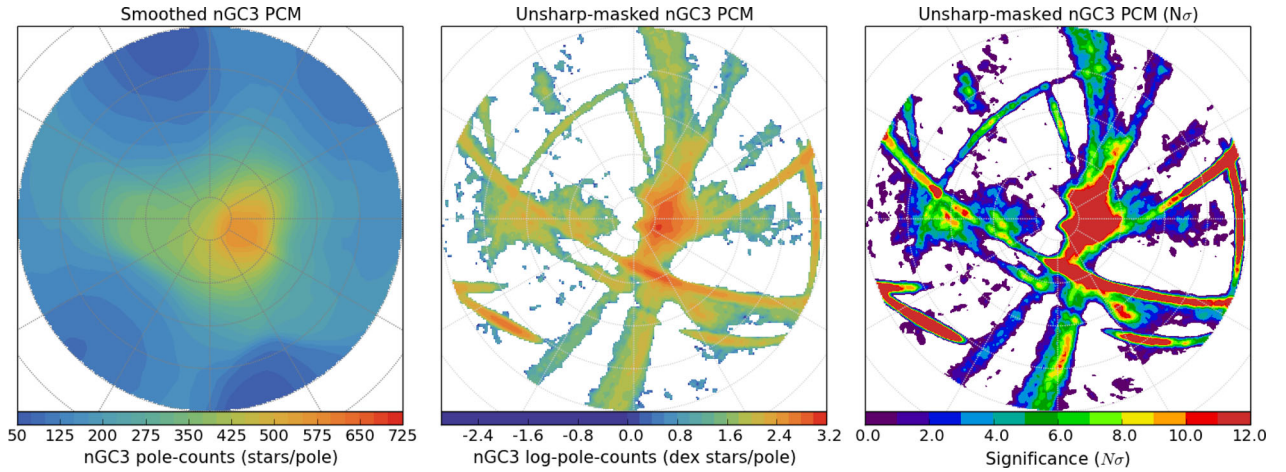


Figure 5. The process of unsharp masking illustrated for the nGC3 PCM of Aquarius halo A2, with the proper motion error cuts (right-hand panel of Fig. 4). Left: smoothed (median-filtered) PCM. Middle: unsharp-masked PCM computed by subtracting the smoothed map from the original one. Right: unsharp-masked PCM in $N\sigma$ units, computed dividing the subtracted PCM by the square root of the smoothed PCM.

of the kinematic data worsens. An effect due to distance errors is also present, but here it is minimized by our choice of tracers with reasonably small photometric distance errors (Section 3.2.1). This severe stretching of the peaks into great circles is problematic because it will increase the contamination in the features detected, and will also make the detection of spurious peaks more likely at the intersection of great circles. To mitigate these effects, we keep only those stars with proper motion errors less than 50 per cent. The resulting PCM is shown in the right-hand panel of Fig. 4. Some features are inevitably lost because some progenitors do not give rise to streams that have stars sufficiently close or bright enough to have *Gaia* proper motions with errors smaller than our 50 per cent cut. That is the case of peak e in the left-hand panel of Fig. 4, which has completely disappeared in the right-hand panel. On the other hand, some other features like d and g would merge into one great circle if the cut were not imposed, so we believe this relatively relaxed cut offers a good compromise.

6.2 Detecting peaks in PCMs

We begin by first removing the contribution of the smooth background by *unsharp masking*, as in M11. A smoothed map is produced by applying a median filter to the PCM, assigning to each pixel the median counts computed in a neighbourhood of fixed size, selected to be $\sim 20^\circ$ – 22° (~ 15 times the great-circle tolerance $\delta\theta$), i.e. much larger than the typical size of the peaks one is interested in finding. The left-hand panel in Fig. 5 shows the smoothed PCM for the Aquarius A2 halo example in Fig. 4 (right-hand panel). This smoothed map reflects the contribution of the well-mixed halo background stars to the PCM, with the effect of the selection function folded in. The middle panel of Fig. 5 shows the unsharp-masked PCM, obtained by subtracting the smoothed map in the left-hand panel. In this unsharp-masked PCM, the peaks are clearly highlighted. The colour scale in this panel is proportional to the log-counts. To give a sense of the significance of the peak height with respect to the background, the right-hand panel in Fig. 5 shows the unsharp-masked PCM now in $N\sigma$ units. This is computed dividing pixel by pixel the unsharp-masked PCM (middle panel) by the square root of the smoothed PCM (left-hand panel), which assumes that the pole counts follow a Poisson distribution.

Peaks are detected in the unsharp-masked PCM using the FellWalker⁷ algorithm from Berry (2015). As explained in detail in this reference, FellWalker uses a watershed algorithm that divides the pixels in an image into disjoint clumps, each of which contains one local maximum. This is done only for those pixels above some noise threshold, so background pixels below it are not assigned to any clump. This is a very efficient and general algorithm that allows detecting peaks without any particular shape, a crucial point since stream signatures in PCMs can significantly differ from simple Gaussian peaks as illustrated in Fig. 3.

Ideally, one would want the peak detection algorithm to exploit the fact that peak signatures in PCMs tend to stretch along great circles (see Fig. 3), particularly since very elongated features are frequently fragmented into multiple peaks by the detection algorithm. However, implementing this is out of the scope of the present paper, so we defer it for a future work. To reduce this excessive fragmentation, we simply apply the position-only GC3 method (see equation 1 and Section 5.1), but this time on the ϕ – θ coordinates of the FellWalker peak detections in the PCM themselves. This way we can merge peaks for which the majority of pixels lie on great circles within a tolerance of $\sim 1^\circ$. This last step is done in an interactive way to ensure that only the detections that lie along the most obvious great circles are merged into a single detection. The end result of the peak detection is shown in the top panel of Fig. 6 for the Aquarius A2 halo PCM, where each of the identified peaks is marked with a labelled circle.

6.3 The choice of tracers

The choice of tracer will have different effects on the PCM. To illustrate this, the top and bottom panels of Fig. 6 show respectively the nGC3 PCM for the Aquarius A2 halo K giants and RRLs, observable with *Gaia* with errors and after the cut in relative proper motion error. The numbers of both tracer stars per halo, as observable by *Gaia*, with and without the proper motion error cut, are summarized in Table 1.

RRLs are less numerous than K giants by factors ranging from ~ 4 to 10 (see Table 1). Consequently, the RRL PCMs appear noisier

⁷ The FellWalker algorithm is part of the Starlink Software Distribution.

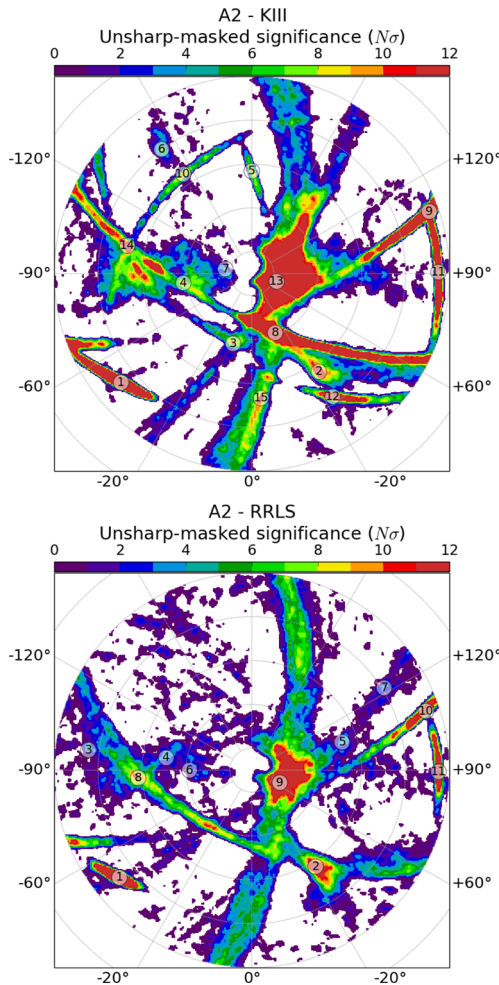


Figure 6. Peak detections in the unsharp-masked nGC3 PCM for the Aquarius A2 halo. Top: K giants. Bottom: RRLSs. Individual peaks are labelled. The colour scale corresponds to the pixel’s significance in $N\sigma$ units.

Table 1. Number of K giants and RRLS observable by *Gaia* and with relative proper motion errors smaller than 50 percent, for the Aquarius haloes.

Halo	KIII		RRLS	
	$G \leq 20$	$\Delta\mu/\mu \leq 0.5$	$G \leq 20$	$\Delta\mu/\mu \leq 0.5$
A2	458 626	283 419	98 870	64 126
B2	771 795	641 795	140 511	114 983
C2	511 452	282 512	75 506	45 176
D2	1815 093	1103 009	264 739	190 288
E2	1586 759	1338 395	155 834	132 549

than their K giant counterparts due to Poisson noise. RRLSs are also fainter than K giants (the latter are giants brighter than the HB), so several peaks for the most distant structures that are observable with K giants are absent in the RRLS PCM; e.g. peaks 5, 6, and 12 in the top panel are not present in the bottom one. On the other hand, typical RRLS distance errors are smaller (see Section 3.1.2) producing sharper features in the PCMs for the progenitors that do contain observable stars, e.g. compare peaks 1 and 11 in the top panel to peaks 1 and 11 in the bottom panel, respectively. Producing more concentrated features could also translate into some peaks being detectable with RRLSs and not with KIII stars, either because they are easier to resolve or because the contrast between the peak

signal and the background noise is larger, as is the case for peak 5 that is just detectable in the bottom panel and not in the top one.

The relative importance of these competing effects cannot be gauged a priori, as it will depend on the particular accretion history of each halo, as we will show in Section 7. Simply, each tracer offers its own advantages: K giants can probe a larger halo volume, whereas RRLSs can provide a more detailed view within the inner halo.

7 FULL HALO PCMS: ALL HALOES

So far we have used one halo, Aquarius A2, as a fiducial example to show in detail what the PCM of a cosmological halo looks like, the effects of the *Gaia* selection function and errors, and the procedures we use to detect maxima in the PCMs. In what follows, we will discuss the PCMs of all Aquarius (Section 7.1) and HYDRO-zoom haloes (Section 7.2).

7.1 Aquarius PCMs

Fig. 7 shows the nGC3 PCMs for the four remaining Aquarius haloes B2, C2, D2 and E2 from left to right. The top and bottom rows correspond to K giant stars and RRLSs, respectively. The halo-to-halo variation in the degree of substructure is evident, ranging from halo E2 with very little substructure to halo D2 with the most.

The effect of the choice of tracer is also illustrated in Fig. 7. In all haloes, but more noticeably in haloes B2 and E2, the predominant effect is that more peaks are detected with K giants than with RRLSs. The notable differences, in particular for E2, come from very strong great-circle peaks present in the K giant PCMs that are absent in the RRLS PCMs. These correspond to fully bound and distant progenitors that lie beyond the reach of *Gaia*’s RRLS. By contrast, in haloes C2 and D2, there are some examples of peaks detected with RRLSs and not K giants, e.g. peaks 12 in halo C2 and 19 in halo D2. In this case, it is the better precision of RRLS distances that makes these peaks detectable with RRLSs and not K giants. These examples show that each tracer has its own advantages and disadvantages, as we had discussed in Section 6.3.

Since nGC3 and all great-circle methods are linear, it is possible to combine the data by simply adding the PCMs for different tracers. It would be advisable to do this after unsharp masking, i.e. to combine the PCMs *after the smooth background has been subtracted*. In the cases like those presented here, where one tracer is much more numerous than the other, this would prevent the dilution of peaks detected only with the more scarce tracer by the overall background of the more numerous one. In what follows, however, we consider the data for different tracers separately in order to analyse which progenitors are recovered with each tracer.

