Hunting for the fastest stars in the Milky Way

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 10 oktober 2019
klokke 10:00 uur

door

Tommaso Marchetti
geboren te Rome, Italië
in 1991

Promotiecommissie

Promotor:	Prof. dr. K.H. Kuijken	
Co-promotor:	Dr. E.M. Rossi	
Overige leden:	Prof. dr. M. Franx	
	Prof. dr. H.J.A. Röttgering	
	Prof. dr. P.P. van der Werf	
	Prof. dr. P.T. de Zeeuw	
	Dr. W.R. Brown	(Harvard University)
	Dr. K. Hawkins	(University of Texas at Austin)
	Dr. T. Prusti	(European Space Agency)

ISBN: 978-94-6380-528-5

Cover design, illustration and coloring by Maria Cristina Fortuna and Tommaso Marchetti, freely inspired by the artwork of Aaron Turner for the album *In the Absence of Truth* by the band Isis

Printed by ProefschriftMaken `www.proefschriftmaken.nl`

Contents

1 Introduction **1**
1.1 High velocity stars . . . 3
1.1.1 Runaway stars . . . 4
1.1.2 Hypervelocity Stars . . . 5
1.2 The ESA mission *Gaia* . . . 11
1.2.1 The first *Gaia* data release . . . 11
1.2.2 The second *Gaia* data release . . . 12
1.2.3 Future *Gaia* data releases . . . 13
1.2.4 Warnings and caveats while using *Gaia* data . . . 14
1.2.5 *Gaia* and HVSs . . . 14
1.3 Methods used in this thesis . . . 15
1.3.1 Bayes' Theorem . . . 15
1.3.2 Machine Learning . . . 16
1.4 Thesis content . . . 18

2 Predicting the hypervelocity star population in *Gaia* **21**
2.1 Introduction . . . 22
2.2 The "Vesc" Mock Catalogue: A Simple Approach . . . 25
2.2.1 Astrometric Characterization of a HVS . . . 26
2.2.2 Photometric Characterization of a HVS . . . 27
2.2.3 *Gaia* Error Estimates . . . 29
2.2.4 Number Density of HVSs . . . 29
2.2.5 *"Vesc"* Catalogue: Number Estimates of HVSs in *Gaia* 31
2.3 The "Hills" Catalogue . . . 36
2.3.1 Velocity Distribution of HVSs . . . 36
2.3.2 Flight Time Distribution of HVSs . . . 37
2.3.3 Initial Conditions and Orbit Integration . . . 39
2.3.4 "Hills" Catalogue: Number Estimates of HVSs in *Gaia* 39

2.4 The "MBHB" Catalogue 45
2.4.1 Ejection of HVSs by the MBHB 45
2.4.2 "MBHB" Catalogue: Number Estimates of HVSs in *Gaia* 50
2.5 Prospects for the Current Sample of HVSs 52
2.6 Discussion and Conclusions 55

3 An artificial neural network to discover hypervelocity stars: candidates in *Gaia* DR1/TGAS 61
3.1 Introduction 62
3.2 Data Mining algorithm 64
3.2.1 Artificial Neural Networks 65
3.2.2 Building the Training Set 67
3.2.3 Optimization of the Algorithm 69
3.2.4 Performance of the Algorithm 70
3.3 Application to *Gaia* DR1 70
3.4 Acquiring spectral information 72
3.4.1 Catalogue cross-matching 72
3.4.2 Follow-up observations with the INT 72
3.5 Distance estimation 74
3.6 Results 79
3.6.1 Total Galactocentric velocity 82
3.6.2 Orbital traceback 84
3.7 Discussion of Individual Candidates 87
3.7.1 HVS and BHVS Candidates 87
3.7.2 Runaway Star Candidates 92
3.7.3 Uncertain Candidates 94
3.7.4 HD 5223: Most Likely Not a HVS 95
3.8 Discussion and Conclusions 95
.1 *Gaia* Identifiers 97
.2 Assuming Different Priors on Distance 97

4 *Gaia* DR2 in 6D: Searching for the fastest stars in the Galaxy101
4.1 Introduction 102
4.2 Distance and Total Velocity Determination 105
4.2.1 The "*low-f* Sample" 105
4.2.2 The "*high-f* Sample" 107
4.3 The Total Velocity Distribution of Stars in *Gaia* DR2 110
4.4 High Velocity Stars in *Gaia* DR2 111
4.4.1 Orbital Integration 114

4.5 Unbound Stars: Hypervelocity and Hyper Runaway Star Candidates . . . 118
4.5.1 Galactic Stars . . . 121
4.5.2 Extragalactic Stars . . . 122
4.6 Conclusions . . . 122
.1 Choice of the Prior Probability on Distances . . . 125
.2 Content of the Distance and Velocity Catalogue . . . 128
.3 List of High Velocity Stars with $0.5 < P_{\rm ub} \leqslant 0.8$. . . . 128
.4 Global Parallax Offset . . . 128
.5 Systematic Errors in Parallax . . . 134

5 Joint constraints on the Galactic dark matter halo and Galactic Centre from hypervelocity stars 139
5.1 Introduction . . . 140
5.2 Ejection velocity distributions . . . 143
5.3 Predicting velocity distributions in the halo: first approach. . 146
5.3.1 Velocity distribution in the halo: global description of the potential . . . 146
5.3.2 Comparison with data . . . 150
5.3.3 Results . . . 151
5.4 Second approach: assuming a Galactic Potential model . . . 154
5.4.1 The low-velocity tail . . . 156
5.4.2 Results . . . 160
5.4.3 Impact of different disc and bulge models . . . 162
5.5 Discussion and conclusions . . . 164
.1 Markov Chain Monte Carlo to fit the observed circular velocity 168

Bibliography 171

Nederlandse Samenvatting 181

English Summary 187

List of Publications 193

Curriculum Vitae 195

Acknowledgements 197

1 Introduction

"Quod tertio loco a nobis fuit observatum, est ipsiusmet LACTEI Circuli essentia, seu materies, quam Perspicilli beneficio adeo ad sensum licet intueri, ut et altercationes omnes, quæ per tot sæcula philosophos excruciarunt, ab oculata certitudine dirimantur, nosque a verbosis disputationibus liberemur. Est enim GALAXIA nihil aliud, quam innumerarum Stellarum coacervatim consitarum congeries: in quamcumque enim regionem illius Perspicillum dirigas, statim Stellarum ingens frequentia sese in conspectum profert, quarum complures satis magnæ ac valde conspicuæ videntur; sed exiguarum multitudo prorsus inexplorabilis est."
(Galileo Galilei, Sidereus Nuncius, 1610)

"What was observed by us in the third place is the nature or matter of the Milky Way itself, which, with the aid of the spyglass, may be observed so well that all the disputes that for so many generations have vexed philosophers are destroyed by visible certainty, and we are liberated from wordy arguments. For the Galaxy is nothing else than a congeries of innumerable stars distributed in clusters. To whatever region of it you direct your spyglass, an immense number of stars immediately offer themselves to view, of which very many appear rather large and very conspicuous but the multitude of small ones is truly unfathomable." [1]

It was the year 1610 when, using the telescope he constructed, Galileo Galilei first showed that the bright band on the sky whose origin and composition fascinated ancient cultures is a collection of multiple stars, whose majority cannot be resolved by the naked eye. This was the first step towards a modern scientific approach to the study of the Milky Way (MW), the Galaxy we are living in. Today, with the help of large ground- and space-based tele-

[1] English translation from Albert Van Helden, University of Chicago Press, Chicago, Illinois, 1989.

scopes, we have made huge steps forward to understand our Galaxy, but still, we are far from a comprehensive, complete and self-consistent picture, and many questions are still open. What are the accretion and evolutionary history of the MW? How do stars behave in the proximity of the central massive black hole (MBH)? What is the shape and extent of the dark matter halo? More than four hundred years after Galileo Galilei's breakthrough discovery, we are still looking up staring at the night sky, building new telescopes and satellites to better understand our Galaxy. In the light of these open questions, we present here our work on searching for the fastest objects in the MW: stars whose speed is so high that they are flying away from it on unbound trajectories. We show how these remarkable objects can help us decipher the Galaxy, giving us insights into its structure, its building components, and on some of its most energetic phenomena.

