Mon. Not. R. Astron. Soc. **000**, 1–11 (2017) Printed 7 November 2018 (MN LATEX style file v2.2)

Gaia DR2 in 6D: Searching for the fastest stars in the Galaxy

T. Marchetti[1]⋆, E. M. Rossi[1] and A. G. A. Brown[1]
[1]*Leiden Observatory, Leiden University, PO Box 9513 2300 RA Leiden, the Netherlands*

7 November 2018

ABSTRACT

We search for the fastest stars in the subset of stars with radial velocity measurements of the second data release (DR2) of the European Space Agency mission *Gaia* . Starting from the observed positions, parallaxes, proper motions, and radial velocities, we construct the distance and total velocity distribution of more than 7 million stars in our Milky Way, deriving the full 6D phase space information in Galactocentric coordinates. These information are shared in a catalogue, publicly available at `http://home.strw.leidenuniv.nl/~marchetti/research.html`. To search for unbound stars, we then focus on stars with a median total velocity in the Galactic rest frame > 450 km s^{-1} . This cut results in a clean sample of 165 sources with reliable astrometric parameters and radial velocities. Of these, 28 stars have probabilities greater than 50 % of being unbound from the Galaxy. On this latter sub-sample, we perform orbit integration to characterize the stars' orbital parameter distributions. We find 2 to 5 hypervelocity star candidates, stars that are moving on orbits consistent with coming from the Galactic Centre, and 9 hyper-runaway star candidates, coming from the Galactic disk. Surprisingly, the remaining unbound stars cannot be traced back to the Galaxy, including our two fastest stars (above 700 km s^{-1}). These may constitute the tip of the iceberg of a large extragalactic population or the extreme velocity tail of stellar streams.

Key words: Galaxy: kinematics and dynamics, Galaxy: stellar contents, Galaxy: centre.

1 INTRODUCTION

Stars with extremely high velocities have been long studied to probe our Galaxy. The interest in the high velocity tail of the total velocity distribution of stars in our Milky Way is twofold. First, it flags the presence of extreme dynamical and astrophysical processes, especially when the velocity of a star is so high that it approaches (or even exceeds) the escape speed from the Galaxy at its position. Secondly, high velocities stars, spanning a large range of distances, can be used as dynamical tracers of integral properties of the Galaxy. The stellar high velocity distribution has for example been used to trace the local Galactic escape speed and the mass of the Milky Way (e.g. Smith et al. 2007; Gnedin et al. 2010; Piffl et al. 2014). To put the concept of *high velocity* in context, the value of the escape speed is found to be ~ 530 km s^{-1} at the Sun position, it increases up to ~ 600 km s^{-1} in the central regions of the Galaxy, and then falls down to $\lesssim 400$ km s^{-1} at Galactocentric distances ~ 50 kpc (Williams et al. 2017). We adopt a minimum threshold of 450 km s^{-1} to define the "high velocity" tail of the Galactic velocity distribution.

A first class of objects that can be found in this tail is fast halo star. Their measured dispersion velocity is around 150 km s^{-1} (Smith et al. 2009; Evans et al. 2016), therefore 3-σ outliers can exceed this threshold, while remaining bound. Halo stars could also reach unbound velocities, when they are part of the debris of tidally disrupted satellite galaxies, like the Sagittarius Dwarf galaxy, that has not yet virialized (e.g. Abadi et al. 2009). Velocities outliers in the bulge and disc velocity distribution may also exist and become apparent in a large data set.

"Runaway stars" (RSs) form an another class of high velocity stars. They were originally introduced as O and B type stars traveling away from the Galactic disk with velocities higher than 40 km s^{-1} (Blaauw 1961). Theoretically, there are two main formation channels: i) dynamical encounters between stars in dense stellar systems such as young star clusters (e.g. Poveda et al. 1967; Leonard & Duncan 1990; Gvaramadze et al. 2009), and ii) supernova explosions in stellar binary systems (e.g. Blaauw 1961; Portegies Zwart 2000). Both mechanisms have been shown to occur in our Galaxy (Hoogerwerf et al. 2001). Typical velocities attained by the two formation channels are of the order of several hundreds of km s^{-1} (Portegies Zwart 2000; Przybilla et al. 2008; Gvaramadze et al. 2009; Gvaramadze & Gualandris 2011; Silva & Napiwotzki 2011), but simulations indicate that the majority of runaway stars from dynamical encounters have ejection velocities $\lesssim 200$ km s^{-1} (Perets & Šubr 2012). Recent results show that it is possible to achieve ejection velocities up to ~ 1300 km s^{-1} for low-mass G/K type stars in very compact binaries (Tauris 2015). Nevertheless, the rate of production of unbound RSs, referred to as *hyper runaway*

⋆ E-mail: marchetti@strw.leidenuniv.nl

stars (HRSs), is estimated to be as low as $8 \cdot 10^{-7}$ yr^{-1} (Perets & Šubr 2012; Brown 2015).

As a class, the fastest stars in our Galaxy are expected to be hypervelocity stars (HVSs). These have been first theoretically predicted by Hills (1988) as the result of a three-body interaction between a binary star and the massive black hole in the Galactic Centre (GC), Sagittarius A*. Following this close encounter, a star can be ejected with a velocity $\sim$ 1000 km s^{-1} , sufficiently high to escape from the gravitational field of the Milky Way (Kenyon et al. 2008; Brown 2015). The first HVS candidate was discovered by Brown et al. (2005): a B-type star with a velocity more than twice the Galactic escape speed at its position. Currently about $\sim$ 20 unbound HVSs with velocities $\sim$ 300 - 700 km s^{-1} have been discovered targeting young stars in the outer halo of the Milky Way (Brown et al. 2014). In addition, tens of mostly bound candidates have been found at smaller distance range but uncertainties prevent the precise identification of the GC as their ejection location (e.g. Hawkins et al. 2015; Vickers et al. 2015; Zhang et al. 2016; Marchetti et al. 2017; Ziegerer et al. 2017). HVSs are predicted to be ejected from the GC with an unknown rate around 10^{-4} yr^{-1} (Yu & Tremaine 2003; Zhang et al. 2013), two orders of magnitude larger than the rate of ejection of runaway stars with comparable velocities from the stellar disc (Brown 2015). Because of their extremely high velocities, HVS trajectories span a large range of distances, from the GC to the outer halo. Thus HVSs have been proposed as tools to study the matter distribution in our Galaxy (e.g. Gnedin et al. 2005; Sesana et al. 2007; Kenyon et al. 2014; Rossi et al. 2017; Fragione & Loeb 2017, Contigiani et al., in preparation) and the GC environment (e.g. Zhang et al. 2013; Madigan et al. 2014), but a larger and less observationally biased sample is needed in order to break degeneracies between the GC binary content and the Galactic potential parameters (Rossi et al. 2017). Using the fact that their angular momentum should be very close to zero, HVSs have also been proposed as tools to constrain the Solar position and velocity (Hattori et al. 2018). Other possible alternative mechanisms leading to the acceleration of HVSs are the encounter between a single star and a massive black hole binary in the GC (e.g. Yu & Tremaine 2003; Sesana et al. 2006, 2008), the interaction between a globular cluster with a single or a binary massive black hole in the GC (Capuzzo-Dolcetta & Fragione 2015; Fragione & Capuzzo-Dolcetta 2016), and the tidal interaction of a dwarf galaxy near the center of the Galaxy (Abadi et al. 2009). Another possible ejection origin for HVSs and high velocity stars in our Galaxy is the Large Magellanic Cloud (LMC, Boubert & Evans 2016; Boubert et al. 2017; Erkal et al. 2018), orbiting the Milky Way with a velocity $\sim$ 380 km s^{-1} (van der Marel & Kallivayalil 2014).

In addition to the unbound population of HVSs, all the ejection mechanisms mentioned above predict also a population of *bound* HVSs (BHVSs): stars sharing the same formation scenario as HVSs, but with an ejection velocity which is not sufficiently high to escape from the whole Milky Way (e.g. Bromley et al. 2006). Most of the deceleration occurs in the inner few kpc due to the bulge potential (Kenyon et al. 2008), and the minimum velocity necessary at ejection to be unbound is of the order of $\sim$ 800 km s^{-1} (a precise value depends on the choice of the Galactic potential, Brown 2015; Rossi et al. 2017). If we consider the Hills mechanism , this population of bound stars is expected to be dominant over the sample of HVSs (Rossi et al. 2014; Marchetti et al. 2018).

At the moment, the fastest star discovered in our Galaxy is US 708, traveling away from the Milky Way with a total velocity $\sim$ 1200 km s^{-1} (Hirsch et al. 2005). Its orbit is not consistent with coming from the GC (Brown et al. 2015), and the most likely mechanism responsible for its acceleration is the explosion of a thermonuclear supernova in an ultra-compact binary in the Galactic disk (Geier et al. 2015).

The second data release (DR2) of the European Space Agency satellite *Gaia* (Gaia Collaboration et al. 2016, 2018a) gives us the first opportunity to look for extremely high velocity stars in our Milky Way, using an unprecedented sample of precisely and accurately measured sources. On 2018 April 25, *Gaia* provided positions (α, δ), parallaxes ϖ and proper motions $(\mu_{\alpha*}, \mu_\delta)$ for more than 1.3 billion of stars, and, notably, radial velocities $v_{\rm rad}$ for a subset of 7224631 stars brighter than the 12th magnitude in the *Gaia* Radial Velocity Spectrograph (RVS) passband (Cropper et al. 2018; Katz et al. 2018). Radial velocities are included in the *Gaia* catalogue for stars with an effective temperature $T_{\rm eff}$ from 3550 to 6990 K, and have typical uncertainties of the order of few hundreds of m s^{-1} at the bright end of the magnitude distribution (*Gaia G* band magnitude $\approx$ 4), and of a few km s^{-1} at the faint end ($G \approx 13$).