7.2 HYDRO-zoom PCMs

In the following, we will focus on the *Gaia* mock distribution of K giant stars in the HYDRO-zoom haloes. Table 2 summarizes, for each halo, the number of K giants observable within the *Gaia* magnitude limit and with an additional cut $\Delta\mu/\mu \leq 0.5$, respectively, shown separately for the accreted (left) and *in situ* components (right). This shows that the number of *in situ* K giant stars is quite large, more than double in most cases than the number of K giant accreted stars. This is to be expected since not only halo, but also the disc stars are included.

Since the disc is the most important source of contamination, we introduce a cut to eliminate all low-latitude ($|b| \leq 10^\circ$) stars inside a

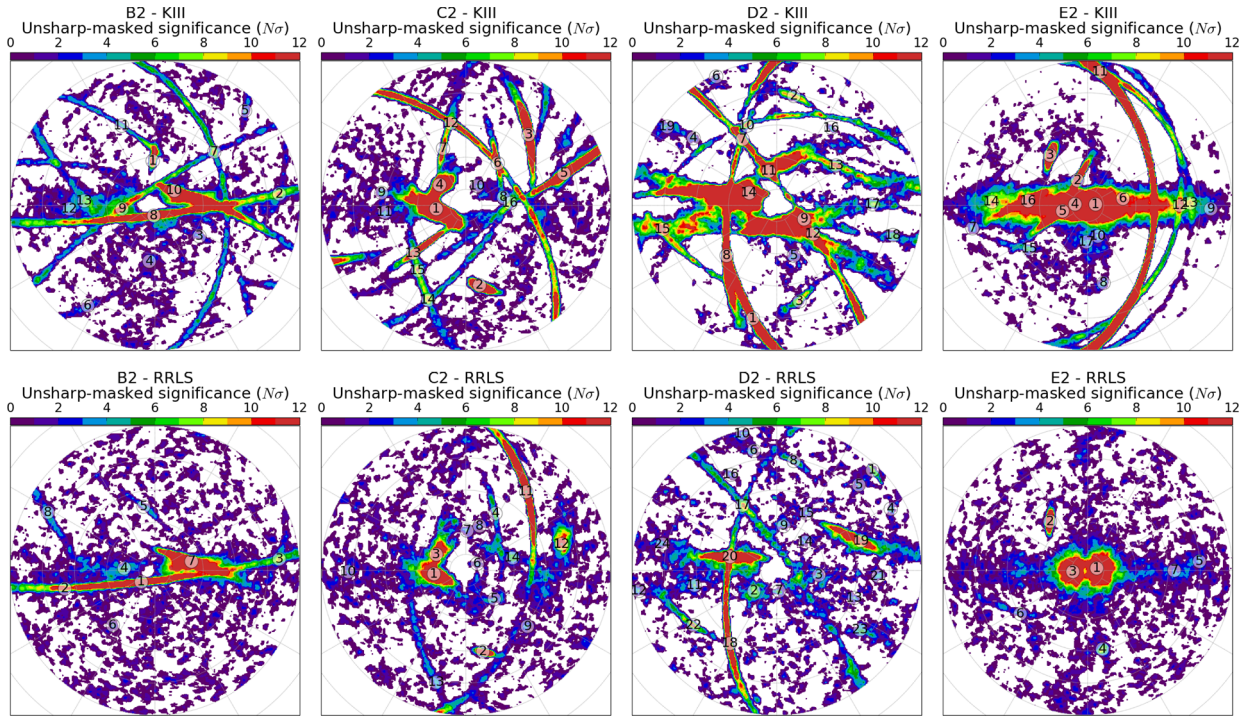


Figure 7. Unsharp-masked nGC3 PCM for Aquarius haloes B, C, D and E from left to right, for *Gaia* observable K giants (top) and RRLSs (bottom), with errors, after proper motion error cut ($G \leq 20$, $\Delta\mu/\mu \leq 0.5$). The colour scale corresponds to the pixel’s significance in $N\sigma$ units. Labelled circles indicate the peaks detected using the procedure described in Section 6.2.

Table 2. Number of K giants observable by *Gaia* and with relative proper motion errors ϵ_μ smaller than 50 per cent in total and excluding stars with $|b| \leq 10^\circ$ and $R \leq 20$ kpc, for the HYDRO-zoom haloes.

Halo	KIII (accreted)		KIII (<i>in situ</i>)		$\Delta\mu/\mu \leq 0.5$ + excl. zone
	$G \leq 20$	$\Delta\mu/\mu \leq 0.5$	$G \leq 20$	$\Delta\mu/\mu \leq 0.5$	
001	2048 136	1789 181	9953 712	9496 893	5540 500
004	2476 549	2053 819	7595 529	7059 015	3642 743
006	803 328	602 513	8577 449	7825 379	3101 889
008	1170 131	549 300	7438 652	7241 107	4418 790
009	1937 388	1806 513	6123 721	5750 542	3078 387

given Galactocentric (cylindrical) radius ($R \leq 20$ kpc). This way we avoid eliminating distant stars that may belong to streams beyond the disc radius. The cut is introduced in Galactocentric radius, taking advantage of the fact that our assumed tracer provides reasonably precise distances (errors < 20 per cent); if this were not the case, it would be preferable to define the cuts using a direct observable (see e.g. 88M11). Fig. 8 shows the nGC3 PCM for halo 006 using all *in situ* stars (left) and including the disc avoidance zone (right), which clearly illustrates how the overall pole counts are reduced and several features are revealed in the PCM using this simple cut.

Fig. 9 shows the K giant nGC3 unsharp-masked PCMs for all gas dynamical HYDRO-zoom haloes (001 to 009), with the *in situ* background and excluding stars with $|b| \leq 10^\circ$ and $R \leq 20$ kpc. The peak detections were made following the procedure described in Section 6.2, and are shown with labelled circles. As seen here, there seems to be less substructure overall in the HYDRO-zoom PCMs than in the Aquarius haloes. Even so, there is a large range in the amount of substructure present in the different HYDRO-zoom haloes. Haloes 008 and 009 exhibit little substructure and a few very luminous progenitors producing strong great-circle maxima in

each case, whereas halo 006 shows a level of substructure similar to some of the Aquarius haloes.

To check whether the lower mass resolution translates into fewer progenitors overall, Table 3 lists the number of progenitors with masses higher than $10^6 M_\odot$ (the mass limit of the HYDRO-zooms). The table shows that, on average, there are *more* progenitors per halo in the HYDRO-zooms compared to Aquarius, so this is not the reason why less substructure is visible. In addition, the stellar mass of the accreted component in the HYDRO-zoom haloes is, in general, higher than that of the Aquarius haloes, with three haloes (006,008,009) being as massive as the most massive Aquarius halo D2, and with haloes 001 and 004 being close to two and four times as massive. This causes the number of K giants in HYDRO-zoom haloes to be generally higher than in the Aquarius haloes, as Tables 1 and 2 show, which translates into much higher backgrounds in the PCMs (this will be clearly illustrated in the next section by N_{BG} in Tables 4 and 5). Note that here we are comparing the number of stars in the *accreted* halo, so this effect is present even without taking the *in situ* component into account.

Up to this point, it is not possible to tell whether this will prevent us from using the HYDRO-zoom simulations for our predictions on the number of detectable streams. We will address this in the next section, as we analyse the progenitors recovered in the two types of simulations.

8 RECOVERING STREAMS IN COSMOLOGICAL SIMULATIONS

8.1 Which progenitors can we recover?

First, we need to decide when a progenitor is considered ‘recovered’, as well as one or more quantities that will help us define the quality of the recovery. Two useful definitions are the *fraction of recovered*

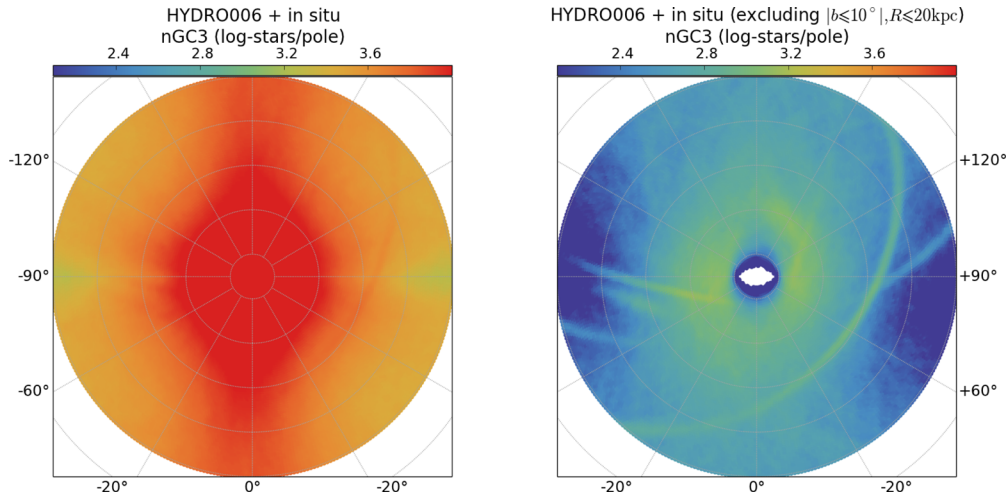


Figure 8. PCMs for HYDRO-006 illustrating the effect of including the *in situ* component. Left: full *in situ* background. Right: excluding stars with $|b| \leq 10^\circ$ and $R_{\text{gal}} \leq 20$ kpc. The colour scale is the same for both plots and shows nGC3 pole counts. Note how the PCM pole counts are reduced by about an order of magnitude after including the disc avoidance zone.

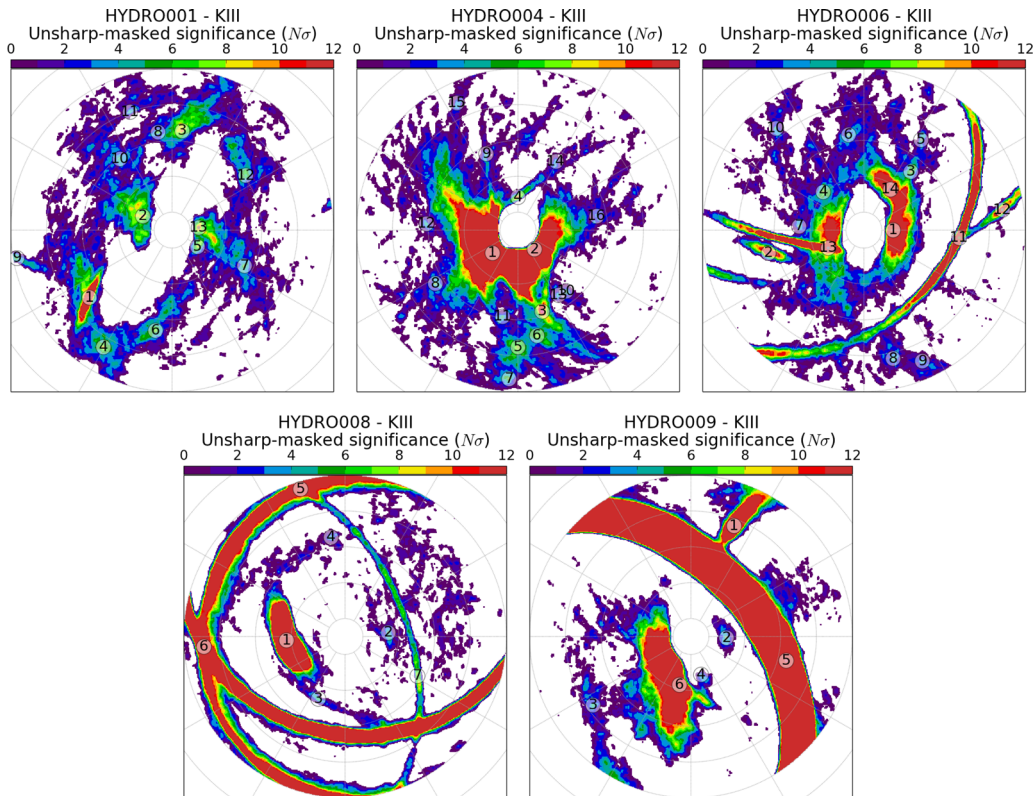


Figure 9. Unsharp-masked nGC3 PCM for HYDRO-zoom haloes 001 to 009, for *Gaia* observable K giants with errors, after proper motion error cut ($G \leq 20$, $\Delta\mu/\mu \leq 0.5$). The colour scale corresponds to the pixel's significance in $N\sigma$ units. Labelled circles indicate the peaks detected using the procedure described in Section 6.2.