The MW is a barred spiral galaxy, and our Sun is only one of the hundreds of billions of stars orbiting inside it. Jan Oort in 1927 first discovered that the majority of these stars rotate coherently around the Galactic Centre (GC) in the shape of a flattened disk (Oort 1927). The principal stellar components of the Galaxy are the central box/peanut bulge, the stellar disk (composed of the thin and thick disks), and a diffuse stellar halo. The MW is embedded in a vast dark matter halo, which constitutes most of the MW mass, and extends up to hundreds of kpc from the GC (Bland-Hawthorn & Gerhard 2016).

Thanks to the exquisite quality of the recent imaging of the centre of the galaxy M87 with the Event Horizon Telescope (Event Horizon Telescope Collaboration et al. 2019), it has been definitely proven that MBHs exist at the centre of galaxies. In our MW, the location of the MBH coincides with the radio source Sagittarius A* (often abbreviated as Sgr A*, Balick & Brown 1974; Reid et al. 2009). The Sun is located at a distance of 8.127 kpc from the GC (Gravity Collaboration et al. 2018). Observations have shown the presence of several dozens of main sequence B-type stars orbiting around Sgr A*, the so-called S stars (Ghez et al. 2003). The orbits of these stars represent the best proof for the existence of our MBH (and provide tight constraints on the enclosed mass, Gillessen et al. 2009, 2017). S stars challenge our knowledge of how stars form in this extreme environment: the tidal forces of the MBH are predicted to be too strong to permit star formation within 1 arcsecond of the GC (Morris 1993).

In this introduction, we will discuss, among others, how high velocity stars can provide valuable information on the dynamics and origin of S stars, and how they can constrain global properties of the MW. This chapter

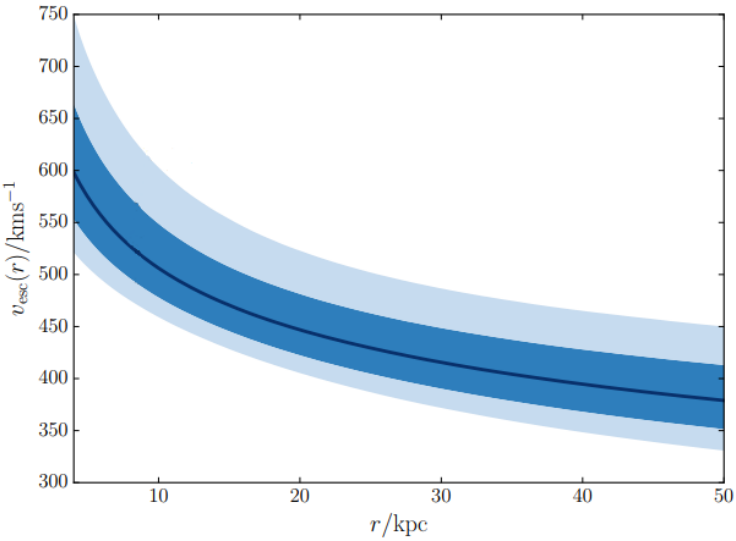


Figure 1.1: Escape speed from the Galaxy as a function of Galactocentric distance. Adapted from Williams et al. (2017).

is organized as follows. In Section 1.1 we introduce the two main classes of high velocity stars that will be studied in this thesis: runaway stars and hypervelocity stars. We will give a theoretical introduction to the acceleration mechanisms, and we will present the current status of the observations. Section 1.2 gives an overview of the European Space Agency (ESA) satellite *Gaia*, which has provided the largest stellar catalogue of the Galaxy ever produced. We use this dataset in three chapters of this thesis. In Section 1.3 we will cover the main methods used in this thesis. Finally, Section 1.4 provides an overview of the content of each of the following scientific chapters.

1.1 High velocity stars

Fast moving stars are intriguing for several reasons. The mechanisms leading to the acceleration of a star above its original velocity can give insights into multiple astrophysical processes, including but not limited to stellar and binary evolution, dynamics in the proximity of (massive) compact objects, and mergers between galaxies. In this Section we will introduce the main classes of high velocity stars. Typical velocities of stars can be com-

pared to the *escape speed* from the Galaxy, which defines the minimum velocity that a star needs to have in order to be *unbound* from the MW. Fig. 1.1 shows a recent result from Williams et al. (2017), showing the derived escape speed across a range of ~ 50 kpc from the GC, inferred using a variety of different kinematic tracers. The value at the Sun position is found to be 521^{+46}_{-30} km s^{-1}, falling to ~ 380 km s^{-1}at a Galactocentric distance of 50 kpc. In a more recent study, Monari et al. (2018) find a slightly higher value at the Sun position, 580 ± 63 km s^{-1}. A measurement of the escape speed can be converted into an estimate of the total mass of the MW (e.g. Smith et al. 2007; Piffl et al. 2014; Monari et al. 2018).

If we consider the encounter of two individual stars, the highest speed that can result is set by the escape velocity from their surface, since higher velocities would require the two stars to orbit at a distance smaller than their physical size (Leonard 1991):

$$v^*_{\rm esc} = \sqrt{\frac{2Gm_*}{r_*}} \simeq 618 \left(\frac{m_*}{M_\odot}\,\frac{R_\odot}{r_*}\right)^{1/2} \text{km /s}, \tag{1.1}$$

where G is the gravitational constant, m_* is the mass of the star, and r_* is its radius. Because of the approximately linear relation between m_* and r_* for stars on the main sequence, it follows that $v^*_{\rm esc} \simeq 600$ km s^{-1}in the mass range $m_* \in [0.4, 4]\ M_\odot$. Higher velocities can be achieved for compact objects such as white dwarfs and neutron stars. It turns out that equation (1.1) is an overestimate of the value of the escape velocity from a star: more precise calculations including binary evolution and mass transfer result into lower values of $v^*_{\rm esc}$.

1.1.1 Runaway stars

The term *runaway star* has first been coined by Blaauw (1961) to refer to the young, O and B-type stars observed out of the Galactic plane. Two main mechanisms have been introduced to predict the excess of velocity with respect to the Galaxy at their location. Blaauw (1961) proposed that runaway stars form as the result of a supernova explosion in a binary system. The more massive star in the binary evolves faster, transferring mass to the companion. When the donor explodes as a supernova, it can eject the companion star with a high velocity, forming a runaway star. The other proposed mechanism is dynamical encounters between stars in a dense stellar system (Poveda et al. 1967). In systems such as a young open cluster, interaction between binaries can lead to the ejection of one star from the cluster.