Using *Gaia* DR2 data, Boubert et al. (2018) show that almost all the previously discovered late-type HVS candidates are most likely bound to the Galaxy, and their total velocity was previously overestimated because of inaccurate parallaxes and/or proper motions. Only one late-type star, LAMOST J115209.12+120258.0 (Li et al. 2015), is most likely unbound, but the Hills mechanisms is ruled out as a possible explanation of its extremely high velocity. The majority of B-type HVSs from (Brown et al. 2014, 2015) are still found to be consistent with coming from the GC when using *Gaia* DR2 proper motions (Erkal et al. 2018).

In this paper we search for the fastest stars in the Milky Way, within the sample of $\sim$ 7 million stars with a six-dimensional phase space measurement in *Gaia* DR2. Since the origin of high velocity stars in our Galaxy is still a puzzling open question, we simply construct the total velocity distribution in the Galactic rest-frame in order to identify and characterize the high velocity tail. In doing so, we do not bias our search towards any specific class of high velocity stars.

This manuscript is organized as follows. In Section 2, we explain how we determine distances and total velocities in the Galactic rest frame for the whole sample of stars. We presents results in terms of stellar total velocity in Section 3. In Section 4, we focus on the high velocity stars in the sample, and then in Section 5 we concentrate on the stars with a probability greater than 50% of being unbound from the Galaxy, discussing individually the most interesting candidates. Finally, we conclude and discuss our results and findings in Section 6.

2 DISTANCE AND TOTAL VELOCITY DETERMINATION

The *Gaia* catalogue provides parallaxes, and thus a conversion to a distance is required to convert the apparent motion of an object on the celestial sphere to a physical motion in space, that is needed to determine the total velocity of a star. Bailer-Jones (2015) discusses in details how this operation is not trivial when the relative error in parallax, $f \equiv \sigma_\varpi / \varpi$, is either above 20% or it is negative. We therefore separate the discussion on how we determine distances and total velocities of stars with $0 < f \leq 0.2$ (the "*low-f* sample") and of those with either $f > 0.2$ or $f < 0$ (the "*high-f* sample"). There are 7183262 stars with both radial velocity and the astrometric parameters (parallax and proper motions) in *Gaia* DR2, therefore in the following we will focus on this subsample of stars.

2.1 The "*low-f* Sample"

6376803 out of 7183262 stars ($\sim$ 89%) with radial velocity measurement in *Gaia* DR2 have a relative error in parallax $0 < f \leq 0.2$. For this majority of stars we can get an accurate determination of their distance just by inverting the parallax: $d = 1/\varpi$ (Bailer-Jones 2015). We then model the proper motions and parallax distribution as a multivariate Gaussian with mean vector:

$$\mathbf{m} = [\mu_{\alpha*}, \mu_\delta, \varpi] \tag{1}$$

and with covariance matrix:

$$\Sigma = \begin{pmatrix} \sigma^2_{\mu_{\alpha*}} & \sigma_{\mu_{\alpha*}}\sigma_{\mu_\delta}\rho(\mu_{\alpha*},\mu_\delta) & \sigma_{\mu_{\alpha*}}\sigma_\varpi\rho(\mu_{\alpha*},\varpi) \\ \sigma_{\mu_{\alpha*}}\sigma_{\mu_\delta}\rho(\mu_{\alpha*},\mu_\delta) & \sigma^2_{\mu_\delta} & \sigma_{\mu_\delta}\sigma_\varpi\rho(\mu_\delta,\varpi) \\ \sigma_{\mu_{\alpha*}}\sigma_\varpi\rho(\mu_{\alpha*},\varpi) & \sigma_{\mu_\delta}\sigma_\varpi\rho(\mu_\delta,\mu_\varpi) & \sigma^2_\varpi \end{pmatrix}, \tag{2}$$

where $\rho(i, j)$ denotes the correlation coefficient between the astrometric parameters i and j, and it is provided in the *Gaia* DR2 catalogue. Radial velocities are uncorrelated to the astrometric parameters, and we assume them to follow a Gaussian distribution centered on $v_{\rm rad}$, and with standard deviation $\sigma_{v_{\rm rad}}$. We then draw 500 Monte Carlo (MC) realizations of each star's observed astrometric parameters, and we simply compute distances by inverting parallaxes.

Total velocities in the Galactic rest frame are computed correcting radial velocities and proper motions for the solar and the local standard of rest (LSR) motion (Schönrich 2012). In doing so, we assume that the distance between the Sun and the GC is $d_\odot = 8.2$ kpc, and that the Sun has an height above the stellar disk of $z_\odot = 25$ pc (Bland-Hawthorn & Gerhard 2016). We assume a rotation velocity at the Sun position $v_{\rm LSR} = 238$ km s^{-1} and a Sun's orbital velocity vector $\mathbf{v}_\odot = [U_\odot, V_\odot, W_\odot] = [14.0, 12.24, 7.25]$ km s^{-1} (Schönrich et al. 2010; Schönrich 2012; Bland-Hawthorn & Gerhard 2016). We then derive Galactic rectangular velocities (U, V, W) adopting the following convention: U is positive when pointing in the direction of the GC, V is positive along the direction of the Sun rotation around the Galaxy, and W is positive when pointing towards the North Galactic Pole (Johnson & Soderblom 1987). Starting from the MC samples on proper motions, distances, and radial velocities, we then compute total velocities in the Galactic rest frame $v_{\rm GC} = v_{\rm GC}(\alpha, \delta, \mu_{\alpha*}, \mu_\delta, d, v_{\rm rad})$ summing in quadrature the three velocity components (U, V, W).

Finally, for each star we compute the probability $P_{\rm ub}$ of being unbound from the Galaxy as the fraction of MC realizations which result in a total velocity $v_{\rm GC}$ greater than the escape speed from the MW at that given position. We make use of the median escape speed from Williams et al. (2017), which is measured across a range of $\sim$ 40 kpc in Galactocentric distances using a variety of tracers extracted from the Sloan Digital Sky Survey (SDSS, Ahn et al. 2012).

2.2 The "*high-f* Sample"

A more careful analysis is required for 806459 stars with either $f > 0.2$ or with a negative measured parallax. For these stars, we follow the approach outlined in Bailer-Jones (2015); Astraatmadja & Bailer-Jones (2016a,b); Luri et al. (2018). We use a full Bayesian analysis to determine the posterior probability $P(d|\varpi, \sigma_\varpi)$ of observing a star at a distance d, given the measured parallax ϖ and its Gaussian uncertainty σ_ϖ. The authors show how the choice of the prior probability on distance $P(d)$ can seriously affect the shape of the posterior distribution, and therefore lead to significantly different values for the total velocity of a star. We decide to adopt an *exponentially decreasing prior*:

$$P(d) \propto d^2 \exp\left(-\frac{d}{L}\right), \tag{3}$$

which has been shown to perform best for stars further out than $\sim$ 2 kpc (Astraatmadja & Bailer-Jones 2016b), that is the expected distance of stars with a large relative error on parallax (see Appendix A). The value of the scale length parameter L is fixed to 2600 pc, and we refer the reader to the discussion in Appendix A for the reasons behind our choice of this particular value. By means of Bayes' theorem we can then express the posterior distribution on distances as:

$$P(d|\varpi, \sigma_\varpi) \propto P(\varpi|d, \sigma_\varpi)P(d), \tag{4}$$

where the likelihood probability $P(\varpi|d, \sigma_\varpi)$ is a Gaussian distribution centered on $1/d$:

$$P(\varpi|d, \sigma_\varpi) \propto \exp\left[-\frac{1}{2\sigma^2_\varpi}\left(\varpi - \frac{1}{d}\right)\right]. \tag{5}$$

In our case, we decide to fully include the covariance matrix between the astrometric properties, following the approach introduced in Marchetti et al. (2017). In this case, for each star the likelihood probability is a three dimensional multivariate Gaussian distribution with mean vector:

$$\mathbf{m} = [\mu_{\alpha*}, \mu_\delta, 1/d] \tag{6}$$

and covariance matrix given by equation (2). The prior distribution on distance is given by equation (3), and we assume uniform priors on proper motions. We then draw proper motions and distances from the resulting posterior distribution using the affine invariant ensemble Markov chain Monte Carlo (MCMC) sampler EMCEE (Goodman & Weare 2010; Foreman-Mackey et al. 2013). We run each chain using 32 walkers and 100 steps, for a total of 3200 random samples drawn from the posterior distribution. We initialize the walkers to random positions around the mean value of the proper motions and of the inverse of the mode of the posterior distribution in distance, equation (4), to achieve a fast convergence of the chain. We then directly use this MC sampling to derive a distribution for the total velocity in the Galactic rest frame of each star, assuming the same parameters for the Sun presented in Section 2.1. We check for the mean acceptance fraction as a test for the convergence of each MC chain.

3 THE TOTAL VELOCITY DISTRIBUTION OF STARS IN *Gaia* DR2

Using the approach discussed in Section 2, we publish a catalogue with distances and velocities in the Galactocentric frame for all the 7183262 stars analyzed in this paper. This is publicly available at `http://home.strw.leidenuniv.nl/~marchetti/research.html`. A full description of the catalogue content can be found in Appendix B.