$\text{stars } f_{\text{rec}}$, i.e. the fraction of progenitor stars in a given pole detection; and the *purity*, defined as the number of progenitor stars within the detected peak, divided by the total number of stars within that detection. With these definitions $f_{\text{rec}} = 0.4$ means 40 per cent of the total (observable) progenitor stars are recovered in the pole detection and $\text{purity} = 1$ means there are no contaminant stars from other progenitors (or the *in situ* halo).

Since different progenitors can produce maxima that overlap in the PCM, any given pole detection can be associated with more

than one progenitor and vice versa. We allow for multiple progenitors to be associated with any given pole, and hence considered as detected, provided a minimum fraction $f_{\text{rec}} > 0.1$ of progenitor stars is recovered. For a given pole, we will consider the progenitor recovered with the highest purity as the primary detection, and the remaining as secondary detections, so there will be as many primary detections as poles detected in the PCM. Although secondary detections are indistinguishable from primary ones with nGC3, we will consider them as valid detections as we expect that follow-up

Table 3. Stellar mass and number of all progenitors N_{all} and of streams N_{streams} with stellar masses $>10^6 M_{\odot}$.

Halo	M_{stellar} (M_{\odot})	N_{all} ($>10^6 M_{\odot}$)	N_{streams} ($>10^6 M_{\odot}$)
A2	7.2e+08	41	19
B2	6e+08	23	15
C2	1.3e+09	36	20
D2	2.3e+09	39	21
E2	1.2e+09	28	14
001	8.1e+09	52	32
004	4e+09	33	21
006	2.5e+09	37	34
008	2e+09	24	17
009	2.6e+09	32	28

methods will be able to disentangle them, e.g. through radial velocities or colour–magnitude diagram analyses.

The distribution of recovered and unrecovered progenitors in the Aquarius and HYDRO-zoom haloes is illustrated in Fig. 10. The main plots (central and bottom panels) show the infall redshift z_{infall} versus N_{KIII} , the number of *Gaia* observable K giants with proper motion errors $\Delta\mu/\mu \leq 0.5$. We define the infall redshift as the time of the first simulation output at which a progenitor is identified as a subhalo of the main halo. Different symbols represent different haloes, as shown in the plot legend. Recovered progenitors are represented as filled coloured symbols, large and small representing, respectively, primary and secondary detections and with a colour proportional to the purity. Unrecovered progenitors are shown with filled grey symbols. The labelled points indicate examples of recovered (a–f) and unrecovered (g) progenitors, for which the corresponding spatial distribution (X – Z plot) and nGC3 PCMs (with errors) are shown in the top row of the figure. Labelled points indicate progenitors selected to illustrate examples of detections, primary and secondary, and non-detections.

Table 4. Statistics of recovered progenitors in the Aquarius haloes (primary + secondary detections combined). The columns are the median PCM background counts N_{BG} ; the overall fraction of all progenitors and streams recovered f_{all} and f_{str} , respectively; the number of progenitors inside the detection boundary ($\Delta\theta \leq 15^\circ$) in total and recovered, respectively, for bound progenitors $N_{\text{T}}^{\text{bnd}}$, $N_{\text{rec}}^{\text{bnd}}$ and for streams $N_{\text{T}}^{\text{str}}$, $N_{\text{rec}}^{\text{str}}$; and the numbers of recovered streams detected in common with RRLSs and K giants $N_{\text{RR+K}}$, only RRLS N_{KnotRR} , only K giants N_{KnotRR} and combined $N_{\text{RR+K}}$.

Halo	N_{BG}	f_{all}	f_{str}	RRLS				N_{BG}	f_{all}	f_{str}	KIII				RRLS+KIII ($N_{\text{rec}}^{\text{str}}$)			
				$N_{\text{T}}^{\text{bnd}}$	$N_{\text{rec}}^{\text{bnd}}$	$N_{\text{T}}^{\text{str}}$	$N_{\text{rec}}^{\text{str}}$				$N_{\text{T}}^{\text{bnd}}$	$N_{\text{rec}}^{\text{bnd}}$	$N_{\text{T}}^{\text{str}}$	$N_{\text{rec}}^{\text{str}}$	$N_{\text{RR+K}}$	N_{RRnotK}	N_{KnotRR}	$N_{\text{RR+K}}$
A2	45	0.71	0.67	1	1	6	4	168	0.86	0.75	6	6	8	6	1	3	5	9
B2	55	1.00	1.00	0	0	7	7	251	0.91	1.00	3	2	8	8	7	0	1	8
C2	22	0.83	0.83	0	0	12	10	114	0.67	0.67	6	4	12	8	5	5	3	13
D2	111	0.78	0.88	1	0	8	7	624	0.85	0.88	5	4	8	7	5	2	2	9
E2	56	0.80	1.00	2	1	3	3	537	1.00	1.00	6	6	3	3	2	1	1	4

Table 5. Statistics of recovered progenitors in the HYDRO-zoom haloes, with K giant stars (primary + secondary detections combined). The columns are the median PCM background counts N_{BG} ; the overall fraction of all progenitors and streams recovered f_{all} and f_{str} , respectively; the number of progenitors inside the detection boundary ($\Delta\theta \leq 15^\circ$) in total and recovered, respectively, for bound progenitors $N_{\text{T}}^{\text{bnd}}$, $N_{\text{rec}}^{\text{bnd}}$ and for streams $N_{\text{T}}^{\text{str}}$, $N_{\text{rec}}^{\text{str}}$.

Halo	N_{BG}	f_{all}	f_{str}	No <i>in situ</i> component				N_{BG}	f_{all}	f_{str}	With <i>in situ</i> excl. $ b \leq 10^\circ$ and $R \leq 20$ kpc				$N_{\text{T}}^{\text{str}}$	$N_{\text{rec}}^{\text{str}}$
				$N_{\text{T}}^{\text{bnd}}$	$N_{\text{rec}}^{\text{bnd}}$	$N_{\text{T}}^{\text{str}}$	$N_{\text{rec}}^{\text{str}}$				$N_{\text{T}}^{\text{bnd}}$	$N_{\text{rec}}^{\text{bnd}}$	$N_{\text{T}}^{\text{str}}$	$N_{\text{rec}}^{\text{str}}$		
001	1441	0.67	1.00	1	0	2	2	703	1.00	1.00	1	1	2	2	2	2
004	1673	1.00	1.00	0	0	1	1	815	0.67	0.50	1	1	2	1	1	1
006	445	1.00	1.00	1	1	7	7	446	1.00	1.00	1	1	7	7	7	7
008	138	0.70	1.00	4	1	6	6	223	0.60	0.67	2	3	6	4	4	4
009	626	0.50	0.50	0	0	4	2	540	0.50	0.75	0	0	4	3	3	3

Fig. 10 shows that recovered streams exhibit a wide variety of morphologies in their spatial distributions as well as in their signatures in PCMs.

Panels a to c show three progenitors recovered with high purity. Progenitor a has produced a bright and well-defined stream, it has a very prominent bound core and tidal tails with several wraps that have undergone some precession, which has spawned a second lobe (light blue) in the PCM. Progenitor c is recovered with a similar purity as a. Even though it was accreted as early as $z_{\text{infall}} \sim 4$, its tidal stream is quite cold and produces a very well defined peak in the PCM. Progenitor b was accreted even earlier than c and has produced a much more disrupted stream with a more complicated signature in the PCM, but that is still recovered with purity >0.3 .

Panels d, e, f show some intermediate cases that illustrate the effect of contamination and overlapping signatures in the PCM. Progenitor d has roughly as many visible stars as c and was accreted slightly earlier ($z_{\text{infall}} \sim 3$). The low purity (<0.1) of this detection is caused by its main peak overlapping in the PCM with the signature of the much brighter progenitor b. This can also be clearly seen in Fig. 6 (top), where pole detection 4, which corresponds to progenitor d, is located in a PCM region with a higher-than-average background (see also Fig. 5, left-hand panel). Progenitor e is an example of a secondary detection. The signature it produces in the PCM is readily evident in Fig. 7 (top row, second panel) and is detected as pole 14. This pole detection, however, is associated with the much brighter progenitor f; a completely disrupted progenitor that produces a PCM signature, which although diffuse, dominates pole counts around progenitor e’s peak. Hence, progenitor f also serves as a false positive example, as it is a spurious detection of a completely disrupted stream that we should not expect to recover with our method.

Panel g shows a progenitor that is not recovered. It was accreted at a relatively high redshift $z_{\text{infall}} > 3$ and is an example of a stream that has been completely phase mixed, which produces no clear signature in the PCM and hence, as expected, is not detectable.