Tracing back the orbit of known runaway star candidates to their natal cluster, both these mechanisms have been observed to take place in the MW (Hoogerwerf et al. 2001). Maximum ejection velocities for both channels are typically $\lesssim 300 - 400$ km s^{-1}(e.g. Leonard & Duncan 1990; Portegies Zwart 2000; Przybilla et al. 2008; Gvaramadze et al. 2009; Renzo et al. 2019), even if values up to ~ 1000 km s^{-1}are possible (Leonard 1991; Tauris 2015), but should be extremely rare for runaway stars (Brown 2015).

1.1.2 Hypervelocity Stars

The first observation of a hypervelocity star

With previous results for the ejection velocity of runaway stars in mind, it was a great surprise when, in 2005, a B-type star was observed in the outer halo of the MW with a heliocentric radial velocity of ~ 830 km s^{-1}(Brown et al. 2005, 2014). This value, once corrected for the motion of the Sun and the local standard of rest (LSR), corresponds to a lower limit on the total velocity of the star of 673 km s^{-1}(Brown et al. 2014), which is sufficiently high to escape the gravitational field of the MW at the star's position. The authors, targeting blue horizontal branch stars to trace the stellar halo, found this star to be a 6σ outlier from the radial velocity distribution. This *unbound* star, SDSS J090745.0+024507, is the first hypervelocity star (HVS) observed, and was referred to as HVS1. As a hint of its puzzling origin, the radial velocity vector of HVS1 points at $\sim 175°$ from the GC, suggesting an origin in the central region of our Galaxy. This intriguing possibility will now be further discussed.

The Hills mechanism

One possible way to explain the surprising velocity of HVS1 involves the interaction with a massive compact object. According to the Hills mechanism, the tidal field of the MBH in the centre of our Galaxy can disrupt a binary system passing sufficiently close (Hills 1988). This results in one of the stars starting to orbit around the MBH, with the other one being ejected with an incredibly high velocity, of the order of thousands of km s^{-1}. Following Brown (2015), we will now derive with a simple calculation an estimate of the ejection velocity of the HVS, showing how the Hills mechanism can easily explain the acceleration of stars to unbound velocities.

A stellar binary with total mass m_b and semi-major axis a gets disrupted by the gravitational field of a MBH of mass M, if the encounter happens

at a distance closer than the *tidal radius* $r_\bullet$. This characteristic distance is defined as the distance within which tidal forces from the MBH dominate over the binary binding force:

$$r_\bullet = a\left(3\frac{M}{m_b}\right)^{1/3} \simeq 14\,\mathrm{AU}\left(\frac{a}{0.1\,\mathrm{AU}}\right)\left(\frac{M_\odot}{m_b}\right)^{1/3}\left(\frac{M}{10^6\,M_\odot}\right)^{1/3}. \tag{1.2}$$

We can compare this characteristic scale to the Schwarzschild radius of a MBH:

$$r_{\mathrm{MBH}} = \frac{2GM}{c^2} \simeq 0.02\,\mathrm{AU}\left(\frac{M}{10^6 M_\odot}\right), \tag{1.3}$$

where c is the speed of light. We can see that, for MBHs with $M > 10^8\ \mathrm{M}_\odot$, stars fall inside the event horizon before reaching the tidal radius (Hills 1988). This is not the case in our Galaxy, where $M \simeq 4.3\cdot 10^6\ \mathrm{M}_\odot$ (Gillessen et al. 2017).

The typical orbital velocity of stars in an equal mass binary is:

$$v_b = \sqrt{\frac{Gm_b}{a}} \simeq 94\ \mathrm{km\ s^{-1}}\left(\frac{m_b}{M_\odot}\right)^{1/2}\left(\frac{0.1\ \mathrm{AU}}{a}\right)^{1/2}. \tag{1.4}$$

For example, $v_b \simeq 100$ km s^{-1}for a binary consisting of two 3 $\mathrm{M}_\odot$ stars at $a = 0.5$ AU. At the moment of the disruption of the binary, the binary orbital velocity is:

$$v = \sqrt{\frac{GM}{r_\bullet}} = v_b\left(\frac{M}{m_b}\right)^{1/3} \simeq 10^4\ \mathrm{km\ s^{-1}}. \tag{1.5}$$

This velocity is equal to few percent of the speed of light, and is consistent with observations of S stars in the GC (see for example Ghez et al. 2005).

When the binary gets disrupted, the stars experience a change in specific kinetic energy δE that we can compute as:

$$\delta E = \frac{1}{2}(v + v_b)^2 - \frac{1}{2}v^2 \simeq v v_b. \tag{1.6}$$

Using energy conservation, we can therefore estimate the resulting velocity of the star ejected from the binary as:

$$v_{\mathrm{ej}} = \sqrt{2 v v_b} \simeq 10^3\ \mathrm{km\ s^{-1}}. \tag{1.7}$$

Equation (1.7) shows that the Hills mechanism is able to predict ejection velocities in the GC up to thousands of km s^{-1}. These incredibly high velocities allow HVSs to travel across the whole MW on unbound trajectories.

Besides explaining the extreme velocities of the observed HVS, the Hills mechanism also provides a possible solution to the puzzling origin of the S stars in the GC: these stars are the binary companions of the ejected HVS, bound to the central MBH after the disruption. The observed orbit and eccentricity distributions of S stars are consistent with predictions from the Hills mechanism (Gillessen et al. 2009; Madigan et al. 2014), if the relaxation time is shorter than the stellar age (Habibi et al. 2017).

After being ejected in the GC, HVSs travel through the Galaxy on almost radial trajectories. A star moving at ~ 1000km s^{-1}travels a distance of ~ 1 kpc in ~ 1 Myr, a small fraction of the typical main sequence lifetime of a star. The initial velocity $v_{\rm ej}$ will then decrease because of the deceleration induced by the Galactic potential, which acts as a high-pass filter: only the stars with sufficiently high velocity at the ejection can travel to distances large enough to be observable (Kenyon et al. 2008). For example, stars with $v_{\rm ej} > 700$ km s^{-1}can reach the Sun position, stars with $v_{\rm ej} > 800$ travel to the edge of the stellar disk, and only stars with $v_{\rm ej} > 800$ can get to the virial radius of the MW, around 250 kpc from the GC. The radial motion of HVSs is deflected by the non-spherical components of the Galactic potential, namely the stellar disk, a possible triaxiality of the dark matter halo, and the presence of satellite galaxies orbiting the MW (Kenyon et al. 2018).

In addition to the population of unbound HVSs, the Hills mechanisms naturally predicts the existence of *bound HVSs*: stars ejected according to the same three-body interaction in the GC, but with an initial velocity not sufficient to escape from the gravitational field of the whole MW (Bromley et al. 2006; Kenyon et al. 2008). The trajectories of these stars do not follow straight lines anymore, and they can cross the stellar disk multiple times during their lifetime.