In order to filter out the more uncertain candidates, for which it would be difficult to constrain the origin, we will now only discuss and plot results for stars with a relative error on total velocity $\sigma_{v_{\rm GC}}/v_{\rm GC} < 0.2$, where $\sigma_{v_{\rm GC}}$ is estimated summing in quadrature the lower and upper uncertainty on $v_{\rm GC}$. This cut results into a total of 6763506 stars, $\sim$ 94% of the original sample of stars. Figure 1 shows the total velocity distribution of the median Galactic rest frame total velocity $v_{\rm GC}$ for the original sample of 7183262 stars

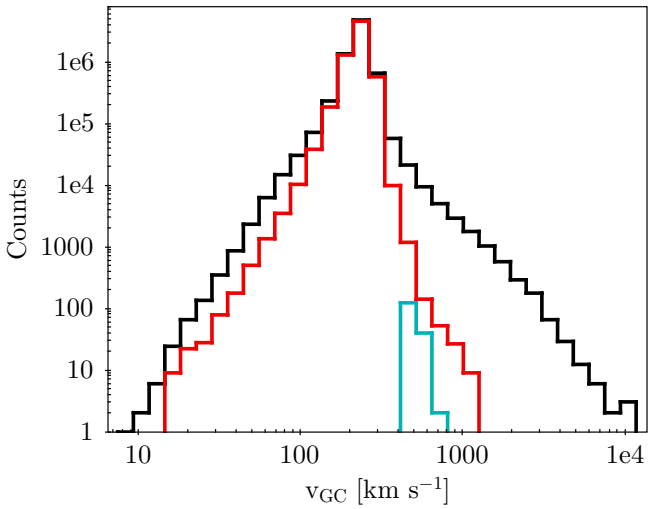

Figure 1. Histogram of median total velocities in the Galactic rest frame for all the ∼ 7 million stars with three-dimensional velocity by *Gaia* DR2 (black). The red line corresponds to those stars with a relative error on total velocity in the Galactic rest-frame below 20%, while the blue line refers to our "clean" sample of high velocity stars (see discussion in Section 4).

(black line) and for the stars with a relative error on total velocity below 20% (red line). We can see how this cut filters out most of the stars with extremely high velocities, which are likely to be outliers with unreliable measurements by *Gaia* . Nevertheless we note the presence of a high velocity tail extending up to ∼ 1000 km s^{-1} surviving the cut. We will now focus only on stars with $\sigma_{v_{GC}}/v_{GC} < 0.2$.

To highlight visually possibly unbound objects, we plot in Figure 2 the total velocity for all stars as a function of the Galactocentric distance r_{GC}, and we overplot the median escape speed from the Galaxy with a green dashed line (Williams et al. 2017). Datapoints correspond to the median of the distributions, with lower and upper uncertainties derived, respectively, from the 16th and 84th percentiles. Most of the stars are located in the solar neighborhood, and have typical velocities of the order of the LSR velocity. We find 173 stars to have probabilities greater than 50% of being unbound from the Galaxy (but note the large errorbars). In particular, 131 (113) stars are more than 1-σ (3-σ) away from the Galactic escape speed.

Figure 3 shows the Toomre diagram for all the ∼ 7 million stars, a plot that is useful to distinguish stellar populations based on their kinematics. On the x-axis we plot the component V of the Galactocentric Cartesian velocity, and on the y-axis the component orthogonal to it, $\sqrt{U^2 + W^2}$. Not surprisingly, most of the stars behave kinematically as disk stars on rotation-supported orbits, with V values around the Sun's orbital velocity (see Gaia Collaboration et al. 2018b). A sub-dominant, more diffuse, population of stars with halo-like kinematics is also present, centered around $V = 0$ and with a larger spread in total velocity.

4 HIGH VELOCITY STARS IN *Gaia* DR2

We now focus our interest towards high velocity stars, which we define as stars with a median $v_{GC} > 450$ km s^{-1} . Since we are interested in kinematic outliers, we have to pay particular attention not to be contaminated by data processing artifacts and spurious measurements. We therefore choose to adopt the following conservative cuts on the columns of the *Gaia* DR2 GAIA_SOURCE catalogue (in addition to the selection $\sigma_{v_{GC}}/v_{GC} < 0.2$ introduced in Section 3):

(i) ASTROMETRIC_GOF_AL < 3;
(ii) ASTROMETRIC_EXCESS_NOISE_SIG ≤ 2;
(iii) −0.23 ≤ MEAN_VARPI_FACTOR_AL ≤ 0.32;
(iv) VISIBILITY_PERIODS_USED > 8;
(v) RV_NB_TRANSITS > 5.

The first cut ensures that statistic astrometric model resulted in a good fit to the data, while the second cut selects only astrometrically well-behaved sources (refer to Lindegren et al. 2012, for a detailed explanation of the excess noise and its significance). The third and the fourth cuts are useful to exclude stars with parallaxes more vulnerable to errors. Finally, the final selection ensures that each source was observed at least for a reasonable amount of times (5) by *Gaia* to determine its radial velocity. Details on the parameters used to filter out possible contaminants can be found in the *Gaia* data model[1]. Applying these cuts and with the further constrain on the median $v_{GC} > 450$ km s^{-1} , we are left with a clean final sample of 165 high velocity stars. We also verify that the quality cuts C.1 and C.2 introduced in Appendix C of Lindegren et al. (2018), designed to select astrometrically clean subsets of objects, are already verified by our sample of high velocity stars. In addition, selection N in Appendix C of Lindegren et al. (2018) does not select any of our candidates. Looking at Fig. 2 we can see how these cuts filter out most of the stars with exceptionally high velocities, which are therefore likely to be instrumental artifacts.

The spatial distribution of these 165 high velocity stars in our Galaxy is shown in Fig. 4, where we overplot the position on the Galactic plane of this subset of stars with a blue colormap above the underlying distribution of the ∼ 7 million stars used in this paper. We can see how the majority of high velocity stars lies in the inner region of the Galaxy, with typical distances $\lesssim 10$ kpc from the GC. This is due to two main factors. First, stars in the inner regions of the Milky Way have higher dispersion velocities as a result of the higher value of the Galactic escape speed. Secondly, most of these stars are on the faint end of the magnitude distribution because of extinction due to dust in the direction of the GC, and thus they have large relative errors on parallax. This in turn translates into larger uncertainties on total velocity, which may cause the stars to be included into our high velocity cut. Another small overdensity lies in correspondence of the Sun position, correlating with the underlying distribution of all the stars. In Fig. 5, we plot the same but in the (x_{GC}, z_{GC}) plane. Most of our high velocity stars lie away from the stellar disc. Fig. 6 shows the position in Galactocentric cylindrical coordinates of the high velocity star candidates only. Most of them are concentrated in the inner region of the Galaxy, but a diffuse population identifiable with the inner stellar halo is observed at large Galactic latitudes. Arrows are proportional to the total velocity of each star in the Galactic rest-frame.

Fig. 7 shows the Hertzsprung-Russell (HR) diagram for all the sources with a radial velocity measurement, with the high velocity star sample overplotted in blue. On the x-axis we plot the color index in the *Gaia* Blue Pass (BP) and Red Pass (RP) bands $G_{BP} - G_{RP}$, while on the y-axis we plot the absolute magnitude in

[1] https://gea.esac.esa.int/archive/documentation/GDR2/Gaia_archive/chap_datamodel/

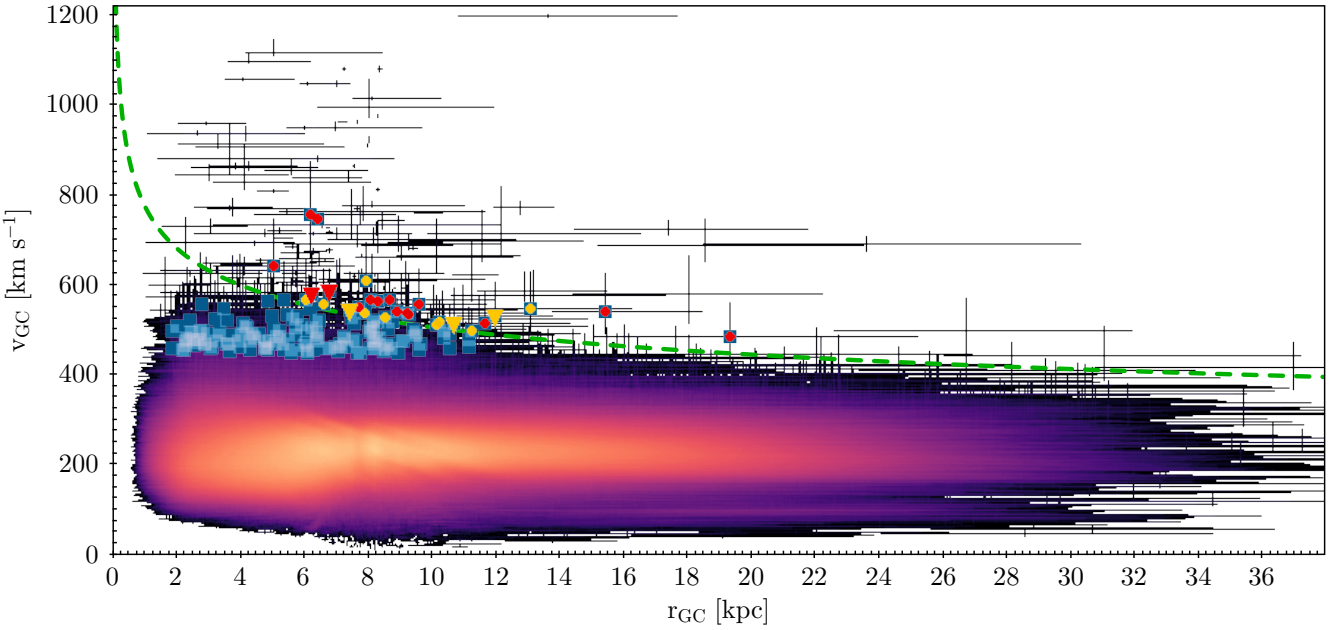

Figure 2. Total velocity in the Galactic rest-frame v_{GC} as a function of Galactocentric distance r_{GC} for all the 6869707 stars in *Gaia* DR2 with relative error on total velocity < 0.2. Colour is proportional to the logarithmic number density of stars. The green dashed line is the median posterior escape speed from the Galaxy from (Williams et al. 2017). We overplot in blue the "clean" high velocity star sample introduced in Section 4. Circles and triangles correspond, respectively, to HRS and HVS candidates discussed in Section 5, colored in yellow (red) if $P_{MW} > 0.5$ ($P_{MW} < 0.5$).