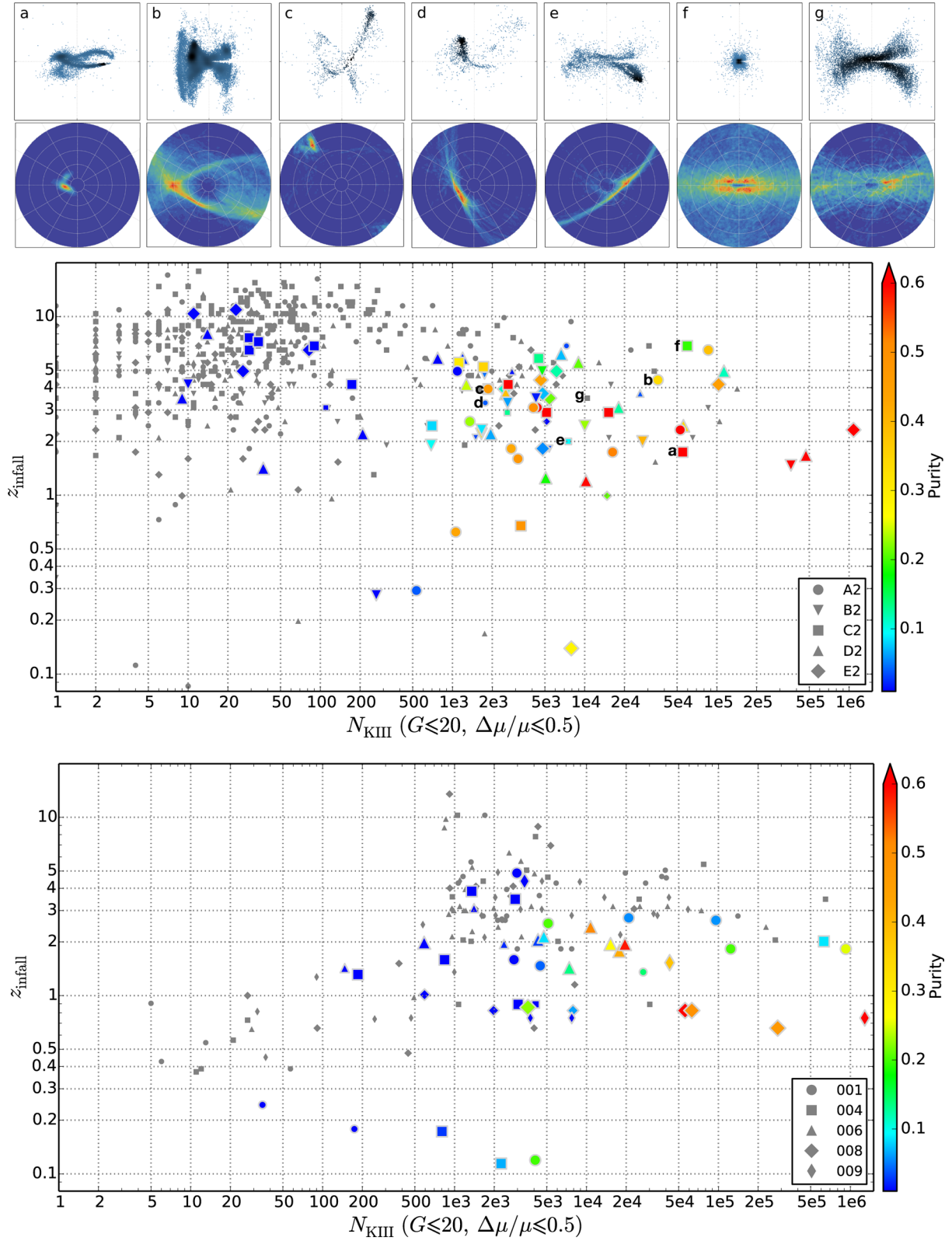


Figure 10. Middle and top panels: infall redshift z_{infall} versus number of observable K giants N_{KIII} for each of the progenitors in the Aquarius haloes. Recovered progenitors are shown with filled symbols, and primary and secondary detections are denoted by large and small symbols. The colour scale is proportional to the purity of the corresponding pole detection, with the upper end of the scale meaning a purity of 0.6 or higher. In the smaller sub-panels at the top, we show the spatial Z versus X distributions and the nGC3 PCMs, respectively. The sub-panels correspond to a few selected progenitors, labelled from a to g in the Aquarius A2 (b, c, d) and C2 (a, e, f, g) haloes, respectively. Darker colours in the spatial distribution plots correspond to higher density. Bottom panel: the infall redshift z_{infall} versus number of observable K giants N_{KIII} for each of the progenitors in the HYDRO-zoom haloes. Again, the colour scale is proportional to the purity of the corresponding pole detection.

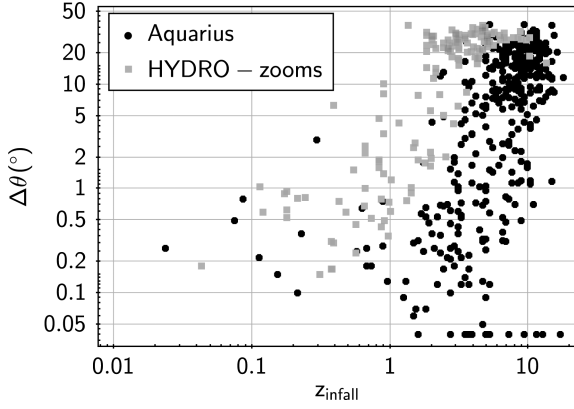


Figure 11. Angular thickness $\Delta\theta$ versus infall redshift z_{infall} for all progenitors in the Aquarius (black circles) and HYDRO-zoom (grey squares) simulations.

The central panel of Fig. 10 shows that, in the Aquarius simulations, streams can be recovered up to infall redshifts as high as ~ 5 – 6 and with relatively good purity (>0.3) for progenitors with more than about a thousand observable stars. The lower panel shows the distribution of the recovered progenitors in the gas dynamical HYDRO-zoom haloes. The HYDRO-zoom simulations have a lower resolution than Aquarius, so in Fig. 10 the plots are more sparsely populated with only the most massive (and hence the most luminous) progenitors. This explains why there are only a few points corresponding to small numbers of observable stars, with a clear deficiency of objects below $N_{\text{KIII}} \sim 500$. This lack of objects is very evident, particularly at redshifts higher than ~ 2 where there are no progenitors with <700 stars. This shows that close to $N_{\text{KIII}} \sim 500$ – 1000 , our results might be hampered by the lower resolution of the HYDRO-zoom simulations.

Overall, infall redshifts are lower for gas dynamical progenitors compared to Aquarius progenitors and, in particular, recovered progenitors are detected up to infall redshifts ~ 2 – 3 , i.e. accreted more recently than those recovered in the Aquarius simulations ~ 5 – 6 . However, it is not possible to tell from this plot whether the recovery of progenitors at lower redshifts in the gas dynamical simulations is simply due to the fact that, overall, there are fewer progenitors accreted at larger redshifts, or whether at a particular redshift, progenitors are more easily disrupted due to the presence of a disc making its detection harder. Hence, we should look at the angular thickness of a stream, which *will* have a direct effect on its detectability with our method and serves as a proxy for the dynamical age of a stream.

To estimate the angular thickness $\Delta\theta$, we rotate each stream so that its mid-plane coincides with the Galactic equator and look at the distribution of stars in latitude. This can give us a sense of how thick the stream is in the direction perpendicular to its orbital plane. We fit a four-component Gaussian mixture model to this distribution and compute $\Delta\theta$ as the sum in quadrature of the standard deviations of the two main Gaussian components, weighted by their amplitudes. We find that this gives a good representation of the angular thickness of the streams, as it balances the contribution of outliers, the actual tidal tails and the bound core (where one exists).

Fig. 11 plots the angular thickness $\Delta\theta$ as a function of infall redshift z_{infall} , for all progenitors in the Aquarius and HYDRO-zoom simulations. This plot clearly shows that, at a given infall redshift, progenitors in the HYDRO-zoom gas dynamical simulations (grey) produce thicker streams than in the dark-matter-only Aquarius

simulations (black). Hence, progenitors are more effectively disrupted, as we anticipated might be the case due to the effect of the disc, which explains why streams are detected up to lower redshifts in the HYDRO-zooms. Also, for both simulations, the correlation of angular width and infall redshift is clear, albeit with large scatter, which confirms that the angular width is a suitable proxy for the infall redshift, and hence for dynamical age.

In the $z_{\text{infall}}-N_{\text{KIII}}$ plots of Fig. 10, additionally, detections and non-detections are not segregated, as is to be expected, since progenitors accreted at the same z_{infall} but on different orbits will be disrupted to different degrees (for example, d and g). In order to look for a clear boundary that separates detections from non-detections, we will examine a plane of observables that have a direct effect on detectability with our method.

8.2 The detection boundary

The number of observable stars and the angular thickness of a tidal stream are two parameters that directly influence the detectability of a stream with nGC3 or any great-circle method in general. Obviously, the method is more efficient when more tracer stars are available and in the cases when streams are dynamically colder.

Recovered and not recovered progenitors from the Aquarius and HYDRO-zoom haloes shown in Fig. 10 are now shown, respectively, in the top and bottom panels of Fig. 12 in the plane of angular thickness versus number of observable stars. The colour coding, symbols and labels are the same as in Fig. 10. In this plot, a clear segregation is evident as detected progenitors (coloured symbols) are fairly well separated from non-detections (grey symbols) and, in general, low-purity detections tend to be those with fewer observable stars and larger angular thickness.

In this plane, we can estimate a priori where a detection boundary should lie based on how the great-circle methods work. Structures that are thinner than the assumed tolerance and having more than some minimum number of stars above the background should be recovered, since all stars would fit inside a single great-circle cell. For wider structures to be detected, increasingly larger numbers of stars are needed to compensate for the fact that stars are dispersed into more than one great-circle cell, up to a certain angular width for which the method saturates. Thus, we propose that the detection boundary can be expressed as

$$\frac{\Delta\theta}{\delta\theta} = \frac{N_{\text{KIII}} - N_{\text{BG}}}{N_{\text{BG}}} = \frac{N_{\text{KIII}}}{N_{\text{BG}}} - 1. \quad (4)$$

In equation (4), $\delta\theta$ is the tolerance used in producing the nGC3 PCMs (see equations 1 and 2) and N_{BG} is the number of stars in the background, which can be estimated from the PCM itself using the smoothed map computed during the unsharp masking (e.g. Fig. 5, left, for Aquarius A2). Therefore, the location of this boundary can be predicted without any free parameters, based on observables alone. The solid black lines in Fig. 12 represent the detection boundary given by equation (4), taking N_{BG} to be the median of the counts in the smoothed PCM for each halo. The shaded regions were computed for each halo using the 25th and 75th percentiles of the respective smoothed PCM counts, so as to represent the uncertainty of this boundary due to the non-uniformity of the PCM smooth background.

The eight progenitors labelled in Fig. 10 are also shown in Fig. 12. This clearly shows that all successfully recovered progenitors (a, b, c, d, e) are well inside the detection boundary. Note that progenitor d is a secondary detection, highly contaminated by the more luminous progenitor b, and is located in the PCM (Fig. 6) in a region

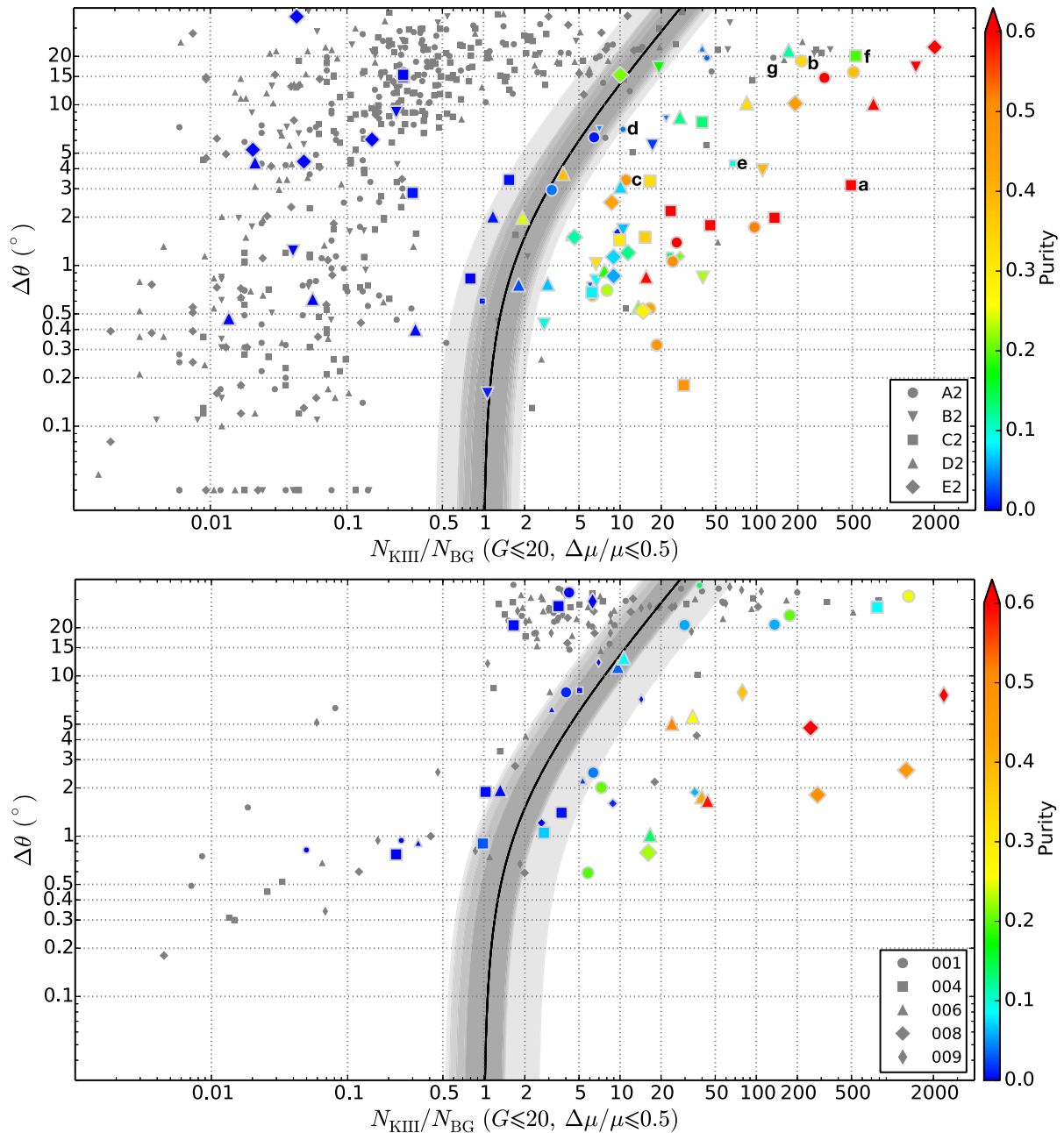


Figure 12. Angular thickness $\Delta\theta$ versus number of observable K giants N_{KIII} for each of the progenitors of the Aquarius haloes. Large and small symbols denote progenitors recovered as primary and secondary detections, and grey symbols indicate progenitors that are not recovered. The colour scale is proportional to purity, with the upper limit indicating a purity = 0.6 or higher. The labelled points correspond to the same progenitors as in Fig. 10. The black solid line and shaded regions in the right-hand panel indicate, respectively, the median detection boundary and its edges computed from equation (4) using the 25th and 75th percentiles of the background counts.

with a relatively higher background. In Fig. 12, it lies close to the grey bands marking the 75th percentile of the background counts, showing that it is almost a border-line detection. Progenitors f and g, classified as a false positive and a non-detection, respectively, are clearly seen here to be bright but very diffuse features with angular widths above 20° .