HVS observations

Following the first detection, a dedicated spectroscopic survey with the MMT telescope was performed to find HVS candidates (Brown et al. 2014). The survey targeted young stars in the outer halo of the MW, which are not expected to be found so far from an active star forming region (such as the GC), unless they traveled there with an extremely high velocity. The survey identified 21 unbound late B-type HVSs, with masses in the range $[2.5,\ 4]$ $M_\odot$, at distances $50 - 120$ kpc from the GC. All these stars are unbound

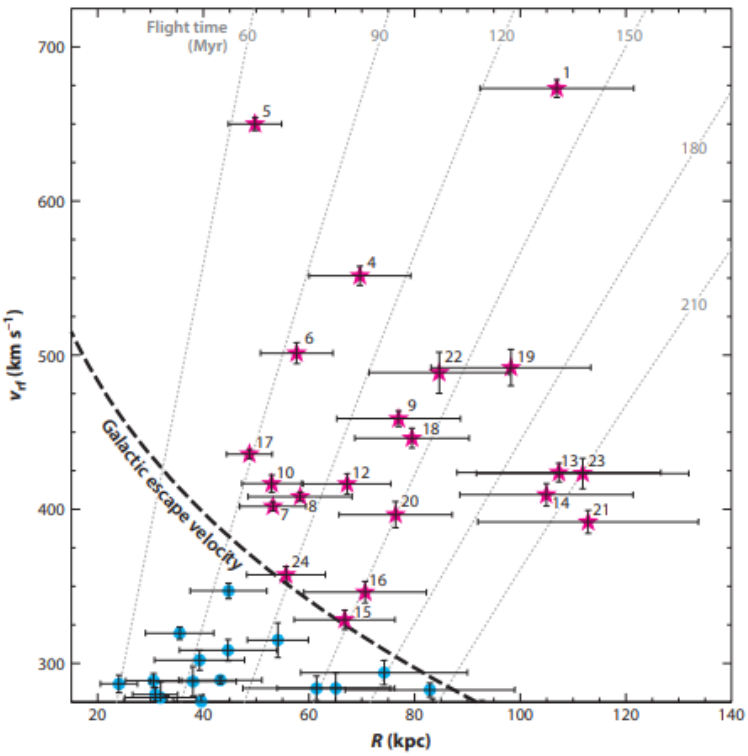


Figure 1.2: Total velocity as a function of Galactocentric distance for the HVS candidates discovered in the outer halo of the MW by the MMT HVS survey. Magenta stars mark the unbound candidates, while blue dots the bound ones. The dashed line marks the escape velocity from the Galaxy. From Brown (2015).

from radial velocity alone, and are moving outward (consistent with the prediction from the Hills mechanism). Fig. 1.2 shows the total velocity in the Galactic rest-frame as a function of distance from the GC for the stars found in the survey (Brown 2015). The dashed line is a choice for the escape speed from the Galaxy (Kenyon et al. 2008). Magenta stars are the unbound HVSs, while blue dots are the bound HVS candidates.

In addition to the population of young stars in the outer halo, many works focused on finding late-type, low mass HVS candidates in the Solar neighbourhood and the inner Galactic halo. For example, Palladino et al. (2014) discovered 20 HVS candidates in the G and K samples of the Sloan Extension for Galactic Understanding and Exploration (SEGUE), and Li et al. (2015) found 19 F, G, and K type candidates using LAMOST data. Most of the known late-type HVSs are likely to be bound to the MW, or not to originate from the GC (e.g. Zheng et al. 2014; Hawkins et al. 2015; Ziegerer et al. 2015, 2017; Boubert et al. 2018). Chemical tagging with high resolution spectroscopy can help to narrow down the ejection location of HVS candidates, by determining their precise chemical composition (e.g. Hawkins & Wyse 2018).

The search for HVSs is complicated by the fact that HVSs are extremely rare objects, with an ejection rate from the GC between 10^{-5} and 10^{-4} yr^{-1} (Brown et al. 2015). The advent of new astrometric and spectroscopic surveys will change dramatically our view on the fastest stars in our Galaxy (see Section 1.2).

Alternative ejection mechanisms for HVSs

In addition to the Hills mechanism, discussed in Section 1.1.2, other ejection scenarios have been proposed to explain the unbound velocities of observed HVSs. Yu & Tremaine (2003) first discussed the chance that HVSs could be ejected following the interaction between a single star and a massive black hole binary (BHB) in the GC. The possibility of a intermediate mass black hole orbiting around Sagittarius A^* cannot be excluded by observations in the GC, with current upper limits on its mass around 10^4 $\mathrm{M}_\odot$ (Gillessen et al. 2017). The presence of a fixed, preferential plane in the geometry of the encounter (the plane of the BHB) introduces an anisotropy in the expected spatial distribution of HVSs, which is flattened along the inspiral plane of the BHB. The degree of flattening is expected to decrease as the BHB hardens, leading to a more isotropic distribution (Sesana et al. 2006). HVSs produced by these mechanisms might be slower compared to the Hills mechanism, depending on the system parameters (Rasskazov

et al. 2019).

Recently it has been proposed that the known B-type HVS could be runaway stars ejected from the Large Magellanic Cloud (LMC), the most massive satellite galaxy orbiting the MW (Boubert et al. 2017a). The LMC is an active star forming region, so runaway stars ejected from supernova explosions in binary systems, summing their velocity to the orbital velocity of the LMC, can easily become unbound to the Galaxy. Recently, a HVS has been shown to originate almost from the centre of the LMC (Erkal et al. 2019), suggesting the presence of a MBH (Boubert & Evans 2016).

Other proposed mechanisms to produce HVSs include tidal interaction between dwarf galaxies infalling in the gravitational field of the MW (Abadi et al. 2009), which might accelerate stars to unbound velocities. Also, massive globular clusters infalling towards the centre of the Galaxy, interacting with the MBH or with a BHB, can produce a population of high velocities stars, with an unbound tail (Capuzzo-Dolcetta & Fragione 2015; Fragione & Capuzzo-Dolcetta 2016). Another possibility is the scatter between single stars and stellar black holes in the proximity of Sgr A* (O'Leary & Loeb 2008).

Different mechanisms predict different spatial and velocity distributions, therefore a large sample of HVSs can be used to investigate the dynamical processes responsible for the acceleration of these stars to unbound velocities.

HVSs as tools to investigate the Milky Way

HVSs are a unique probe to study our Galaxy as a whole. HVSs are predicted to originate in the centre of the MW, and then, because of their extremely large velocities, travel through the Galaxy on unbound trajectories. Therefore they provide a connection between the inner center and the outskirts of the Galaxy. The GC is difficult to observe because of dust extinction and stellar crowding, so HVSs can be used to probe the stellar population in the proximity of the quiescent MBH. A large sample of HVSs, for example, can be used to constrain the mass function and metallicity distribution in the inner parsec of the Galaxy. On the other hand, HVS trajectories are affected by the way the mass is distributed in the MW, therefore they can be used as probes of the Galactic Potential (e.g. Gnedin et al. 2005; Sesana et al. 2007; Yu & Madau 2007; Perets et al. 2009). In particular, the mass and orientation of the halo are still a matter of debate, and there is no general consensus on its shape (e.g. Wang et al. 2015; Bovy et al. 2016; Posti & Helmi 2019). Gnedin et al. (2005) first proposed HVSs to study the dark

matter halo of the MW. The authors show how precise proper motions of the first HVS candidate, SDSS J090745.0+024507, can provide constraints on the triaxiality of the halo, as predicted from cosmological simulations of structure formation. A recent work from Contigiani et al. (2019) shows how a sample of ~ 200 HVSs can be used to nail down the Galactic halo potential parameters with percent precision. In particular, HVSs are found to be extremely sensitive to the axis-ratio of the spheroidal, because of the spherical symmetry of the ejection in the Hills mechanism. A joint constraint on both the GC and the dark matter halo was first performed by Rossi et al. (2017), but tight constraints have been hampered by the low number of known HVSs. Recently, HVSs have also been proposed to constrain the Solar parameters, relying on the condition of zero azimuthal angular momentum (Hattori et al. 2018b).