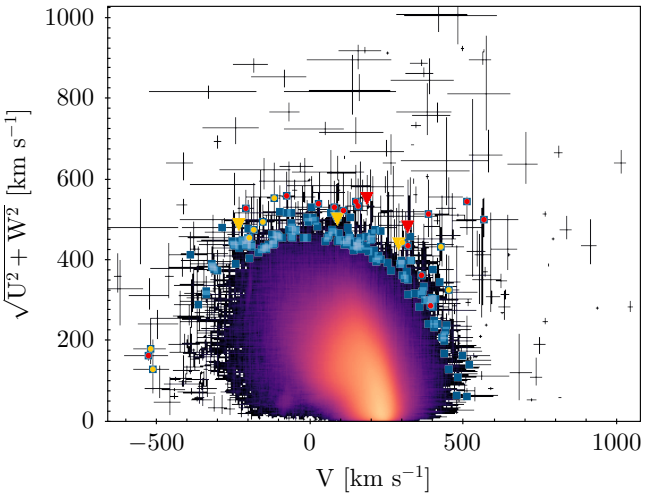

Figure 3. Toomre diagram for the same stars plotted in Fig. 2.

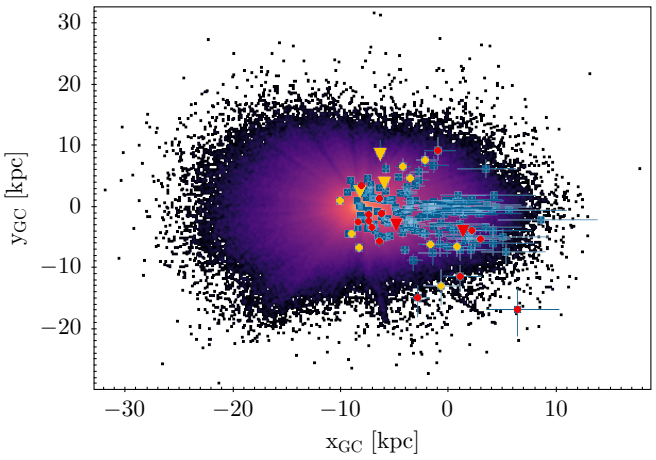

Figure 4. Distribution of the ∼ 7 million stars on the Galactic plane. The Sun is located at $(x_{GC}, y_{GC}) = (-8.2, 0)$ kpc. Colours are the same as in Fig. 2.

the *Gaia G* band M_G, computed assuming the median of the posterior distance distribution. We can see that the great majority of our stars are giants stars. Note that we did not consider extinction to construct the HR diagram, because of the caveats with using the line-of-sight extinction in the G band A_G for individual sources (Andrae et al. 2018).

4.1 Orbital Integration

In order to get hints on the ejection location of our sample of high velocity stars, we perform numerical orbit integration of their trajectories back in time using the python package GALA (Price-Whelan 2017). For each star we use 1000 random samples from the proper motions, distance, and radial velocity MC sampling discussed in Section 2. We integrate each orbit back in time for a total time of 1 Gyr, with a fixed time-step of 0.1 Myr, using the GALA potential *MilkyWayPotential*. This is a four components Galactic potential model consisting of a Hernquist bulge and nucleus (Hernquist 1990), a Miyamoto-Nagai disk (Miyamoto & Nagai 1975), and a Navarro-Frenk-White halo (Navarro et al. 1996). The parameters are chosen to fit the enclosed mass profile of the Milky Way (Bovy 2015). We then derive the pericenter distance and, for bound MC realizations, the apocenter distance and the eccentricity of the

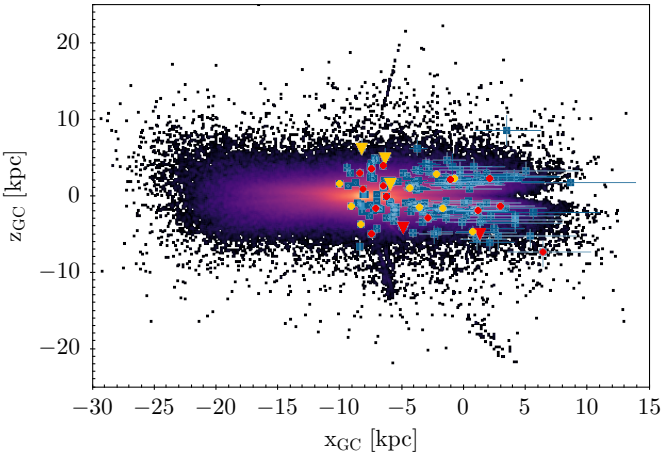

Figure 5. Same as Fig. 4, but showing the distribution of the stars in the (x_{GC}, z_{GC}) plane. The Sun is located at $(x_{GC}, z_{GC}) = (-8200, 25)$ pc. Colors are the same as in Fig. 2.

orbit. We also record the energy and the angular momentum of each MC orbit. We check for energy conservation as a test of the accuracy of the numerical integration.

In Fig. 8, we plot the maximum height above the Galactic disk Z_{max} as a function of the eccentricity of the orbit for our sample of high velocity stars. This plot is useful to identify similar stars based on their orbits (e.g. Boeche et al. 2013; Hawkins et al. 2015). The dashed red line at $Z_{max} = 3$ kpc denotes the typical scale height of the thick disk (Carollo et al. 2010). Not surprisingly, high velocity stars are on highly eccentric orbits, with a mean eccentricity of the sample ~ 0.9. Most of these stars span a large range of Z_{max}, with values up to hundreds of kpc, reflecting the large amplitude of the vertical oscillations.

In our search for HVSs, we keep track of each disk crossing (Cartesian Galactocentric coordinate $z_{GC} = 0$) in the orbital traceback of our high velocity star sample. For each MC realization, we then define the crossing radius r_c as:

$$r_c = \sqrt{x_c^2 + y_c^2}, \tag{7}$$

where x_c and y_c are the Galactocentric coordinates of the orbit (x_{GC}, y_{GC}) at the instant when $z_{GC} = 0$. In the case of multiple disk crossings during the orbital trace-back, we define r_{min} as the minimum crossing radius attained in that particular MC realization of the star's orbit. This approach allows us to check for the consistency of the GC origin hypothesis for our sample of high velocity stars. We also record the ejection velocity v_{ej}: the velocity of the star at the minimum crossing radius, and the flight time t_f: the time needed to travel from the observed position to the disc crossing happening closest to the GC.

In Fig. 9, we plot r_{min} as a function of the orbital energy E. The red dashed line coincides with the separation region between bound and unbound orbits[2]. We can see how the majority of stars are traveling on bound orbits ($E < 0$), but we can see a few stars with remarkably high values of the energy.

5 UNBOUND STARS: HYPERVELOCITY AND HYPER RUNAWAY STAR CANDIDATES

We now focus our search to possible unbound stars, defined as the subsample of clean high velocity stars with $P_{ub} > 50\%$. This amounts to a total of 28 objects. Observed properties from *Gaia* DR2, distances, and total velocities for these stars are summarized in Table 1.

If a star on an unbound orbit was ejected either from the stellar disk (HRS) or from the GC (HVS), then its distribution of minimum crossing radii r_{min} should fall within the edge of the Milky Way disk. To maximize the probability of a disk crossing during the orbital traceback, we integrate the orbits of these stars for a maximum time of 5 Gyr. We then define the probability P_{MW} for a star to come from the Milky Way as the fraction of MC realizations resulting in $r_{min} < 16$ kpc. This probability is useful to flag candidates of possible extragalactic origin, which we define as those stars with $P_{MW} < 0.5$. This subset of 16 stars, if their high velocity is confirmed, could either originate as RS/HRS/HVS from the LMC (Boubert & Evans 2016; Boubert et al. 2017; Erkal et al. 2018), or could be the result of the tidal disruption of a dwarf galaxy interacting with the Milky Way (Abadi et al. 2009).

We then classify a star as a HVS (HRS) candidate if we cannot (can) exclude the hypothesis of GC origin, which we define by the condition $r_{min} - \sigma_{r_{min,l}} < 1$ kpc ($r_{min} - \sigma_{r_{min,l}} > 1$ kpc), where r_{min} denotes the median of the distribution, and $\sigma_{r_{min,l}}$ is the lower uncertainty on the minimum crossing radius. In this way we are testing whether, within its errorbars, a star is consistent with coming from the central region of the Galaxy. Figure 10 shows the histogram of the median minimum disk crossing r_{min} minus the lower uncertainty $\sigma_{r_{min,l}}$ for all the 28 stars with $P_{ub} > 0.5$. A vertical red dashed line corresponds to the value 1 kpc, which we use to define HVS candidates.

5.1 Hypervelocity Star Candidates

According to our classification criterion, there are 5 stars classified as HVS candidates (triangles in Fig. 2 and following plots). Their properties are summarized in Table 1. Typical velocities for these stars are above 500 km s^{-1} , with a maximum value $\sim$ 590 km s^{-1} . We note that 2 of these stars have low probabilities of being ejected from the Milky Way, because the great majority of the MC realizations resulted in no disk crossings. A further careful analysis is needed in order to identify their ejection location, since the Hills mechanism most likely is not the one responsible for their high velocity.

The number of HVS candidates and their spatial distribution are consistent with predictions from Marchetti et al. (2018). The majority of these stars are concentrated in the central region of the Milky Way, with marginally unbound velocities.

5.2 Hyper-Runaway Star Candidates

We find a total of 23 stars whose orbit, when integrated back in time, is not consistent with coming from the GC. These stars are HRS candidates (circles in Fig. 2 and following plots). 14 of these stars have probabilities $< 50\%$ of intersecting the Milky Way stellar disk when traced back in time, therefore an extra-Galactic origin is

[2] Rigorously, this definition for unbound stars is not consistent with the probability of being unbound P_{ub} introduced in Section 2.1. In fact the latter follows directly from the Galactic escape speed inferred from data, while the condition on the energy depends on the adopted Galactic potential model. Nevertheless, we check that the escape speed resulting from the chosen potential falls within the 68 per cent credible interval of the one in presented in Williams et al. (2017) for Galactocentric distances $\gtrsim 1$ kpc.