The two panels in Fig. 12 show the detection boundary effectively separates detections from non-detections in both simulations, with most non-detections (grey points) and low-purity detections (blue points) lying to the left of the boundary and good detections (green

to red points) lying to its right. The qualitative behaviour of these plots is quite similar for both simulations. In both panels, most non-detections (grey symbols) lie above $\Delta\theta \sim 15^\circ$, i.e. 10 times the great-circle tolerance, so we take this to be the anticipated saturation limit of the method. Below this limit, there are few non-detections, which shows that the overall recovery rate is quite good.

The statistics summarizing the fractions and numbers of recovered progenitors are presented in Tables 4 and 5 for the Aquarius and HYDRO-zoom haloes, respectively. All statistics include both primary and secondary detections. In Table 4, for each tracer, the

columns show the median PCM background counts N_{BG} ; the overall fraction of streams and progenitors recovered f_{str} , and in total, f_{all} ; the number of bound progenitors and streams inside the detection boundary in total $N_{\text{T}}^{\text{bnd}}$, $N_{\text{T}}^{\text{str}}$ and recovered $N_{\text{T}}^{\text{bnd}}$ and $N_{\text{T}}^{\text{str}}$. Here, we label as ‘streams’ those progenitors with a fraction of bound stars $f_{\text{bound}} \leq 0.9$. This limit on f_{bound} is arbitrary but results are not very sensitive to the specific choice because, as noted by Cooper et al. (2010) for the Aquarius haloes, most progenitors are either completely bound ($f_{\text{bound}} = 1.0$) or almost completely unbound ($f_{\text{bound}} < 0.1$), which is also true for the HYDRO-zooms. Table 4 also summarizes the numbers of streams recovered with both RRLSs and K giants $N_{\text{RR+K}}$, those recovered with RRLSs only N_{RRnotK} or K giants only N_{KnotRR} , and when both tracers are combined $N_{\text{RR+K}}$, which is simply the sum of the previous numbers ($N_{\text{RR+K}} = N_{\text{RR+K}} + N_{\text{RRnotK}} + N_{\text{KnotRR}}$).

Table 5 presents the same summary statistics for the HYDRO-zooms. The results with the *in situ* background and the exclusion zone defined in Section 7.2 are presented on the right-hand side of the table and, for comparison, results that would be obtained leaving out the *in situ* component are shown on the left-hand side. The results obtained without the *in situ* component (left) are presented here to emphasize that the simple cuts used to filter out disc stars are very effective, since with them we recover as many progenitors (or more, in the case of halo 004) as in the case where we exclude the entire *in situ* background. Hence, the conclusions derived from the gas dynamical simulations are not driven by the presence of the *in situ* background. For haloes 001, 006, 008 and 009, not only is the same number of progenitors recovered, but they are the same progenitors in both cases. For halo 004, as noted, two more progenitors are detected (one bound, one unbound) compared to the case when the *in situ* component is left out. In this particular case, this happens because many stars from a massive and very heavily disrupted progenitor are removed by the exclusion zone, lowering the background enough for these peaks to be revealed.

For the Aquarius haloes, the overall fractions f_{str} and f_{all} in Table 4 show that using K giants as tracers, a median 88 per cent of streams and 86 per cent of all progenitors (bound and unbound) inside the boundary are recovered below the angular width limit of 15° . Out of these, a median 77 per cent are primary detections. When RRLSs are used, the total recovery rate is only slightly lower but still very good, yielding a median of 88 per cent for streams and 80 per cent for all progenitors, with 75 per cent of progenitors recovered as primary detections. The difference between the two tracers therefore does not lie in the relative efficiency, which is the same, but in the total number of progenitors that can be observed.

For the HYDRO-zoom haloes, using K giants, the median recovery fraction for streams is 75 per cent, and for all progenitors it is 67 per cent, out of which a median 67 per cent are primary detections. Note that these recovery fractions are only slightly lower than obtained with Aquarius with the same tracer. This confirms that, within the boundary, the detectability is not significantly affected by the fact that progenitors are more easily disrupted in these simulations.

The detection boundary for the HYDRO-zooms lies at $N_{\text{KIII}} \sim N_{\text{BG}}$ for small $\Delta\theta$, which Table 5 shows is around 200–800. This overlaps with the limit of $N_{\text{KIII}} \sim 500$ –1000, found in the previous section, where we estimate that HYDRO-zoom results might be hampered by a lack of progenitors due to the lower resolution of the simulations. Therefore, in what follows, we will only use results from the Aquarius simulations in our analysis of the number of streams expected within the detection boundary.

The Aquarius simulation results show that a total of 3–8 and 3–10 streams would be recovered successfully with *Gaia*+nGC3 when observed with K giants and with RRLSs, respectively. Note that, in the case of the MW, this implies that *Gaia* could potentially double the number of known dwarf galaxy streams in the halo. Since the detection limit of our method is well above the typical stellar mass associated with haloes at the resolution limit of Aquarius (as shown by Fig. 12), we do not expect that these results would be significantly different in a simulation with even higher resolution. The HYDRO-zoom results imply that a total of two to seven streams would be recovered successfully with *Gaia*+nGC3 with K giants.

A total of 2–6 or 0–1 bound progenitors would be recovered with K giants or RRLSs, respectively. The difference between results with different tracers is more notable in the number of recovered bound progenitors than in the number of tidal streams. This is due to the combination of two factors: that K giants are observable up to distances twice as large as RRLSs and that partially or completely unbound structures such as streams tend to spread stars out over larger ranges of (heliocentric) distance, making it more likely for these structures to have observable RRLSs, in comparison to a bound progenitor, for which all stars lie at approximately the same distance. This is an interesting result as it shows that approximately the same number of streams can be recovered with RRLSs as with K giants, even though RRLSs probe a substantially smaller volume.

In addition to the number of streams recovered being similar, there is the question of whether the two tracers recover the same streams or not. The last four columns of Table 4 provide this information. N_{RRnotK} shows that there can be up to five streams recovered with RRLSs that are not recovered with K giants. This demonstrates that there is something to be gained by using both tracers, instead of just the brighter one. When results from RRLSs and K giants are combined, 4–13 streams are recovered successfully ($N_{\text{RR+K}}$), which implies a median gain of two extra streams compared to results obtained with K giants alone.

The streams we are considering are tidal streams produced by dwarf galaxies, which is why the search tolerance has been tuned to the relatively large value of $\delta\theta = 1.5^\circ$ (see M11). With a lower tolerance, nGC3 could also identify the much narrower globular cluster streams, of which many are known in the MW. For example, the Pal 5 tidal stream has a full width at half-maximum (FWHM) of 0.3° (Odenkirchen et al. 2003) and the GD-1, Cocytos, Acheron and Lethe streams, all thought to have been produced by disrupted globular clusters, have FWHM of 0.5° (Grillmair 2006), 0.7° , 0.9° and 0.4° , respectively (Grillmair 2009). The resolution of the Aquarius (and HYDRO-zoom) simulations is not sufficient, however, to simulate globular cluster analogues. We therefore leave the exploration of the detectability of globular cluster streams for a future work.

Evidently, the estimates presented here are made under the assumption that the simulations we have used are representative of the MW. Although this is not exactly the case (see Sections 2.1 and 2.2), they still provide a useful estimate of the number of streams we can expect to detect with *Gaia*+nGC3 and the selected tracers.

The works of Sharma et al. (2011) and Elahi et al. (2013) have also explored the performance of other stream-finding algorithms, EnLink and S-tracker, VELOCIRAPTOR, ROCKSTAR and HOT-6D, respectively; and used the same definition of purity, among other statistics, to quantify their results. Our findings regarding the purities of the streams recovered by nGC3 (see Fig. 12) are competitive with their results: we find median purities of 0.44 and 0.55 for K giants and RRLSs, respectively; Sharma et al. (2011) obtain purities of ~ 0.66 – 0.70 for their 2MASS M giant and RRLS Large Synoptic Survey Telescope (LSST) synthetic samples; and Elahi

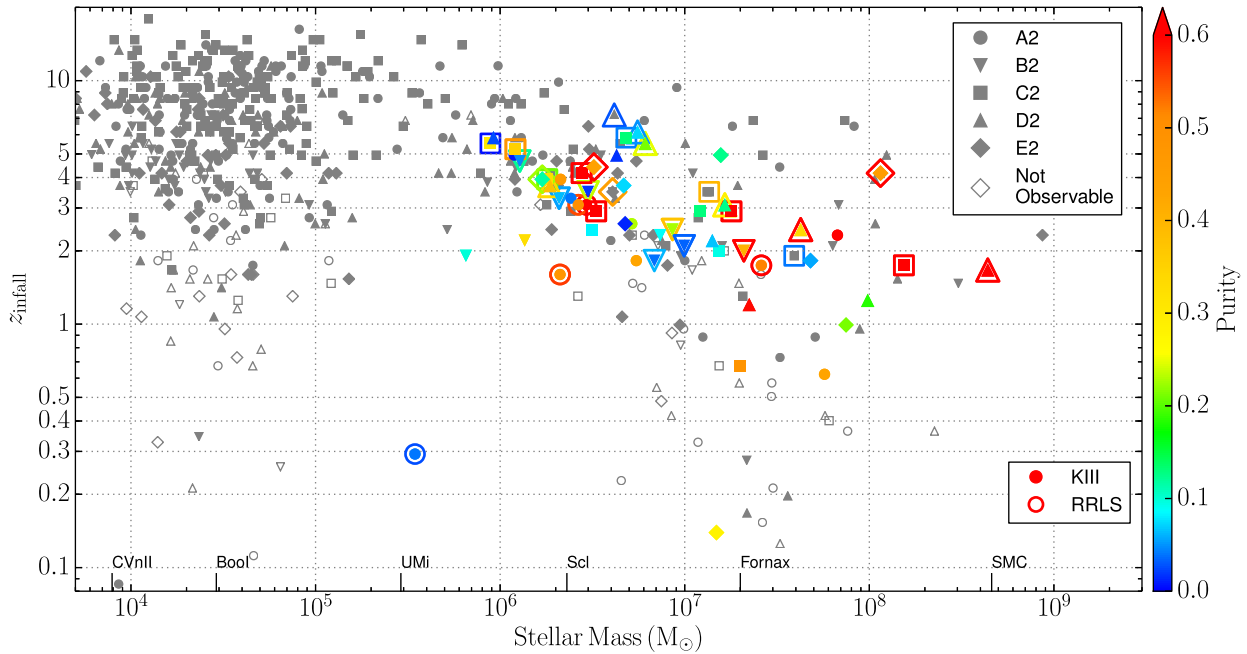


Figure 13. Infall redshift versus stellar mass for progenitors in the Aquarius haloes. Progenitors successfully recovered (with $\Delta\theta \leq 15^\circ$) with K giants and RRLSs are shown, respectively, with filled and open coloured symbols, with a colour proportional to the purity of the detection. Small grey symbols denote progenitors that are observable ($G \leq 20$ and $\Delta\mu/\mu \leq 0.5$) but not recovered (grey filled), or not observable at all (grey open). The stellar masses of the Small Magellanic Cloud and a few classical (Fornax, Scl, UMi) and ultra-faint (BooI, CVnII) dwarf spheroidal MW satellites from McConnachie (2012) are shown in the bottom axis for reference. The stellar mass scale divided by 3 is roughly equivalent to an L_V luminosity scale, for galaxies dominated by an old metal-poor population (>10 Gyr, $[\text{Fe}/\text{H}] = -1.5$).

et al. (2013) find purities of ~ 0.85 for tidally disrupted subhaloes and 0.40 for *completely* disrupted subhaloes. This is remarkable, as our simulations include the effects of the *Gaia* selection function and observational errors, while the Elahi et al. simulations are error-free and although Sharma et al. simulate their M giant sample with similar distance errors (18 per cent) as our K giants, for the RRLSs they assume the much deeper LSST selection function ($m_r = 24.5$) and assume slightly better distance errors (5 per cent).