1.2 The ESA mission *Gaia*

The ESA satellite *Gaia* was launched on 9 December 2013 from the European spaceport in French Guiana, and a few weeks later it arrived at the Lagrangian point L2 for a planned 5 years operations (Gaia Collaboration et al. 2016b). The goal of *Gaia* is to provide the largest three dimensional stellar catalogue ever produced of the Galaxy, providing positions, parallaxes, and proper motions for more than 1 billion sources, and radial velocities for a subset of bright stars. Here we outline the main contents of its data releases.

1.2.1 The first *Gaia* data release

The first data release (DR1) of the ESA satellite *Gaia* was delivered to the general public on the 14th of September 2016, and is based on observations collected between the 25th of July 2014 and the 16th of September 2015, for a total of almost 14 months (Gaia Collaboration et al. 2016b,a). Here we summarize the main contents of *Gaia* DR1:

- Coordinates (right ascension α and declination δ) and magnitudes in the *Gaia G* band for 1142679769 sources;
- The five parameters astrometric solution (positions, parallax ϖ, and proper motions μ_α, μ_δ) for 2057050 sources.

The presence of parallaxes and distances for more than 2 million stars was possible thanks to a joint Tycho-Gaia astrometric solution (TGAS),

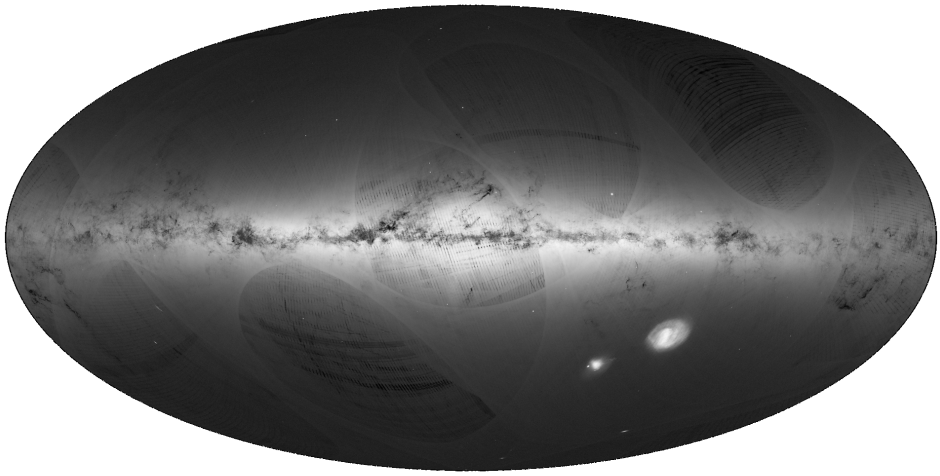


Figure 1.3: First full sky map released by *Gaia*, using data from DR1 (credits: ESA).

performed on the sources in common between *Gaia* and the Tycho-2 Catalogue (Michalik et al. 2015; Lindegren et al. 2016).

Fig. 1.3 shows the first full sky map made using data from *Gaia* DR1. A quick look at the map reveals the presence of characteristic arches and patterns in the density distribution. Those are a unique imprint of the *Gaia* scanning strategy on the sky, and disappeared in future data releases.

1.2.2 The second *Gaia* data release

The second data release (DR2) of *Gaia* happened on the 25th of April 2018, containing observations collected between the 25th of July 2014 and the 23rd of May 2016, spanning a period of 22 months (Gaia Collaboration et al. 2018a). DR2 represents a huge improvement over DR1, both in terms of number of sources observed, and of quality of the measurements. It contains:

- Position and *Gaia G* band magnitude for 1692919135 stars;
- Magnitudes in the *Gaia* blue pass (BP) $G_{\rm BP}$ and red pass (RP) $G_{\rm RP}$ band for 1381964755 and 138551713 sources, respectively;

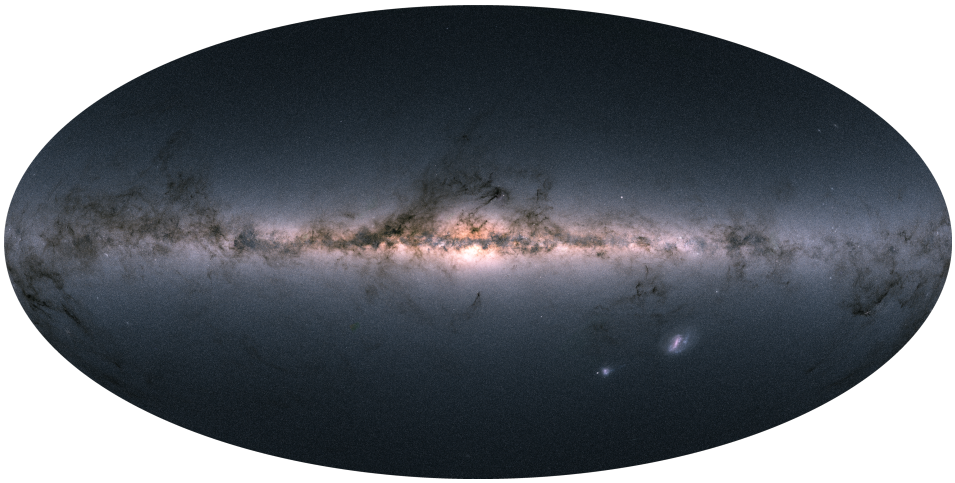


Figure 1.4: Full sky map released by *Gaia* DR2 (credits: ESA).

- The five parameters astrometric solution for 1331909727 sources;
- Radial velocity for 7224631 stars with $4 \lesssim G \lesssim 13$ and with effective temperatures $3550 \lesssim T_{\rm eff} \lesssim 6900$ K;
- Effective temperature for 161497595 stars;
- Extinction and reddening for 87733672 objects;
- Radius and luminosity for 76956778 sources.

Figure 1.4 shows the full sky map for the ~ 1.7 billion sources in *Gaia*, obtained combining the magnitudes in the G, $G_{\rm BP}$ and $G_{\rm RP}$ passbands. Comparing this to Figure 1.3 shows how all the arches due to the scanning law of the satellite have now disappeared, thanks to the longer baseline and more homogeneous sky coverage.

1.2.3 Future *Gaia* data releases

The third data release (DR3) of *Gaia* is currently planned to be split into two different releases. An early data release (EDR3) is expected in the third quarter of 2020 and will contain updated parallaxes and proper motions, with uncertainties reduced by the longer baseline (34 months of data). *Gaia* DR3 is expected in the second half of 2021 and will contain astrophysical parameters and radial velocities for all the spectroscopically well behaved

sources. The final *Gaia* data release, which has not been announced yet, will consist of the full photometric, astrometric, and radial velocity catalogues[2]. This will be the largest and most precise stellar catalogue ever produced and will allow understanding the history of the MW and its stellar population with unprecedented detail.

1.2.4 Warnings and caveats while using *Gaia* data

Gaia is the largest stellar catalogue ever produced, and the most recent data release (DR2) has provided astrometric measurements for more than 1.3 billion sources. There are known issues with *Gaia* astrometry and radial velocities, which have not been corrected for during the raw data reduction. Taking this into account while analyzing the data is essential. As an example, a possible wrong determination of the parallax can severely affect the distance determination, and therefore the total velocity of a star. Lindegren et al. (2018a) pointed out the existence of a global zero point in parallax of -0.029 mas, derived looking at the parallax distribution of distant quasars. This offset is expected to be different for bright sources, and asteroseismic and spectroscopic observations report a global offset of -0.05 mas for $G < 14$ (Zinn et al. 2019; Khan et al. 2019). Parallax uncertainties are also affected by systematics, which can be included inflating the quoted measurement errors by a magnitude-dependent factor (Lindegren et al. 2018a). Spurious astrometry from *Gaia* DR2 can be filtered out using the renormalised unit weight error (Lindegren et al. 2018a). In a recent paper, Boubert et al. (2019) show that *Gaia* spectra for stars in crowded regions could be contaminated by the light coming from nearby sources, causing a shift in the radial velocity measurement. The authors propose further quality cuts to select a clean sample of *Gaia* stars with reliable astrometric and spectroscopic measurements.