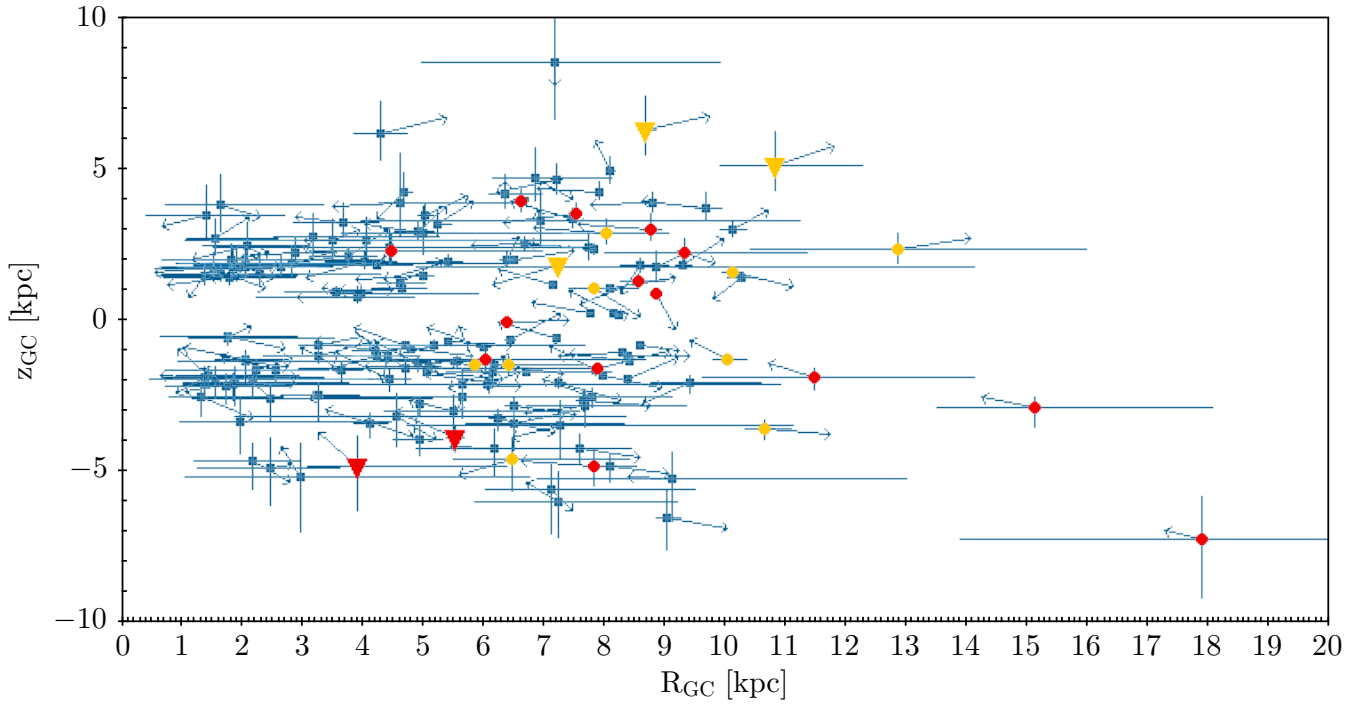

Figure 6. Position of the 438 high velocity stars in Galactocentric cylindrical coordinates (R_{GC}, z_{GC}). Arrows point to the direction of the velocity vector of the stars in this coordinate system, and the arrows' length is proportional to the total velocity of the star in the Galactic rest-frame. Colour is the same as in Fig. 2. The Sun is located at (R_{GC}, z_{GC}) = (8200, 25) pc.

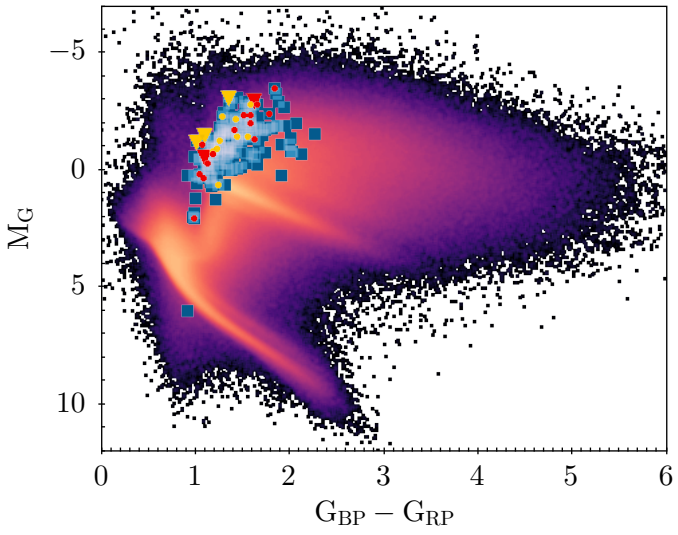

Figure 7. HR diagram for all the ∼ 7 million stars in *Gaia* DR2 with a radial velocity measurement. Colours are the same as in Fig. 2.

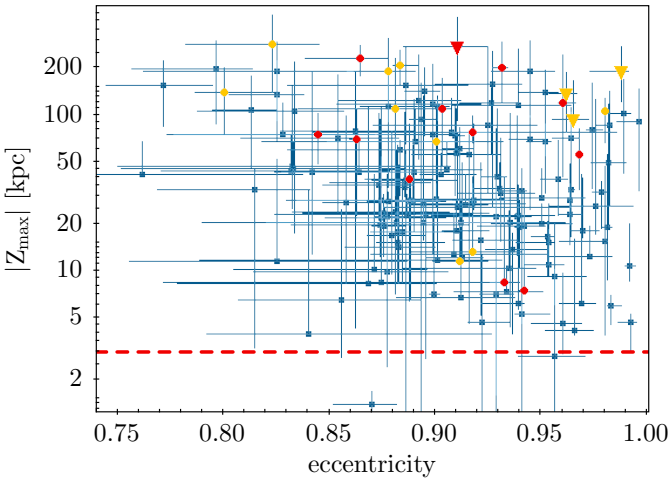

Figure 8. Absolute value of the maximum height above the Galactic plane $|Z_{max}|$ as a function of eccentricity for the high velocity sample of stars. The yellow horizontal dashed line corresponds to Z_{max} = 3 kpc, the edge of the thick disk (Carollo et al. 2010). Colours are the same as in Fig. 2.

preferred. The most likely ejection location could be the LMC, or otherwise spatial correlations with the density of surrounding stars could help identifying them as the high velocity tail of a stellar stream produced by the effect of the gravitational field of the Milky Way on a dwarf satellite galaxy.

Two particular HRS candidates that are worth mentioning are *Gaia* DR2 5932173855446728064 and *Gaia* DR2 5935868592404029184. The first star has an exceptionally well constrained total velocities, $v_{GC} = 747^{+2}_{-3}$ km s^{-1} , which results in a probability of being unbound ≈ 1. Notably, this star most likely does not originate in the Milky Way. The second source has a total velocity $v_{GC} = 755^{+118}_{-94}$, resulting in a probability $P_{ub} = 0.98$. We note that such exceptionally high velocities are thought to be very uncommon in our Galaxy for HRSs, which are predicted to be much rarer than HVSs (Brown 2015) .

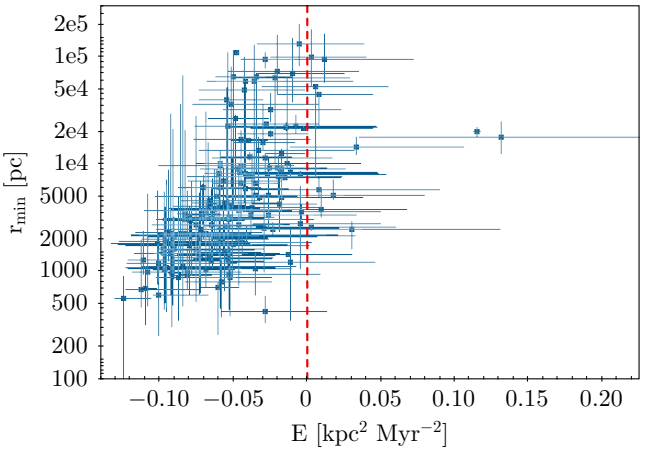

Figure 9. Minimum crossing radius $r_{\rm min}$ versus energy E for the 165 high velocity stars. The vertical dashed line separates unbound ($E > 0$) from bound ($E < 0$) orbits.

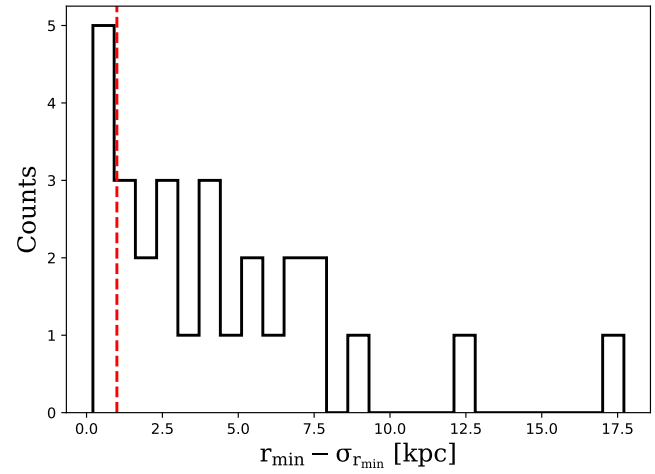

Figure 10. Histogram of the median minimum crossing radius $r_{\rm min}$ minus the correspondent lower uncertainty $\sigma_{r_{\rm min,l}}$ for the sample of 28 high velocity stars with $P_{\rm ub} > 0.5$. The vertical dashed line corresponds to $(r_{\rm min} - \sigma_{r_{\rm min,l}}) = 1$ kpc, our boundary condition for not rejecting the GC origin hypothesis for the HVS candidates (see discussion in Section 5).