8.3 The progenitor stellar masses

A plot of the infall redshift versus total stellar mass for all progenitors in the Aquarius haloes is shown in Fig. 13. All progenitors with at least one observable star ($G \leq 20$ and $\Delta\mu/\mu \leq 0.5$) are plotted with filled grey symbols, and progenitors with no observable stars are plotted as open grey symbols. Recovered progenitors inside the detection boundary, below the angular width threshold $\Delta\theta = 1.5^\circ$, are shown with coloured symbols, the colour scale being proportional to the purity. Spurious detections (left of the detection boundaries in Fig. 12) are not shown. Filled and open coloured symbols denote progenitors recovered with K giants and RRLSs, respectively.

Fig. 13 shows that recovered progenitors have masses down to a few times $10^6 M_\odot$, similar to or lower than that of the Sculptor dwarf spheroidal galaxy (McConnachie 2012), and even below $10^6 M_\odot$ in a handful of cases. This mass limit is of the order of the least massive ‘classical’ dwarf spheroidal satellites of the MW ($\sim 3 \times 10^5 M_\odot$, Ursa Minor and Draco). Interestingly, progenitors can be recovered down to this mass limit and in the same mass range with both tracers, as evidenced in the plot by the fact that there are open and filled coloured symbols spanning the same mass range and overlapped in most cases. In fact, the only clear difference

between detection with both tracers, in terms of their distribution in this plane, is that progenitors accreted relatively recently $z_{\text{infall}} \lesssim 1$ are detected only with K giants. Note also that the majority of progenitors that are not observable also lie in this redshift range; this is precisely because, having been accreted only recently by the main halo, most of these progenitors are almost completely bound and very distant so only their brightest stars are observable.

As we have seen in Section 8.2 and Table 4, the same number of *streams* (unbound progenitors) in total are detectable with RRLSs as with K giants, but the number of *bound progenitors* is smaller because RRLSs are intrinsically fainter and thus probe a smaller volume. Fig. 13 shows that progenitors recovered with RRLSs are not limited to the most massive/luminous ones, but span the same mass range as those recovered with K giants.

8.4 The progenitor distance distribution

The heliocentric distance distribution of progenitors is illustrated in Fig. 14. The panels show the infall redshift as a function of the median heliocentric distance for progenitors more massive than $10^6 M_\odot$ in the Aquarius (left) and HYDRO-zoom (right) haloes. As in the previous figures, filled and open coloured symbols denote progenitors recovered, respectively, with KIII and RRLSs and filled and open grey symbols progenitors that are not recovered or not observable, respectively. Error bars depict the interquartile range of the heliocentric distance distribution of stars in each progenitor, so points with no visible error bars correspond mostly to bound progenitors and, less commonly, to very distant streams in almost perfectly circular orbits.

The two panels in Fig. 14 show that progenitors overall are recovered in the ~ 20 –130 kpc distance range. At the lower distance end, the plot shows detections down to ~ 20 –30 kpc in both types of

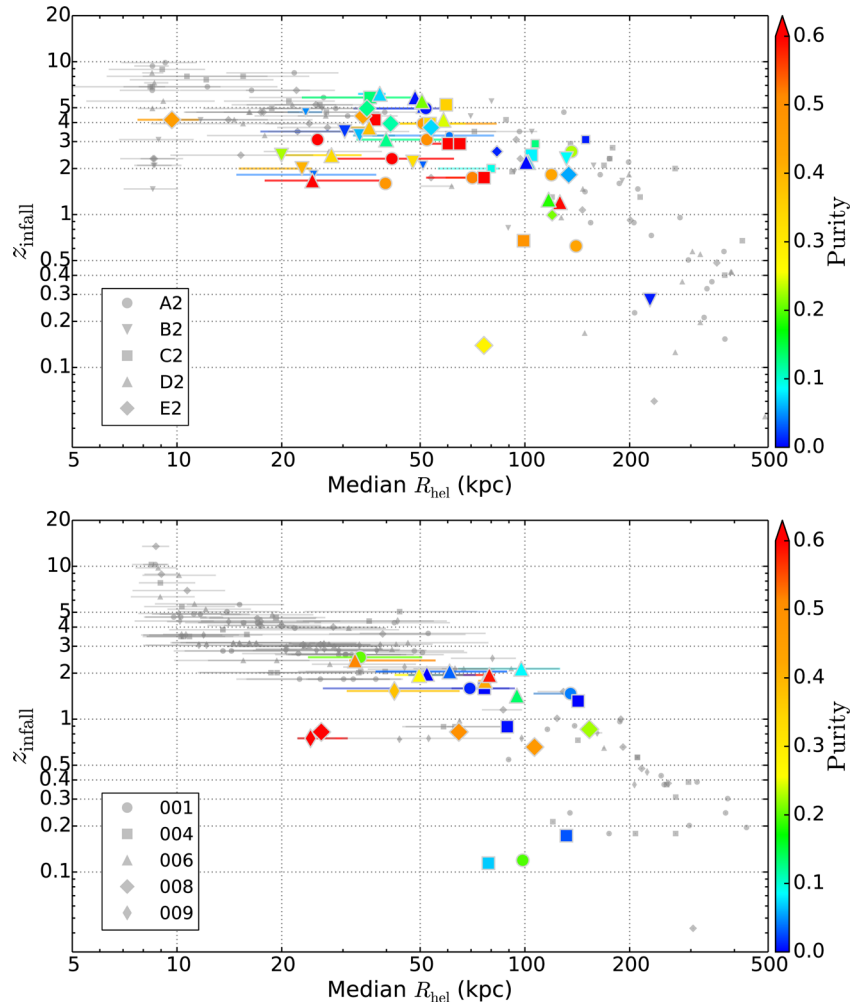


Figure 14. Infall redshift versus median heliocentric distance for progenitors in the Aquarius (left) and HYDRO-zoom (right) haloes. Only progenitors more massive than $10^6 M_\odot$ are shown. As in the previous figure, progenitors successfully recovered (with $\Delta\theta \leq 15^\circ$) with K giants and RRLs are shown, respectively, with filled and open coloured symbols, with a colour proportional to the purity of the detection. Small grey symbols denote progenitors that are observable but not recovered (grey filled). The error bars denote the interquartile range of the heliocentric distance distribution of the stars in each progenitor.

simulations, the dark-matter-only Aquarius and the gas dynamical HYDRO-zooms. As noted in M11 and Smith (2016), great-circle methods are expected to work best at intermediate to large distances ($>20\text{--}30$ kpc), where tidal streams are less affected by the presence of the disc and by phase mixing given the longer dynamical time-scales. This is reinforced by the $z_{\text{infall}}\text{--}R_{\text{hel}}$ trend observed in the plot that shows streams in the inner Galaxy (<20 kpc) were accreted at higher z_{infall} , which translates into larger stream widths $\Delta\theta$ (Fig. 11) making detection more difficult. At the higher distance end, progenitors are much more recent infallers and there are no streams (i.e. unbound progenitors) beyond 100 kpc in the Aquarius simulations, and only a couple in the HYDRO-zooms.

The nGC3 method is most efficient at detecting streams in the distance range from ~ 30 to ~ 90 kpc. As the left-hand panel shows, most progenitors are unbound ones and the majority are recovered (84 and 80 per cent with KIII and RRLs, respectively). Therefore, based on the Aquarius simulation results, in this distance range we expect a search for streams using either KIII or RRLs to be fairly complete (>80 per cent) for progenitors more massive than $10^6 M_\odot$. As discussed in Section 8.2, we cannot draw any conclusions about the completeness of the search with the HYDRO-zoom simulations due to limitations caused by mass resolution.

The distribution of open and filled symbols in the left-hand panel also shows no distance bias in the stream detections made with KIII compared to RRLs up to a median distance of ~ 90 kpc. This is an interesting result as it is rather counter-intuitive since the maximum distance up to which RRLs will be observable since the *Gaia* is ~ 50 kpc (see Section 4 and Figs 1 and 2). However, in streams, stars are scattered across a range of distances making them still detectable at larger median distances. Combined with results from Sections 8.2 and 8.3, this means that RRLs probe the same effective volume ($\sim 20\text{--}90$ kpc) and mass range ($\gtrsim 10^6 M_\odot$), with a similar completeness as KIII stars when it comes to streams.

9 GAIA MISSION LIFETIME EXTENSIONS

In what follows, we will evaluate the effect a possible extension of the *Gaia* mission will have on the number of progenitors we expect to detect with *Gaia*+nGC3.

An extension of the *Gaia* mission lifetime will translate into improved proper motion and parallax precisions. The survey completeness magnitude will remain the same, $G = 20$, as this is set by the available antenna time to download data for the stars detected on-board up to the set magnitude limit. Since we have assumed

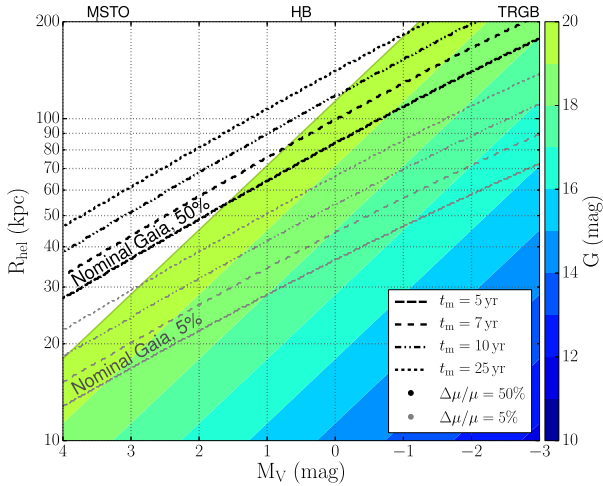


Figure 15. Proper motion relative error horizons for the different *Gaia* mission lifetime scenarios, in the heliocentric distance R_{hel} versus absolute magnitude M_V plane. The colour scale is proportional to the apparent G magnitude and goes up to the *Gaia* magnitude limit ($G \leq 20$), assuming a fixed $V - I = 1$ colour and $A_V = 0$. Grey and black lines, respectively, show horizons for proper motion relative errors of 5 and 50 per cent. The nominal mission lifetime $t_m = 5$ yr horizon is indicated with the long-dashed line (also shown in Fig. 1 with black long dashes); the three possible extension scenarios of $t_m = 7, 10$ and 25 yr are indicated, respectively, with short-dashed, dash-dotted and dotted lines. The absolute magnitudes of MSTO, HB and TRGB stars are shown for reference on the top axis.

throughout this work that photometric distances will be used for our tracers and we have so far neglected radial velocities, in what follows we will only consider the effect of the increase in the proper motion precision.