1.2.5 *Gaia* and HVSs

The advent of the exquisite astrometric data provided by *Gaia* has revolutionized our knowledge on high velocity stars. The combination of *Gaia* with ground-based spectroscopic surveys has enabled the determination of precise and accurate total velocities for millions of stars. Marchetti et al. (2017) first attempted to find HVS candidates in *Gaia* DR1/TGAS, using a data mining routine based on machine learning. Boubert et al. (2018) revisited the origin of previously known unbound objects with the updated *Gaia*

[2] `https://www.cosmos.esa.int/web/gaia/release`

DR2 astrometric information. The authors found that, apart from one star (LAMOST J115209.12+120258.0), all the high velocity late-type candidates are actually bound to the Galaxy, including the ones identified in Marchetti et al. (2017). For what concerns the late B-type HVS, the new *Gaia* proper motions confirm the GC origin for the fastest objects (Brown et al. 2018). Marchetti et al. (2018a) computed total velocities for all the $\sim$ 7 million stars with a radial velocity determination from *Gaia* DR2, finding 20 stars with high probabilities of being unbound, but no HVS candidates from the GC (in agreement with predictions from Marchetti et al. 2018b). Bromley et al. (2018) supported these findings, extending the search to stars with precise parallaxes and high tangential velocity. Hattori et al. (2018a) suggested that this sample is composed of old and metal-poor stars, a result confirmed by Hawkins & Wyse (2018) using high resolution spectroscopy. Thousands of HVSs with precise proper motions are expected to be contained in the *Gaia* catalogue (Marchetti et al. 2018b), but these stars are predicted to be too faint to have a radial velocity from *Gaia*, a fact that has so far prevented their discovery.

1.3 Methods used in this thesis

In this section we will quickly describe some of the methods used in this thesis to analyze and derive properties from the *Gaia* data: Bayes' theorem, which is the basic concept behind Bayesian statistics, and machine learning, which will be used in Chapter 3 to identify HVS candidates.

1.3.1 Bayes' Theorem

Bayes' theorem is a direct consequence of the law of conditional probability. Indicating with $P(A)$ and $P(B)$ the probabilities of two independent events A and B, we can write the conditional probability:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \tag{1.8}$$

We can now express equation (1.8) in a Bayesian fashion. To do that, we consider the case in which we want to fit some model parameter θ given the data $\mathbf{x}$. Equation (1.8) then becomes:

$$P(\theta|\mathbf{x}) = \frac{P(\mathbf{x}|\theta)P(\theta)}{P(\mathbf{x})}. \tag{1.9}$$

This is the most general form of Bayes' theorem. The term $P(\theta|\mathbf{x})$ is called *posterior* probability, and represents the probability distribution of the parameter θ given the data $\mathbf{x}$. The term $P(\mathbf{x}|\theta)$, called *likelihood* probability, is the probability of observing the data $\mathbf{x}$ given a certain model parametrized by θ. The term $P(\theta)$ is the *prior* probability, which represents our prior knowledge on the parameter θ. The advantage of Bayesian statistics is that we can incorporate this prior knowledge on the model parameters, which might come from other experiments. Finally, the term $P(\mathbf{x})$ is called the *model evidence*, and is a normalization factor that is usually not considered (one is interested in relative probabilities), so that equation (1.9) is just expressed as a proportionality.

A common approach is to determine the likelihood using the chosen physical model, assume a prior on the parameters, and then sample the posterior distribution with a Monte Carlo Markov Chain (MCMC) algorithm, such as the affine-invariant ensemble sampler emcee (Foreman-Mackey et al. 2013).

1.3.2 Machine Learning

Machine learning is a data-driven approach to science, in which the algorithms learn from existing data, to make predictions on new data. The machine learning approach is generally divided into two classes: supervised and unsupervised learning. Supervised learning algorithms rely on a training set: a set of data for which ones know the features (the properties used for the training) and the desired output. The goal of a supervised learning algorithm is to learn from the data what is the function that best maps inputs into outputs. In regression algorithms, the output is a single real number, while the goal of classification algorithms is to assign each data-point to a particular class so that the output of the algorithm is the probability that each input belongs to a given class. Unsupervised learning algorithms, on the other hand, do not need a training set for the learning process, but their goal is to find hidden structures in the data. The most common unsupervised learning algorithms are clustering algorithms, that aim to find clustering in a high dimensional space.

The training set comprises of m training examples, each one with n features: $x^{(i)} \in \mathbb{R}^n$, where the superscript (i) refers to the i-th training point. In a supervised learning algorithm, each training example $x^{(i)}$ corresponds to a label $y^{(i)}$, with $y^{(i)} \in \mathbb{R}$ for regression problems, and $y^{(i)} \in \{0, 1, \ldots, M\}$ for a classification problem with M distinct classes. The *hypothesis* function $h_\Theta(x^{(i)})$ represents our best estimate of $y^{(i)}$, which we call $\hat{y}^{(i)}$. For ex-

ample, in multivariate linear regression, we compute the hypothesis for a single data point as:

$$\hat{y}^{(i)} \equiv h_\Theta(x^{(i)}) = \theta_0 + \theta_1 x_1^{(i)} + \cdots + \theta_n x_n^{(i)}, \tag{1.10}$$

where $\Theta = (\theta_0, \ldots, \theta_n) \in \mathbb{R}^{n+1}$ is the parameter vector. In classification algorithms, the output of the hypothesis can be interpreted as the probability that the data point belongs to a certain class. So, for example, in multivariate logistic regression the hypothesis is computed applying a sigmoid function to equation (1.10).

The goal of a supervised machine learning algorithm is to find the values of the parameter vector Θ which minimize the cost function $J(\Theta)$, which is often defined as:

$$J(\Theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)} \right)^2, \tag{1.11}$$

which is the sum over all the training examples of the squared difference between the true labels y and the predicted labels $\hat{y}$. The search for the global minimum of the cost function is usually performed in an iterative fashion using the gradient descent optimization algorithm, but more advanced techniques have been proposed to achieve faster convergence (e.g. Robbins & Monro 1951; Duchi et al. 2011; Singh et al. 2015).

Artificial neural networks are supervised learning algorithms (see Haykin 2009, for an exhaustive description of neural networks). In chapter 3 of this thesis, we will make use of a neural network for a binary classification problem. The advantage of neural networks is their ability to learn highly non-linear mapping functions, more complex than the form presented in equation (1.10). Neural networks are often employed because of their ability to generalize: to provide reasonable outputs for inputs not encountered during the training session. A natural drawback is that overfitting can prevent the algorithm to generalize to new data-points. Overfitting can be avoided in several ways, both splitting the original training set into separate datasets that can be used to tune and test the algorithm, and applying different techniques of regularization by modifying the cost function in equation (1.11).

Neural networks are often used for pattern classification, image recognition, and in general high-dimensional problems with a large number of features. In astronomy, these algorithms are getting popular in different fields, for example for estimating redshifts or for galaxy classification (e.g. Dai & Tong 2018; Stivaktakis et al. 2018; Carrasco-Davis et al. 2018).