6 CONCLUSIONS

We derived distance and total velocities for all the 7183262 stars with a full phase space measurement in the *Gaia* DR2 catalogue, in order to find unbound objects and velocity outliers. We defined our sample of high velocity stars as those stars with a median total velocity in the Galactic rest frame > 450 km s^{-1} , resulting in a total of 165 stars with reliable astrometric parameters and radial velocities. We traced back the high velocity stars in the Galactic potential to derive orbital parameters. Out of the 165 stars, we found the following.

(i) 28 stars have predicted probabilities $P_{\rm ub}$ higher than 50% of being unbound from the Milky Way. The observed and derived kinematic properties of these stars are summarized in Table 1.

(ii) 5 stars have orbits consistent with coming from the GC, and are therefore classified as HVS candidates. The typical velocities are between 500 km s^{-1} and 600 km s^{-1} .

(iii) 23 stars, when traced back in time in the Galactic potential, originate from the stellar disk of the Milky Way. These stars are HRS candidates.

(iv) There is an indication that 2 out of the 5 HVS candidates might not originate from the GC, since the fraction of orbits crossing the Milky disk near the massive black hole is lower than 40%. In addition, 14 out of the 23 HRS candidates also have probabilities $< 50\%$ to originate from the stellar disc of the Galaxy. This surprising and unexpected population of stars could be either produced as RSs/HRSs/HVSs from the LMC, thanks to its high orbital velocity around the Milky Way, or could be members of dwarf galaxies tidally disrupted by the gravitational interaction with the Galaxy. Further analyses are required in order to identify their origin.

This paper is just a first proof of the exciting discoveries that can be made mining the *Gaia* DR2 catalogue. We only limited our search to the $\sim$ 7 million stars with a full phase space information, a small catalogue compared to the full 1.3 billion sources with proper motions and parallaxes. Synergies with existing and upcoming ground-based spectroscopic surveys will be essential to obtain radial velocities and stellar spectra for subsets of these stars (e.g. Dalton 2016; de Jong et al. 2016; Kunder et al. 2017; Martell et al. 2017). For what concerns HVSs, Marchetti et al. (2018) shows how the majority of HVSs expected to be found in the *Gaia* catalogue are actually fainter than the limiting magnitude for radial velocities in DR2. We therefore did not expect to discover the bulk of the HVS population with the method outlined in this paper, but other data mining techniques need to be implemented in order to identify them among the dominant background of bound, low velocity stars (see for example Marchetti et al. 2017). We also show how particular attention needs to be paid to efficiently filter out contaminants and instrumental artifacts, which might mimic high velocity stars at a first inspection.

Table 1. Observed and derived properties for the 28 "clean" high velocity star candidates with a probability > 50% of being unbound from the Galaxy.

Gaia DR2 ID	(RA, Dec.) (°)	ϖ (mas)	$\mu_{\alpha*}$ (mas yr^{-1})	μ_δ (mas yr^{-1})	$v_{\rm rad}$ (km s^{-1})	G (mag)	d (pc)	$r_{\rm GC}$ (pc)	$v_{\rm GC}$ (km s^{-1})	$P_{\rm MW}$	$P_{\rm ub}$
HVS candidates											
6397497209236655872	(333.11342, −68.1686)	0.173 ± 0.017	−18.709 ± 0.022	−6.567 ± 0.026	−8.158 ± 3.613	13.21	5805^{+578}_{-491}	6777^{+115}_{-60}	587^{+49}_{-42}	0.24	0.85
1364548016594914560	(268.77922, 50.57305)	0.096 ± 0.019	−4.386 ± 0.035	7.819 ± 0.038	110.361 ± 0.436	11.93	10309^{+2321}_{-1682}	11951^{+1802}_{-1182}	531^{+77}_{-53}	1.00	0.74
6639557580310606976	(287.9278, −57.80599)	0.081 ± 0.033	−7.267 ± 0.058	2.431 ± 0.05	−257.081 ± 1.196	12.26	11288^{+3426}_{-2327}	6209^{+2749}_{-1349}	581^{+94}_{-57}	0.38	0.62
4593398670455374592	(274.89655, 33.81894)	0.201 ± 0.022	−1.185 ± 0.036	−25.738 ± 0.041	−312.966 ± 1.157	12.24	4982^{+587}_{-483}	7437^{+124}_{-69}	545^{+71}_{-57}	0.55	0.58
1508756353921427328	(210.20492, 45.86615)	0.148 ± 0.024	−15.547 ± 0.033	4.975 ± 0.035	128.221 ± 1.289	12.63	6795^{+1214}_{-889}	10693^{+817}_{-548}	515^{+78}_{-53}	0.99	0.58
HRS candidates											
5932173855446728064	(244.1181, −54.44045)	0.454 ± 0.029	−2.676 ± 0.043	−4.991 ± 0.034	−614.286 ± 2.492	13.81	2205^{+126}_{-134}	6391^{+103}_{-96}	747^{+2}_{-3}	0.02	1.00
5935868592404029184	(253.90291, −53.29868)	0.074 ± 0.021	5.47 ± 0.032	6.358 ± 0.026	308.412 ± 1.212	13.08	12373^{+3060}_{-2480}	6165^{+2667}_{-1768}	755^{+118}_{-94}	0.45	0.98
5244448023850619648	(153.90438, −69.19268)	0.054 ± 0.013	5.429 ± 0.024	−1.216 ± 0.022	39.876 ± 3.525	13.21	16038^{+3405}_{-1976}	15388^{+3024}_{-1672}	540^{+84}_{-49}	0.18	0.90
6431596947468407552	(274.68792, −70.24932)	0.084 ± 0.016	4.551 ± 0.019	4.974 ± 0.023	259.079 ± 1.653	13.09	11910^{+2778}_{-1779}	7936^{+2203}_{-1184}	607^{+86}_{-56}	0.74	0.87
5212817273334550016	(107.19916, −76.21933)	0.262 ± 0.021	12.169 ± 0.039	35.917 ± 0.045	159.866 ± 0.278	10.89	3826^{+294}_{-288}	8039^{+63}_{-51}	568^{+52}_{-51}	0.35	0.79
3705761936916676864	(192.7642, 4.94109)	0.266 ± 0.024	15.037 ± 0.053	−32.292 ± 0.032	88.655 ± 1.877	13.19	3736^{+397}_{-303}	8289^{+108}_{-70}	564^{+63}_{-47}	0.30	0.79
6077622510498751616	(187.9836, −53.91915)	0.054 ± 0.02	−6.685 ± 0.022	1.212 ± 0.022	623.609 ± 0.81	13.06	15014^{+3626}_{-3152}	13073^{+3161}_{-2502}	548^{+81}_{-59}	0.99	0.79
2233912206910720000	(299.2838, 55.49696)	0.278 ± 0.018	27.848 ± 0.031	−5.483 ± 0.034	−343.939 ± 1.714	12.97	3595^{+227}_{-236}	8895^{+91}_{-89}	540^{+30}_{-31}	0.42	0.76
4531575708618805376	(281.8599, 22.13939)	0.08 ± 0.018	4.073 ± 0.017	5.153 ± 0.031	−417.36 ± 0.895	13.06	11898^{+2687}_{-2049}	9587^{+2104}_{-1372}	558^{+79}_{-58}	0.34	0.73
5374177064347894272	(169.49883, −47.83129)	0.168 ± 0.025	7.244 ± 0.041	−17.28 ± 0.037	143.173 ± 0.453	12.19	5975^{+1013}_{-837}	8644^{+470}_{-315}	566^{+87}_{-72}	0.35	0.72
6010197124582216832	(239.83557, −37.75698)	0.084 ± 0.027	8.038 ± 0.054	−1.178 ± 0.034	294.687 ± 2.935	13.95	11267^{+3017}_{-2103}	5009^{+2534}_{-1310}	643^{+105}_{-71}	0.28	0.71
5779439836114210304	(234.18614, −78.30489)	0.014 ± 0.015	1.207 ± 0.022	2.469 ± 0.025	12.004 ± 1.049	13.36	23333^{+6165}_{-4603}	19311^{+5899}_{-4265}	485^{+74}_{-55}	0.33	0.69
4916199478888664320	(23.38253, −51.92318)	0.184 ± 0.023	−11.091 ± 0.034	−17.585 ± 0.036	86.865 ± 1.326	12.61	5457^{+691}_{-599}	9211^{+345}_{-267}	538^{+62}_{-53}	0.44	0.67
3784964943489710592	(169.3563, −5.81538)	0.259 ± 0.039	22.575 ± 0.077	−16.335 ± 0.048	126.155 ± 1.27	12.25	3883^{+711}_{-494}	9273^{+354}_{-222}	534^{+87}_{-59}	0.49	0.63
1268023196461923712	(225.78358, 26.24632)	0.224 ± 0.022	−29.641 ± 0.039	−18.877 ± 0.041	−276.839 ± 1.644	13.00	4482^{+507}_{-410}	7698^{+105}_{-61}	551^{+80}_{-63}	0.49	0.62
1042515801147259008	(129.79902, 62.50127)	0.389 ± 0.034	−33.08 ± 0.035	−41.029 ± 0.067	73.916 ± 1.144	12.72	2588^{+217}_{-219}	10228^{+182}_{-182}	518^{+50}_{-49}	0.56	0.60
5847044923481848192	(210.57, −69.36009)	0.062 ± 0.015	3.313 ± 0.022	3.851 ± 0.023	74.408 ± 1.095	13.53	14846^{+3075}_{-2370}	11628^{+2668}_{-1892}	513^{+71}_{-55}	0.49	0.60
4248140165233284352	(299.668, 4.51105)	0.147 ± 0.021	−17.336 ± 0.034	−0.186 ± 0.026	−358.1 ± 2.283	13.21	6725^{+1179}_{-772}	6017^{+314}_{-83}	567^{+89}_{-56}	0.50	0.56
5823263311601403392	(231.31938, −68.50629)	0.111 ± 0.026	7.405 ± 0.027	−4.895 ± 0.041	−78.245 ± 0.699	12.64	9078^{+2148}_{-1580}	6600^{+1279}_{-614}	557^{+78}_{-56}	1.00	0.55
5508788005188732416	(107.05511, −48.32651)	0.216 ± 0.022	22.734 ± 0.043	−13.23 ± 0.039	257.033 ± 2.069	12.55	4623^{+533}_{-442}	10124^{+331}_{-261}	510^{+66}_{-55}	0.62	0.54
4536536670713820288	(276.23382, 24.0701)	0.1 ± 0.015	−8.548 ± 0.017	−8.547 ± 0.026	−501.403 ± 2.331	13.60	10177^{+1649}_{-1364}	8505^{+1121}_{-783}	527^{+80}_{-62}	0.60	0.53
2036303544778830080	(288.22674, 26.8623)	0.131 ± 0.02	12.347 ± 0.027	−0.824 ± 0.035	−100.043 ± 1.191	13.02	7632^{+1459}_{-1081}	7883^{+744}_{-414}	536^{+83}_{-61}	0.54	0.54
5482348392671802624	(91.55732, −60.34047)	0.133 ± 0.012	3.182 ± 0.021	13.026 ± 0.023	434.118 ± 2.079	13.16	7609^{+790}_{-583}	11247^{+555}_{-391}	497^{+44}_{-33}	0.99	0.52

Note. Distances and total velocities are quoted in terms of the median of the distribution, with uncertainties derived from the 16th and 84th percentiles.