We will consider the following three scenarios:

- (i) a two-year extension of *Gaia*, increasing the total mission time to 7 yr;
- (ii) a five-year extension of *Gaia*, increasing the total mission time to 10 yr;
- (iii) a *Gaia* twin mission launched in 20 yr time, increasing the overall mission baseline to 25 yr.

The first two are realistic short-term scenarios, depending on the satellite’s fuel budget and instrument performance at the end of the nominal mission lifetime of 5 yr in 2019 (Brown, private communication). In these two scenarios, we assume that the proper motion errors will decrease by the expected factor of $(t_{\text{nom}}/t_m)^{3/2}$, where t_m is the new mission lifetime and $t_{\text{nom}} = 5$ yr is the nominal mission duration.⁸

The third scenario is a medium-term possibility. In this case, we assume that the proper motion errors will decrease by a more conservative factor of (t_{nom}/t_m) , as even though there would be a much longer baseline of 25 yr, there will be a gap in the data for the ~ 20 yr in between the two missions (Brown and de Bruijne, private communication).

Fig. 15 shows a new version of Fig. 1 illustrating the proper motion error prescriptions expected for each of the three scenarios. The scaling factors for the proper motion errors are 0.60, 0.35 and 0.20 for the $t_m = 7, 10$ and 25 yr scenarios, respectively.

⁸ The factor of $t^{3/2}$ comes from a factor of $t^{1/2}$ due to photon noise reduction and a factor t due to the longer time baseline (Brown and de Bruijne, private communication).

Table 6. Statistics of the progenitor recovery for the different *Gaia* lifetime scenarios. N_T is the total number of detectable progenitors and f_{70} the fraction of progenitors with more than 70 per cent of stars observable by *Gaia*.

Halo	$t_m = 5$ yr		$t_m = 7$ yr		$t_m = 10$ yr		$t_m = 25$ yr		N_∞
	N_T	f_{70}	N_T	f_{70}	N_T	f_{70}	N_T	f_{70}	
KIII									
A2	14	0.66	15	0.75	16	0.77	17	0.84	26
B2	11	0.75	13	0.75	14	0.82	14	0.90	16
C2	18	0.66	18	0.71	18	0.82	19	0.86	26
D2	13	0.51	14	0.63	14	0.69	14	0.78	18
E2	9	0.51	10	0.66	11	0.75	12	0.81	15
RRLS									
A2	7	0.70	9	0.77	9	0.86	9	0.93	9
B2	8	0.70	8	0.83	9	0.87	11	0.93	11
C2	12	0.59	12	0.67	13	0.85	13	0.94	13
D2	10	0.60	11	0.76	11	0.86	11	0.95	11
E2	5	0.67	6	0.86	6	0.95	7	0.95	7

For each scenario, we produce *Gaia* mock catalogues of the Aquarius simulations, rescaling the proper motion error prescriptions by the appropriate factors. As we have already shown, the nGC3 has a clear detection boundary; we simply need to compute the full nGC3 PCM for each halo to estimate the median pole counts in the smoothed PCM (see Section 6.2) and, using equation (4), count how many progenitors lie inside the detectability boundary below the $\Delta\theta = 15^\circ$ threshold we have used so far. This gives us N_T , the total number of detectable progenitors. We also compute f_{70} , the fraction of progenitors that have more than 70 per cent of their stars observable by *Gaia* with proper motion precision better than 50 per cent. These results are summarized in Table 6 for RRLSs and KIII, starting with the nominal mission lifetime $t_m = 5$ yr for comparison (leftmost columns), followed by the three scenarios considered.

Table 6 shows that, in general, the number of detectable progenitors N_T increases by only one in each scenario with respect to the previous one, so that for any halo an overall increase of two to three progenitors at most is expected in the last scenario of a second *Gaia* mission in 25 yr time, with respect to the nominal 5 yr mission lifetime. This seems like a relatively small gain, but f_{70} shows why this should be the case: in all the Aquarius haloes, more than half the progenitors will have over 70 per cent of their tracer stars meeting the proper motion criterion, even for the nominal 5 yr mission lifetime. This fraction increases up to ~ 70 per cent for the $t_m = 7$ yr mission lifetime scenario and up to ~ 80 – 95 per cent, depending on the halo, for the $t_m = 25$ yr scenario of a second *Gaia* mission.

The last column of Table 6 gives N_∞ , the number of detectable progenitors expected in the limiting case considering *Gaia* observable stars but without any observational errors. This gives us an estimate of how many more progenitors we could expect to get in an ideal case, with a fixed $G = 20$ *Gaia* completeness limit. This shows that a maximum number of 15–26 progenitors, depending on the halo, could be detectable with KIII stars and *Gaia*+nGC3 in the ideal error-free case. So, with K giants, there would still be room for improvement since in the $t_m = 25$ yr scenario the number of detectable progenitors could be increased by 20–50 per cent depending on the halo.

On the other hand, for RRLSs it is clear that, even for the nominal mission time, the number of detectable progenitors is quite close to N_∞ ; and it reaches this limit, for all five Aquarius haloes, in the

$t_m = 25$ yr scenarios. In other words, all progenitors that could be detected with RRLs in an ideal *Gaia* error-free case are indeed detectable as they do lie inside the method's detection boundary. The fact that this happens for RRLs and not K giants is most likely due to the notably smaller distance errors RRLs have in comparison to K giants, which, even though the GC3 methods have been implemented so as to minimize the effect of distance errors, are still expected to have an impact (see M11). This result highlights the importance of obtaining radial velocities for RRLs identified during the nominal mission time, which will be necessary for removing contaminants and for detailed modelling of the stream candidates found.

Although there is an increase in the number of detectable progenitors, the expected improvement in the future *Gaia* scenarios explored is relatively modest, considering that these numbers could improve by up to a further ~ 50 per cent for K giants. In the next section, we comment on possible strategies to improve upon these results.

10 PUSHING THE DETECTION BOUNDARY: RECOMMENDATIONS AND IMPROVEMENTS

Throughout this work, we have analysed the performance of the nGC3 method detecting tidal streams and satellites in cosmological simulations and we have discussed some possible recommendations and improvements for the time this can be applied to real data. Our suggestions and recommendations can be summarized as follows.

(i) Radial velocities will help reduce foreground/background contamination, in as much as they can be obtained for large samples of stars, and they are also necessary to disentangle different streams that share an orbital plane. Spectroscopic surveys planned and ongoing like LAMOST, WEAVE, 4MOST and DESI (Dalton et al. 2012; Liu et al. 2014; de Jong & Consortium 2015; Eisenstein & DESI Collaboration 2015) will make an important contribution in this respect, providing radial velocities for K giants spanning large portions of the volume probed by *Gaia*.

(ii) *Gaia* radial velocities could also be incorporated, when available, by combining the mGC3 PCM for stars with full 6D information with nGC3 PCMs for the remaining stars.

(iii) Great-circle-cell count methods are linear, so PCMs from different tracers could be combined by simple addition. As we discuss in Section 6.3, for a given survey, it would be optimal to add PCMs from different tracers *after* unsharp masking.

(iv) The use of simple cuts can effectively reduce background contamination minimizing its effect on progenitor detectability, as we have shown in Section 8.2. Thus, it would prove useful to analyse the use of other cuts that can help reduce the background even further.

(v) Further improvements of this method can be made by combining it with the chemical abundance information. For example, knowing that many intermediate and metal-rich stars in the halo belong to tidal debris from massive satellite galaxies or from those accreted more recently (Gilbert et al. 2009), the number of detections can be maximized by targeting this metallicity range preferentially. A broad classification as metal-poor, intermediate or metal-rich will be feasible with *Gaia* BP/RP spectrophotometry, which can be used to produce separate PCMs in each metallicity bin. We intend to test in the future how the combination of chemical abundances and GC3 methods can improve on the recovery of substructure in the halo.

Other benefits will come from improvements in the peak detection algorithm and the pole-counting strategy such as (i) using a

deblending algorithm in the peak detection and incorporating the fact that peaks in PCMs stretch along great-circle arcs (Torii & Imiya 2005), (ii) weighing the contribution of stars to poles proportionally to the observational errors, (iii) assigning pole-membership probabilities to each star and (iv) using the full sphere in pole space to differentiate structures with different sense of rotation.

11 CONCLUSIONS

Tidal streams are widely recognized for their usefulness in the inference of the Galactic accretion history, one of the key science drivers for the *Gaia* mission (de Bruijne 2012). However, any such inference demands a thorough understanding of the selection biases that may affect tidal stream detection methods.

Motivated by this, and the prospects that the *Gaia* mission opens up for all-sky homogeneous stream surveying, we have explored the detectability of tidal streams in *Gaia* mock catalogues using nGC3, a great-circle-cell count method that uses positional information and proper motions (Abedi et al. 2014). We have built mock catalogues for two standard candle tracers: K giants and RRLs, reproducing the *Gaia* selection function and observational errors, and assuming photometric distance errors of 20 and 7 per cent, respectively, for each tracer. These mock catalogues were made from a set of five haloes from the Aquarius *N*-body simulations and five haloes from the HYDRO-zoom gas dynamical simulations. The diversity of orbits and progenitors in these allows us to characterize the nGC3 method's completeness and detection limits in a realistic setting. We have also explored how the *in situ* stellar halo background in HYDRO-zoom gas dynamical simulations affects the detection of streams, and the improvements in proper motion errors expected for three possible extensions of the *Gaia* mission.

We summarize our results as follows.

(i) The nGC3 method is able to identify realistic tidal streams produced in cosmological *N*-body and gas dynamical simulations, even when contamination from a smooth halo background is included.

(ii) The method has a *well-defined parameter-free detection boundary* in the plane of angular width versus ratio of observable to PCM background stars, defined in equation (4).

(iii) Progenitors are recovered up to infall redshifts as large as $z_{\text{infall}} \sim 3$ based on results with the gas dynamical simulations, in which progenitors are more prone to disruption.

(iv) A total of 9–12 progenitors, bound and unbound, are expected to be detectable with *Gaia*+nGC3 using KIII stars as tracers, and 4–10 using RRLs. These correspond respectively to a median 86 and 80 per cent of all progenitors inside the detection boundary, below our selected threshold of $\Delta\theta = 15^\circ$.

(v) A total of 3–8 *streams* would be recovered successfully with *Gaia*+nGC3 when observed with K giants and 3–10 with RRLs. Depending on the specific merger history of the MW, this means that *Gaia* has the potential to almost double the number of known tidal streams in the halo. Also, approximately the same number of streams can be recovered with RRLs as with K giants, even though RRLs probe a substantially smaller volume.

(vi) When results from RRLs and K giants are combined, 4 to 13 streams are recovered successfully ($N_{\text{RR+K}}$), which implies a median gain of two extra streams when compared to results obtained with K giants alone.

(vii) The stellar masses and luminosities of recovered progenitors go down to $\sim 10^6 M_\odot$ and $\sim 4 \times 10^5 L_\odot$, respectively, i.e. similar to the classical dwarf spheroidal MW satellites.

(viii) Our forecasts in items (iv)–(vii) are based on results from the Aquarius simulations alone, since HYDRO-zoom results may be hampered due to their lower mass resolution.

(ix) Progenitors are recovered down to the same stellar mass limit and the same infall redshift range with either tracer, RRLSs or K giants.

(x) Recovered progenitors span the heliocentric distance range from 20 to 130 kpc, with the best completeness (>80 per cent) achieved in the range from ~ 30 to ~ 90 kpc.