1.4 Thesis content

This thesis focuses on the search for the fastest stars in our Galaxy. We combine modelling, observations, and data mining techniques to identify and characterize these rare objects in the largest and most precise stellar catalogue ever produced: the data released from the ESA satellite *Gaia*.

In **Chapter 2** we create mock catalogues of HVSs to predict the properties of the HVS population in *Gaia*. We build three mock catalogues, adopting different assumptions on the ejection mechanism, including the Hills mechanism and the interaction between a single star and a massive black hole binary. In all cases, we find hundreds to thousands of HVSs to be contained in the final *Gaia* data release with precise proper motions, representing a huge improvement over the few tens of known candidates. We show how their identification is not trivial since the bulk of the population is expected to be too faint to have a radial velocity measurement from *Gaia*. Therefore new, advanced data mining techniques need to be implemented to search for these rare objects.

In **Chapter 3** we develop, implement and apply a novel data mining routine based on machine learning techniques, to identify HVS candidates in the *Gaia* DR1 TGAS subset. We choose to use an artificial neural network, trained on mock populations of HVSs created in Marchetti et al. (2018b), as presented in Chapter 2. Because of the missing radial velocity information, we choose to use the 5 parameters astrometric solution for the training process. The application to the TGAS subset results in the identification of 80 stars with high probabilities of being HVSs. Subsequent spectroscopic follow-ups with the Isaac Newton Telescope in La Palma and cross-match with spectroscopic surveys of the MW resulted in radial velocities for more than half of the candidates. We discovered one possibly unbound HVS, 5 bound HVSs, and 5 runaway star candidates with median velocities up to ~ 780 km s^{-1}.

Chapter 4 focuses on characterizing the high velocity tail of the velocity distribution of stars in the MW, using the subset of ~ 7 million stars with a radial velocity measurement from *Gaia* DR2. We derive distances from *Gaia* parallaxes using a Bayesian approach, and we then compute total velocities for the whole sample of stars. Focusing on the subset of stars with reliable astrometric measurements from *Gaia*, we identify 125 stars with

predicted probability $> 50\%$ of being unbound from the MW, and 20 with a probability $> 80\%$. Thanks to the precise full phase information given by *Gaia*, we can trace back in time these stars in the Galactic potential to identify their ejection location. We discover 7 stars coming from the stellar disk, consistent with being runaway stars. Surprisingly, the remaining 13 stars cannot be traced back to any star forming region. These objects have a preferred extragalactic origin, and they could be the result of the tidal disruptions of satellite galaxies from the gravitational field of the MW, or might be runaway stars originating in MW satellite galaxies, such as the LMC.

In **Chapter 5** we use the sample of ~ 20 unbound late B-type HVSs from Brown et al. (2014) to give joint constraints on the GC binary population and on the dark matter halo of the MW. We model the ejection velocity distribution of HVSs adopting the Hills mechanism, and we compare the resulting observed velocity distribution to the HVS data using a Kolmogorov-Smirnov (KS) statistical test. We find that assuming typical values observed in Galactic star forming regions for the binary properties in the GC, a good fit is achieved for dark matter haloes that result into an escape velocity from the GC to 50 kpc lower than 850 km s^{-1}. For realistic choices of the mass profile, these haloes are consistent with MW circular velocity data out to ~ 100 kpc, and with predictions from the concordance Λ CDM cosmological model. The discovery of hundreds of HVSs will break degeneracies between the GC and potential parameters, allowing a systematic study of these two different but complementary environments.

2 Predicting the hypervelocity star population in *Gaia*

T. Marchetti, *O. Contigiani, E.M. Rossi, J.G. Albert, A.G.A. Brown, A. Sesana 2018, MNRAS, 476, 4697-4712*

Hypervelocity stars (HVSs) are amongst the fastest objects in our Milky Way. These stars are predicted to come from the Galactic centre (GC) and travel along unbound orbits across the Galaxy. In the coming years, the ESA satellite *Gaia* will provide the most complete and accurate catalogue of the Milky Way, with full astrometric parameters for more than 1 billion stars. In this paper, we present the expected sample size and properties (mass, magnitude, spatial, velocity distributions) of HVSs in the *Gaia* stellar catalogue. We build three *Gaia* mock catalogues of HVSs anchored to current observations, exploring different ejection mechanisms and GC stellar population properties. In all cases, we predict hundreds to thousands of HVSs with precise proper motion measurements within a few tens of kpc from us. For stars with a relative error in total proper motion below 10 per cent, the mass range extends to $10\ \mathrm{M}_\odot$ but peaks at $\sim 1\ \mathrm{M}_\odot$. The majority of *Gaia* HVSs will therefore probe a different mass and distance range compared to the current non-*Gaia* sample. In addition, a subset of a few hundreds to a few thousands of HVSs with $M \sim 3\ \mathrm{M}_\odot$ will be bright enough to have a precise measurement of the three-dimensional velocity from *Gaia* alone. Finally, we show that *Gaia* will provide more precise proper motion measurements for the current sample of HVS candidates. This will help identifying their birthplace narrowing down their ejection location, and confirming or rejecting their nature as HVSs. Overall, our forecasts are extremely encouraging in terms of quantity and quality of HVS data that can be exploited to constrain both the Milky Way potential and the GC properties.

2.1 Introduction

A hypervelocity star (HVS) is a star observationally characterized by two main properties: its velocity is higher than the local escape velocity from our Galaxy (it is gravitationally unbound), and its orbit is consistent with a Galactocentric origin (Brown 2015). The term HVS was originally coined by Hills (1988), and the first detection happened only in 2005 (Brown et al. 2005). Currently ~ 20 HVS candidates have been found by the MMT HVS Survey of the northern hemisphere, in a mass range $[2.5, 4]$ $M_\odot$, and at distances between 50 kpc and 100 kpc from the Galactic Centre (GC) (Brown et al. 2014). This restricted mass range is an observational bias due to the survey detection strategy, that targets massive late B-type stars in the outer halo, that were not supposed to be found there (the halo is not a region of active star formation), unless they were ejected somewhere else with very high velocities. Lower mass HVSs have been searched for in the inner Galactic halo, using high proper motion, high radial velocity, and/or metallicity criteria. Most of these candidates are bound to the Galaxy, and/or their trajectories seem to be consistent with a Galactic disc origin (e.g. Heber et al. 2008; Palladino et al. 2014; Zheng et al. 2014; Hawkins et al. 2015; Ziegerer et al. 2015; Zhang et al. 2016; Ziegerer et al. 2017).

One puzzling aspect of the observed sample of B-type HVSs is their sky distribution: about half of the candidates are clumped in a small region of the sky (5 % of the coverage area of the MMT HVS Survey), in the direction of the Leo constellation (Brown 2015). Different ejection mechanisms predict different distributions of HVSs in the sky, and a full sky survey is needed in order to identify the physics responsible for their acceleration.