ACKNOWLEDGEMENTS

We thank E. Zari and the *Gaia* group meeting at Leiden Observatory for useful comments, suggestions and discussions during the preparation of this paper. TM and EMR acknowledge support from NWO TOP grant Module 2, project number 614.001.401. This project was developed in part at the 2017 Heidelberg *Gaia* Sprint, hosted by the Max-Planck-Institut für Astronomie, Heidelberg. This work has made use of data from the European Space Agency (ESA) mission *Gaia* (`https://www.cosmos.esa.int/gaia`), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, `https://www.cosmos.esa.int/web/gaia/dpac/consortium`). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. This research made use of ASTROPY, a community-developed core PYTHON package for Astronomy (Astropy Collaboration et al. 2013). All figures in the paper were produced using MATPLOTLIB (Hunter 2007) and TOPCAT (Taylor 2005). This work would not have been possible without the countless hours put in by members of the open-source community all around the world.

REFERENCES

Abadi M. G., Navarro J. F., Steinmetz M., 2009, ApJ, 691, L63
Ahn C. P., et al., 2012, ApJS, 203, 21
Andrae R., et al., 2018, preprint, (arXiv:1804.09374)
Astraatmadja T. L., Bailer-Jones C. A. L., 2016a, ApJ, 832, 137
Astraatmadja T. L., Bailer-Jones C. A. L., 2016b, ApJ, 833, 119
Astropy Collaboration et al., 2013, A&A, 558, A33
Bailer-Jones C. A. L., 2015, PASP, 127, 994
Blaauw A., 1961, Bull. Astron. Inst. Netherlands, 15, 265
Bland-Hawthorn J., Gerhard O., 2016, ARA&A, 54, 529
Boeche C., et al., 2013, A&A, 553, A19
Boubert D., Evans N. W., 2016, ApJ, 825, L6
Boubert D., Erkal D., Evans N. W., Izzard R. G., 2017, MNRAS, 469, 2151
Boubert D., Guillochon J., Hawkins K., Ginsburg I., Evans N. W., 2018, preprint, (arXiv:1804.10179)
Bovy J., 2015, ApJS, 216, 29
Bromley B. C., Kenyon S. J., Geller M. J., Barcikowski E., Brown W. R., Kurtz M. J., 2006, ApJ, 653, 1194
Brown W. R., 2015, ARA&A, 53, 15
Brown W. R., Geller M. J., Kenyon S. J., Kurtz M. J., 2005, ApJ, 622, L33
Brown W. R., Geller M. J., Kenyon S. J., 2014, ApJ, 787, 89
Brown W. R., Anderson J., Gnedin O. Y., Bond H. E., Geller M. J., Kenyon S. J., 2015, ApJ, 804, 49
Capuzzo-Dolcetta R., Fragione G., 2015, MNRAS, 454, 2677
Carollo D., et al., 2010, ApJ, 712, 692
Cropper M., et al., 2018, preprint, (arXiv:1804.09369)
Dalton G., 2016, in Skillen I., Balcells M., Trager S., eds, Astronomical Society of the Pacific Conference Series Vol. 507, Multi-Object Spectroscopy in the Next Decade: Big Questions, Large Surveys, and Wide Fields. p. 97
Erkal D., Boubert D., Gualandris A., Evans N. W., Antonini F., 2018, preprint, (arXiv:1804.10197)
Evans N. W., Sanders J. L., Williams A. A., An J., Lynden-Bell D., Dehnen W., 2016, MNRAS, 456, 4506
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Fragione G., Capuzzo-Dolcetta R., 2016, MNRAS, 458, 2596
Fragione G., Loeb A., 2017, New A, 55, 32
Gaia Collaboration et al., 2016, A&A, 595, A2
Gaia Collaboration et al., 2018b, preprint, (arXiv:1804.09380)
Gaia Collaboration Brown A. G. A., Vallenari A., Prusti T., de Bruijne J. H. J., Babusiaux C., Bailer-Jones C. A. L., 2018a, preprint, (arXiv:1804.09365)
Geier S., et al., 2015, Science, 347, 1126
Gnedin O. Y., Gould A., Miralda-Escudé J., Zentner A. R., 2005, ApJ, 634, 344
Gnedin O. Y., Brown W. R., Geller M. J., Kenyon S. J., 2010, ApJ, 720, L108
Goodman J., Weare J., 2010, Comm. App. Math. Comp. Sci., 5, 65
Gvaramadze V. V., Gualandris A., 2011, MNRAS, 410, 304
Gvaramadze V. V., Gualandris A., Portegies Zwart S., 2009, MNRAS, 396, 570
Hattori K., Valluri M., Castro N., 2018, preprint, (arXiv:1804.08590)
Hawkins K., et al., 2015, MNRAS, 447, 2046
Hernquist L., 1990, ApJ, 356, 359
Hills J. G., 1988, Nature, 331, 687
Hirsch H. A., Heber U., O'Toole S. J., Bresolin F., 2005, A&A, 444, L61
Hoogerwerf R., de Bruijne J. H. J., de Zeeuw P. T., 2001, A&A, 365, 49
Hunter J. D., 2007, Computing In Science & Engineering, 9, 90
Johnson D. R. H., Soderblom D. R., 1987, AJ, 93, 864
Katz D., et al., 2018, preprint, (arXiv:1804.09372)
Kenyon S. J., Bromley B. C., Geller M. J., Brown W. R., 2008, ApJ, 680, 312
Kenyon S. J., Bromley B. C., Brown W. R., Geller M. J., 2014, ApJ, 793, 122
Kunder A., et al., 2017, AJ, 153, 75
Leonard P. J. T., Duncan M. J., 1990, AJ, 99, 608
Li Y.-B., et al., 2015, Research in Astronomy and Astrophysics, 15, 1364
Lindegren L., Lammers U., Hobbs D., O'Mullane W., Bastian U., Hernández J., 2012, A&A, 538, A78
Lindegren L., et al., 2018, preprint, (arXiv:1804.09366)
Luri X., et al., 2014, A&A, 566, A119
Luri X., et al., 2018, preprint, (arXiv:1804.09376)
Madigan A.-M., Pfuhl O., Levin Y., Gillessen S., Genzel R., Perets H. B., 2014, ApJ, 784, 23
Marchetti T., Rossi E. M., Kordopatis G., Brown A. G. A., Rimoldi A., Starkenburg E., Youakim K., Ashley R., 2017, MNRAS, 470, 1388
Marchetti T., Contigiani O., Rossi E. M., Albert J. G., Brown A. G. A., Sesana A., 2018, MNRAS, 476, 4697
Martell S. L., et al., 2017, MNRAS, 465, 3203
Miyamoto M., Nagai R., 1975, PASJ, 27, 533
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Perets H. B., Šubr L., 2012, ApJ, 751, 133
Piffl T., et al., 2014, A&A, 562, A91
Portegies Zwart S. F., 2000, ApJ, 544, 437
Poveda A., Ruiz J., Allen C., 1967, Boletin de los Observatorios Tonantzintla y Tacubaya, 4, 86
Price-Whelan A. M., 2017, The Journal of Open Source Software, 2
Przybilla N., Fernanda Nieva M., Heber U., Butler K., 2008, ApJ, 684, L103

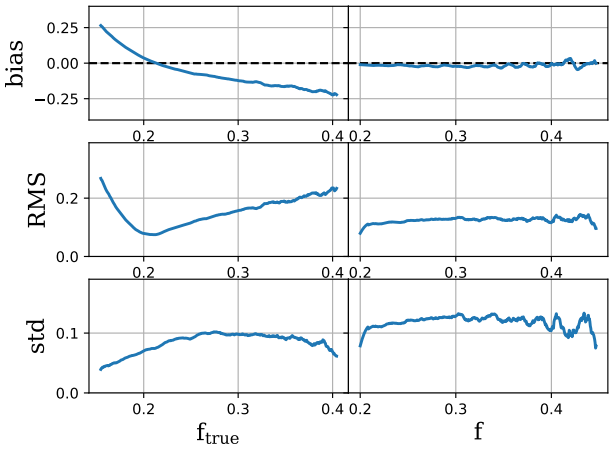

Figure A1. Bias, RMS, and standard deviation of the estimator x_0 as a function of $f_{\rm true} = \sigma_\varpi d_{\rm true}$ (left panel) and $f = \sigma_\varpi/\varpi$ (right panel). The modes of the posterior distributions are estimated using the exponentially decreasing prior with a characteristic scale length $L = 2600$ pc.