(xi) For streams (i.e. partially unbound progenitors), RRLSs probe *the same effective volume* (~ 20 – 90 kpc) and mass range ($\gtrsim 10^6 M_\odot$), with a similar completeness, as KIII stars. For bound progenitors, KIII stars probe a larger volume, reaching out to ~ 130 kpc.

(xii) We analysed the detectability of progenitors also for gas dynamical simulations that naturally include the *in situ* background. Although, as expected, the contamination from this additional background hinders the detections, we find that using a simple cut to exclude the disc stars ($|b| \leq 10^\circ$ and $R \leq 20$ kpc) one can recover as many progenitors (or more in one case) as in the case when the *in situ* component is not taken into account.

(xiii) We analysed how the detectability of progenitors would be improved by the smaller proper motion errors resulting from an extension of the *Gaia* mission lifetime. The three scenarios considered were a two-year extension, a five-year extension and a second *Gaia* mission launched in 20 yr. In these scenarios, proper motion errors would be reduced by factors of 0.6, 0.35 and 0.2, respectively. Increases of about one, two and three progenitors, respectively, are expected in each scenario with respect to the results found for the nominal mission lifetime.

Finally, the K giant and RRLS *Gaia* mock catalogues produced for both the Aquarius and HYDRO-zoom simulations are publicly available at this URL. These catalogues include, for each star, all the position and velocity information in heliocentric spherical and Galactocentric Cartesian coordinate systems, with and without simulated *Gaia* errors, including the pole ID indicating with which pole detection (if any) it is associated in the nGC3 PCM. Each pole detection catalogue thus represents a realistic set of stream detections in which streams may overlap, stars will be missing and there will be contamination from the smooth background and foreground. These catalogues will be a useful benchmark for further studies on the inference of the Galactic accretion history and gravitational potential.

ACKNOWLEDGEMENTS

We thank the anonymous referee for a careful reading of our manuscript and for suggestions that helped improve its clarity. CM and LA acknowledge support from DGAPA/UNAM grant IG100115. CM acknowledges the support of the post-doctoral fellowship of DGAPA-UNAM, Mexico; and the European Commission's Framework Programme 7, through the Marie Curie International Research Staff Exchange Scheme LACEGAL (PIRSES-GA-2010-269264), and is grateful for the hospitality of LJMU and the Institute for Computational Cosmology (ICC), Durham University, where part of this research was carried out. LA also acknowledges the hospitality of the Institute for Computational Cosmology (ICC), Durham University. APC and WW are supported by the COFUND Junior Research Fellowship scheme under EU grant 267209 and acknowledge support from STFC (ST/L00075X/1). *Gaia* error simulations were carried out using ATAI, a high-performance cluster,

at IA-UNAM. The Aquarius Project was carried out at the Leibniz Computing Centre and Computing Centre of the Max-Planck-Society in Garching, Germany; on the ICC COSMA system; and on the STELLA supercomputer of the LOFAR experiment at the University of Groningen. The mock catalogues were generated with the DiRAC Data Centric system at Durham University, operated by the ICC 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. We thank the organizers of the 2nd *Gaia* Challenge Workshop, where this project was initiated. The 2nd *Gaia* Challenge Workshop (2014 October, Max Planck Institute for Astronomy, Heidelberg) was supported by Collaborative Research Center SFB881 'The Milky Way System'. CM acknowledges the use of TOPCAT (Taylor 2005) throughout the course of this investigation.

REFERENCES

- Abedi H., Mateu C., Aguilar L. A., Figueras F., Romero-Gómez M., 2014, *MNRAS*, 442, 3627
- Baker M., Willman B., 2015, *AJ*, 150, 160
- Berry D. S., 2015, *Astron. Comput.*, 10, 22
- Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, *MNRAS*, 370, 645
- Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., 2010, *MNRAS*, 406, 896
- Bressan A., Marigo P., Girardi L., Salasnich B., Dal Cero C., Rubele S., Nanni A., 2012, *MNRAS*, 427, 127
- Brown A. G. A., Velazquez H. M., Aguilar L. A., 2005, *MNRAS*, 359, 1287
- Bullock J. S., Johnston K. V., 2005, *ApJ*, 635, 931
- Colless M. et al., 2001, *MNRAS*, 328, 1039
- Cooper A. P. et al., 2010, *MNRAS*, 406, 744
- Cooper A. P., D'Souza R., Kauffmann G., Wang J., Boylan-Kolchin M., Guo Q., Frenk C. S., White S. D. M., 2013, *MNRAS*, 434, 3348
- Cooper A. P., Parry O. H., Lowing B., Cole S., Frenk C., 2015, *MNRAS*, 454, 3185
- Cooper A. P., Cole S., Frenk C. S., Le Bret T., Pontzen A., 2016, preprint ([arXiv:1611.03497](https://arxiv.org/abs/1611.03497))
- Crain R. A. et al., 2015, *MNRAS*, 450, 1937
- Dalton G. et al., 2012, *Proc. SPIE*, 8446, 84460P
- de Bruijne J. H. J., 2012, *Ap&SS*, 341, 31
- de Bruijne J. H. J., Rygl K. L. J., Antoja T., 2014, in Walton N. A., Figueras F., Balaguer-Núñez L., Soubiran C., eds, *EAS Publ. Ser. Vol. 67, GREAT Science from the Gaia Data Releases*. ESA, Noordwijk, p. 23
- de Jong R. S., Consortium, 2015, *IAU General Assembly*, 22, 2255843
- Deg N., Widrow L., 2014, *MNRAS*, 439, 2678
- Drimmel R., Cabrera-Lavers A., López-Corredoira M., 2003, *A&A*, 409, 205
- Eisenstein D., DESI Collaboration, 2015, *Am. Astron. Soc. Meeting Abstr.*, 225, 336.05
- Elahi P. J. et al., 2013, *MNRAS*, 433, 1537
- Errani R., Peñarrubia J., Laporte C. F. P., Gómez F. A., 2017, *MNRAS*, 465, L59
- Font A. S., McCarthy I. G., Crain R. A., Theuns T., Schaye J., Wiersma R. P. C., Dalla Vecchia C., 2011a, *MNRAS*, 416, 2802
- Font A. S. et al., 2011b, *MNRAS*, 417, 1260
- Gao L., Navarro J. F., Cole S., Frenk C. S., White S. D. M., Springel V., Jenkins A., Neto A. F., 2008, *MNRAS*, 387, 536
- Garrison-Kimmel S. et al., 2017, preprint ([arXiv:1701.03792](https://arxiv.org/abs/1701.03792))
- Gilbert K. M., Font A. S., Johnston K. V., Guhathakurta P., 2009, *ApJ*, 701, 776
- Gómez F. A., Helmi A., 2010, *MNRAS*, 401, 2285
- Grillmair C. J., 2006, *ApJ*, 651, L29

- Grillmair C. J., 2009, *ApJ*, 693, 1118
- Helmi A., de Zeeuw P. T., 2000, *MNRAS*, 319, 657
- Helmi A., White S. D. M., 1999, *MNRAS*, 307, 495
- Helmi A., Cooper A. P., White S. D. M., Cole S., Frenk C. S., Navarro J. F., 2011, *ApJ*, 733, L7
- Johnston K. V., Hernquist L., Bolte M., 1996, *ApJ*, 465, 278
- Johnston K. V., Bullock J. S., Sharma S., Font A., Robertson B. E., Leitner S. N., 2008, *ApJ*, 689, 936
- Le Bret T., Pontzen A., Cooper A. P., Frenk C., Zolotov A., Brooks A. M., Governato F., Parry O. H., 2015, preprint ([arXiv:1502.06371](https://arxiv.org/abs/1502.06371))
- Liu C. et al., 2014, *ApJ*, 790, 110
- Lowing B., Wang W., Cooper A., Kennedy R., Helly J., Cole S., Frenk C., 2015, *MNRAS*, 446, 2274
- Lynden-Bell D., Lynden-Bell R. M., 1995, *MNRAS*, 275, 429
- McCarthy I. G., Schaye J., Font A. S., Theuns T., Frenk C. S., Crain R. A., Dalla Vecchia C., 2012, *MNRAS*, 427, 379
- McConnachie A. W., 2012, *AJ*, 144, 4
- Mateu C., Bruzual G., Aguilar L., Brown A. G. A., Valenzuela O., Carigi L., Velázquez H., Hernández F., 2011, *MNRAS*, 415, 214 (M11)
- Mateu C., Vivas A. K., Downes J. J., Briceño C., Zinn R., Cruz-Díaz G., 2012, *MNRAS*, 427, 3374
- Mignard F. et al., 2008, in Jin W. J., Platais I., Perryman M. A. C., eds, *Proc. IAU Symp. 248, A Giant Step: From Milli- to Micro-arcsecond Astrometry*. Cambridge Univ. Press, Cambridge, p. 224
- Morrison H. L., Mateo M., Olszewski E. W., Harding P., Dohm-Palmer R. C., Freeman K. C., Norris J. E., Morita M., 2000, *AJ*, 119, 2254
- Navarro J. F. et al., 2010, *MNRAS*, 402, 21
- Odenkirchen M. et al., 2003, *AJ*, 126, 2385
- Perryman M. A. C. et al., 2001, *A&A*, 369, 339
- Pillepich A., Madau P., Mayer L., 2015, *ApJ*, 799, 184
- Price-Whelan A. M., Johnston K. V., 2013, *ApJ*, 778, L12
- Robin A. C. et al., 2012, *A&A*, 543, A100
- Romero-Gómez M., Figueras F., Antoja T., Abedi H., Aguilar L., 2015, *MNRAS*, 447, 218
- Rygl K. L. J., Molinari S., Prusti T., Antoja T., Elia D., de Bruijne J., 2014, in Walton N. A., Figueras F., Balaguer-Núñez L., Soubiran C., eds, *EAS Publ. Ser. Vol. 67. GREAT Science from the Gaia Data Releases*. ESA, Noordwijk, p. 397
- Sanderson R. E., 2016, *ApJ*, 818, 41
- Sanderson R. E., Helmi A., Hogg D. W., 2015, *ApJ*, 801, 98
- Schaye J. et al., 2015, *MNRAS*, 446, 521
- Sesar B. et al., 2009, *ApJ*, 708, 717
- Sesar B. et al., 2013, *AJ*, 146, 21
- Sharma S., Steinmetz M., 2006, *MNRAS*, 373, 1293
- Sharma S., Johnston K. V., Majewski S. R., Bullock J., Muñoz R. R., 2011, *ApJ*, 728, 106
- Smith H. A., 1995, *Cambridge Astrophysics Series, Vol. 27, RR Lyrae Stars*. Cambridge Univ. Press, Cambridge
- Smith M. C., 2016, in Newberg H. J., Carlin J. L., eds, *Astrophysics and Space Science Library, Vol. 420, Tidal Streams in the Local Group and Beyond*. Springer, Berlin, p. 113
- Spergel D. N. et al., 2003, *ApJS*, 148, 175
- Springel V. et al., 2008a, *MNRAS*, 391, 1685
- Springel V. et al., 2008b, *Nature*, 456, 73
- Starkenburg E. et al., 2009, *ApJ*, 698, 567
- Taylor M. B., 2005, in Shopbell P., Britton M., Ebert R., eds, *ASP Conf. Ser. Vol. 347, Astronomical Data Analysis Software and Systems XIV*. Astron. Soc. Pac., San Francisco, p. 29
- Torii A., Imiya A., 2005, *Computer Analysis of Images and Patterns*. Springer, Berlin, p. 371
- Vivas A. K., Zinn R., 2006, *AJ*, 132, 714
- Vivas A. K. et al., 2004, *AJ*, 127, 1158
- Xue X.-X. et al., 2014, *ApJ*, 784, 170
- Zolotov A., Willman B., Brooks A. M., Governato F., Hogg D. W., Shen S., Wadsley J., 2010, *ApJ*, 721, 738

This paper has been typeset from a \LaTeX file prepared by the author.