Robin A. C., et al., 2012, A&A, 543, A100
Rossi E. M., Kobayashi S., Sari R., 2014, ApJ, 795, 125
Rossi E. M., Marchetti T., Cacciato M., Kuiack M., Sari R., 2017, MNRAS, 467, 1844
Schönrich R., 2012, MNRAS, 427, 274
Schönrich R., Binney J., Dehnen W., 2010, MNRAS, 403, 1829
Sesana A., Haardt F., Madau P., 2006, ApJ, 651, 392
Sesana A., Haardt F., Madau P., 2007, MNRAS, 379, L45
Sesana A., Haardt F., Madau P., 2008, ApJ, 686, 432
Silva M. D. V., Napiwotzki R., 2011, MNRAS, 411, 2596
Smith M. C., et al., 2007, MNRAS, 379, 755
Smith M. C., et al., 2009, MNRAS, 399, 1223
Tauris T. M., 2015, MNRAS, 448, L6
Taylor M. B., 2005, in Shopbell P., Britton M., Ebert R., eds, Astronomical Society of the Pacific Conference Series Vol. 347, Astronomical Data Analysis Software and Systems XIV. p. 29
Vickers J. J., Smith M. C., Grebel E. K., 2015, AJ, 150, 77
Williams A. A., Belokurov V., Casey A. R., Evans N. W., 2017, MNRAS, 468, 2359
Yu Q., Tremaine S., 2003, ApJ, 599, 1129
Zhang F., Lu Y., Yu Q., 2013, ApJ, 768, 153
Zhang Y., Smith M. C., Carlin J. L., 2016, ApJ, 832, 10
Ziegerer E., Heber U., Geier S., Irrgang A., Kupfer T., Fürst F., Schaffenroth J., 2017, A&A, 601, A58
de Jong R. S., et al., 2016, in Ground-based and Airborne Instrumentation for Astronomy VI. p. 99081O, doi:10.1117/12.2232832
van der Marel R. P., Kallivayalil N., 2014, ApJ, 781, 121

APPENDIX A: CHOICE OF THE PRIOR PROBABILITY ON DISTANCES

In this appendix we discuss the choice of the prior probability on distances $P(d)$ which gives the most accurate results on the subsample of bright stars in *Gaia* DR2 with a large relative error on parallax (the *high-f* sample introduced in Section 2). We cross-match the *Gaia* Universe Model Snapshot (GUMS, Robin et al. 2012) and the *Gaia* Object Generator (GOG, Luri et al. 2014) catalogues based on the value of the source identifier, to get a resulting sample of $7 \cdot 10^6$ stars with $G_{\rm RVS} < 12.2$. We use the latest versions of these mock catalogues, GUMS-18 and GOG-18[3]. The resulting combined catalogue contains positions, parallaxes, proper motions, radial velocities, and distances for all stars, with corresponding uncertainties. We extend the limiting magnitude to $G_{\rm RVS} = 12.2$ to take into account the fact that *Gaia* does take spectra of some stars which are fainter than the limiting magnitude. In particular, these faint stars are the one with the largest error on parallax, so we want to be sure to include them, in order to derive accurate distances for the stars in *Gaia* DR2. We multiply the uncertainties on parallax and radial velocity by a factor $(60/22)^{0.5}$, and the ones on both proper motions by a factor $(60/22)^{1.5}$, to simulate the reduced performance of the *Gaia* satellite on 22 months of collected data.

We find 47001 of the 7 million stars to have $f = \sigma_\varpi/\varpi > 0.2$. We can see that this value is more than one order of magnitude smaller than the one found in *Gaia* DR2 (see Section 2.2 . All these stars are found at distance larger than $\sim$ 4.5 kpc from the Sun, and therefore we choose to adopt the exponentially decreasing prior to derive their distances (Astraatmadja & Bailer-Jones 2016b), see equation (3). The mode of the posterior distribution in equation (4) can be determined by numerically finding the roots of the implicit equation (Bailer-Jones 2015):

$$\frac{d^3}{L} - 2d^2 + \frac{\varpi}{\sigma_\varpi^2} d - \frac{1}{\sigma_\varpi^2} = 0. \tag{A1}$$

We compute the mode $d_{\rm Mo,i}$ for each star i in the simulated catalogue for different values of the scaling length L. We then determine the best fitting value of the parameter L as the one minimizing the quantity $\sum_i x_i^2$, where the scaled residual x_i is computed as (Astraatmadja & Bailer-Jones 2016a):

$$x_i = \frac{d_{\rm Mo,i} - d_{\rm true,i}}{d_{\rm true,i}}, \tag{A2}$$

where $d_{\rm true,i}$ denotes the *true* simulated distance of the i-th star. We find the value for the scale length $L = 2600$ pc to work best on this sample of $\sim$ 47000 simulated stars. In Fig. A1 we plot the mean value of the bias $\bar{x}$, the root mean squared (RMS) $\bar{x^2}^{1/2}$, and the standard deviation of the residual x for each bin of $f_{\rm true} = \sigma_\varpi d_{\rm true}$ (left panel) and f (right panel). We can see that, with this choice of prior, the mode of the posterior distribution on distances is an unbiased estimator for all the range of *observed* relative errors in parallax f, even if it shows a bias of $\sim 20\%$ for stars with low and high values of the *true* relative error $f_{\rm true}$.

APPENDIX B: CONTENT OF THE DISTANCE AND VELOCITY CATALOGUE

Table B1 provides an explanation of the content of the catalogue containing distances and velocities for the 7183262 stars with a radial velocity measurement in *Gaia* DR2. The catalogue is publicly available at `http://home.strw.leidenuniv.nl/~marchetti/research.html`.

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

[3] `https://wwwhip.obspm.fr/gaiasimu/`

Table B1. Catalogue description. Derived distances and velocities correspond to the median of the distribution, and lower and upper uncertainties are derived, respectively, from the 16th and 84th percentiles of the distribution function. Entries labelled [1] are derived in this paper, while entries labelled [2] are taken from the *Gaia* DR2 catalogue (Gaia Collaboration et al. 2018a).

Column	Units	Name	Description
1	-	source_id	*Gaia* DR2 identifier[2]
2	deg	ra	Right ascension[2]
3	deg	dec	Declination[2]
4	mas	parallax	Parallax[2]
5	mas	e_parallax	Standard uncertainty in parallax[2]
6	mas yr^{-1}	pmra	Proper motion in right ascension[2]
7	mas yr^{-1}	e_pmra	Standard uncertainty in proper motion in right ascension[2]
8	mas yr^{-1}	pmdec	Proper motion in declination[2]
9	mas yr^{-1}	e_pmdec	Standard uncertainty in proper motion declination[2]
10	km s^{-1}	vrad	Radial velocity[2]
11	km s^{-1}	e_vrad	Radial velocity error[2]
12	mag	GMag	G-band mean magnitude[2]
13	pc	dist	Distance estimate[1]
14	pc	el_dist	Lower uncertainty on distance[1]
15	pc	eu_dist	Upper uncertainty on distance[1]
16	pc	rGC	Spherical Galactocentric radius[1]
17	pc	el_rGC	Lower uncertainty on spherical Galactocentric radius[1]
18	pc	eu_rGC	Upper uncertainty on spherical Galactocentric radius[1]
19	pc	RGC	Cylindrical Galactocentric radius[1]
20	pc	el_RGC	Lower uncertainty on cylindrical Galactocentric radius[1]
21	pc	eu_RGC	Upper uncertainty on cylindrical Galactocentric radius[1]
22	pc	xGC	Cartesian Galactocentric x-coordinate[1]
23	pc	el_xGC	Lower uncertainty on Cartesian Galactocentric x-coordinate[1]
24	pc	eu_xGC	Upper uncertainty on Cartesian Galactocentric x-coordinate[1]
25	pc	yGC	Cartesian Galactocentric y-coordinate[1]
26	pc	el_yGC	Lower uncertainty on Cartesian Galactocentric y-coordinate[1]
27	pc	eu_yGC	Upper uncertainty on Cartesian Galactocentric y-coordinate[1]
28	pc	zGC	Cartesian Galactocentric z-coordinate[1]
29	pc	el_zGC	Lower uncertainty on Cartesian Galactocentric z-coordinate[1]
30	pc	eu_zGC	Upper uncertainty on Cartesian Galactocentric z-coordinate[1]
31	km s^{-1}	U	Cartesian Galactocentric x-velocity[1]
32	km s^{-1}	el_U	Lower uncertainty on Cartesian Galactocentric x-velocity[1]
33	km s^{-1}	eu_U	Upper uncertainty on Cartesian Galactocentric x-velocity[1]
34	km s^{-1}	V	Cartesian Galactocentric y-velocity[1]
35	km s^{-1}	el_V	Lower uncertainty on Cartesian Galactocentric y-velocity[1]
36	km s^{-1}	eu_V	Upper uncertainty on Cartesian Galactocentric y-velocity[1]
37	km s^{-1}	W	Cartesian Galactocentric z-velocity[1]
38	km s^{-1}	el_W	Lower uncertainty on Cartesian Galactocentric z-velocity[1]
39	km s^{-1}	eu_W	Upper uncertainty on Cartesian Galactocentric z-velocity[1]
40	km s^{-1}	UW	Cartesian Galactocentric xz-velocity[1]
41	km s^{-1}	el_UW	Lower uncertainty on Cartesian Galactocentric xz-velocity[1]
42	km s^{-1}	eu_UW	Upper uncertainty on Cartesian Galactocentric xz-velocity[1]
43	km s^{-1}	vR	Cylindrical Galactocentric R-velocity[1]
44	km s^{-1}	el_vR	Lower uncertainty on cylindrical Galactocentric R-velocity[1]
45	km s^{-1}	eu_vR	Upper uncertainty on cylindrical Galactocentric R-velocity[1]
46	km s^{-1}	vtot	Total velocity in the Galactic rest-frame[1]
47	km s^{-1}	el_vtot	Lower uncertainty on total velocity in the Galactic rest-frame[1]
48	km s^{-1}	eu_vtot	Upper uncertainty on total velocity in the Galactic rest-frame[1]
49	-	P_ub	Probability of being unbound from the Galaxy[1]