The leading mechanism to explain the acceleration of a star up to ~ 1000 km s^{-1}is the Hills mechanism (Hills 1988). According to this scenario, HVSs are the result of a three body interaction between a binary star and the massive black hole (MBH) residing in the centre of our Galaxy, Sagittarius A*. In it simpler version, this mechanism predicts an isotropic distribution of HVSs in the sky. One possible alternative ejection mechanism involves the interaction of a single star with a massive black hole binary (MBHB) in the GC (Yu & Tremaine 2003). Current observations cannot exclude the presence of a secondary massive compact object companion to Sagittarius A*, with present upper limits around 10^4 $M_\odot$ (Gillessen et al. 2017). In this case, the ejection of HVSs becomes more energetic as the binary shrinks, and it typical lasts for tens of Myr. This results in a ring of HVSs ejected in a very short burst, compared to the continuous ejection of

stars predicted by the Hills mechanism (e.g. Gualandris et al. 2005; Sesana et al. 2006, 2008). Other mechanisms involve the interaction of a globular cluster with a super massive black hole (Capuzzo-Dolcetta & Fragione 2015) or with a MBHB (Fragione & Capuzzo-Dolcetta 2016), the interaction between a single star and a stellar black hole orbiting a MBH (O'Leary & Loeb 2008), and the tidal disruption of a dwarf galaxy (Abadi et al. 2009). Recent observations have even shown evidence of star formation inside a galactic outflow ejected with high velocity from an active galactic nucleus (Maiolino et al. 2017), suggesting that HVSs can be produced in other galaxies in such jets (Silk et al. 2012; Zubovas et al. 2013).

A more recent explanation for the observed B-type HVSs is given by Boubert et al. (2017b), which interpret the current sample of candidates clumped in the direction of the Leo constellation as runaway stars from the Large Magellanic Cloud (LMC). Alternatively, HVSs could be produced by an hypothetical MBH in the centre of the LMC with a process that is analogous to the Hills mechanism (Boubert & Evans 2016).

All these mechanisms predict an additional population of stars, called *bound HVSs*. These objects are formed in the same scenario as HVSs, but their velocity is not sufficiently high to escape from the gravitational field of the MW (e.g. Bromley et al. 2006; Kenyon et al. 2008). These slower stars can travel along a wide variety of orbits, making their identification very difficult (Marchetti et al. 2017).

In the past years HVSs have been proposed as tools to study multiple components of our Galaxy. The orbits of HVSs, spanning an unprecedented range of distances from the GC, integrate the Galactic potential, making them powerful tracers to study the matter distribution and orientation of the MW (i.e. Gnedin et al. 2005; Sesana et al. 2007; Yu & Madau 2007; Kenyon et al. 2014; Fragione & Loeb 2017). On the other hand, HVSs come from the GC, therefore they can be used to probe the stellar population near a quiescent MBH (Kollmeier et al. 2009, 2010). It has been shown that a fraction of the original companions of HVSs can be tidally disrupted by the MBH, therefore the ejection rate of HVSs is directly linked to the growth rate of Sagittarius A* (Bromley et al. 2012). A clean sample of HVSs would be also useful to constrain the metallicity distribution of stars in the GC. Rossi et al. (2017), adopting the Hills mechanism, first attempted to constrain both the properties of the binary population in the GC (in terms of distributions of semi-major axes and mass ratios) and the scale parameters of the dark matter halo, using the sample of unbound HVSs from Brown et al. (2014). They show that degeneracies between the parame-

ters are preventing us from giving tight constraints, because of both the restricted number and the small mass range of the HVS candidates.

The ESA satellite *Gaia* is going to revolutionize our knowledge of HVSs, shining a new light on their properties and origin. Launched in 2013, *Gaia* is currently mapping the sky with an unprecedented accuracy, and by its final release (the end of 2022) it will provide precise positions, magnitudes, colours, parallaxes, and proper motions for more than 1 billion stars (Gaia Collaboration et al. 2016b,a). Moreover, the Radial Velocity Spectrometer (RVS) on board will measure radial velocities for a subset of bright stars (magnitude in the *Gaia* RVS band $G_{\rm RVS} < 16$). On the 14th September 2016 the first data (*Gaia* DR1) were released. The catalogue contains positions and G magnitudes for more than 1 billion of sources. In addition, the five parameter astrometric solution (position, parallax, and proper motions) is available for a subset of $\sim 2 \times 10^6$ stars in common between *Gaia* and the Tycho-2 catalogue: the Tycho-*Gaia* Astrometric Solution (TGAS) catalogue (Michalik et al. 2015; Lindegren et al. 2016). The next data release, *Gaia* DR2, is planned for the 25th of April 2018, and will be consisting of the five parameter astrometric solution, magnitudes, and colours for the full sample of stars ($> 10^9$ sources). It will also provide radial velocities for 5 to 7 million stars brighter than the 12th magnitude in the $G_{\rm RVS}$ band. Effective temperatures, line-of-sight extinctions, luminosities, and radii will be provided for stars brighter than the 17th magnitude in the G band (Katz & Brown 2017).

A first attempt to find HVSs in *Gaia* DR1/TGAS can be found in Marchetti et al. (2017), who developed a data-mining routine based on an artificial neural network trained on mock populations to distinguish HVSs from the dominant background of other stars in the Milky Way, using only the provided astrometry and no radial velocity information. This approach avoids biasing the search for HVSs towards particular spectral types, making as few assumptions as possible on the expected stellar properties. They found a total of 14 stars with a total velocity in the Galactic rest frame higher than $400\,{\rm km\,s^{-1}}$, but because of large uncertainties, a clear identification of these candidates as HVSs is still uncertain. Five of these stars have a probability higher than 50% of being unbound from the MW. Because most of the stars have masses of the order of the Solar mass, they form a different population compared to the observed late B-type stars.

In this work, we forecast the sample size and properties of the HVS data expected in the next data releases of *Gaia*, starting in April with DR2. The manuscript is organised as follows. In Section 2.2 we explain how we build

our first mock catalogue of HVSs, the Vesc catalogue, using a simple assumption on the total stellar velocity, and how we simulate *Gaia* observations of these stars. Here we present our first results: how many HVSs we are expecting to find in the *Gaia* catalogue using this first simple catalogue. In Section 2.3 we specialise our estimates on HVSs adopting the Hills mechanism, drawing velocities from a probability distribution, and we show how previous estimates and results change because of this assumption. In Section 2.4 we build the third mock catalogue, the MBHB catalogue, assuming that HVSs are produced following the three-body interaction of a star with a MBHB. Here we also discuss the resulting number estimates. Finally, in Section 2.5 we estimate *Gaia* errors on the current sample of HVS candidates presented in Brown et al. (2015), and in Section 2.6 we summarize our results for the different catalogues, and we discuss their implications and limitations in view of the following data releases from the *Gaia* satellite.

2.2 The "Vesc" Mock Catalogue: A Simple Approach

We create synthetic populations of HVSs in order to assess and forecast *Gaia*'s performance in measuring their proper motions and parallax. We characterise the astrometric and photometric properties of the stars using their position in Galactic coordinates (l, b, r) and mass M, and then estimate *Gaia*'s precision in measuring these properties.

In this section we choose to compute the total velocity v of a HVS adopting a simple conservative approach, i.e. to assume it equal to the escape velocity from the Galaxy at its position:

$$v(l, b, r) = v_{\rm esc}(l, b, r). \tag{2.1}$$

Our decision is motivated by the choice not to focus on a particular ejection mechanism, but just to rely on the definition of a HVS as an unbound object. In addition to that, proper motions for a star travelling away from the GC on a radial orbit are directly proportional to the velocity, see equations (2.2) and (2.3), therefore a higher velocity (e.g. for an unbound star) would result in a lower relative error in total proper motion, making the detection by *Gaia* even more precise (refer to Section 2.2.3). This catalogue does not make any assumption on the nature and origin of HVSs, and the impact of adopting a particular ejection mechanism for modelling the ve-

locity distribution is explored in Sections 2.3 and 2.4, where we also introduce predictions for the expected bound population of HVSs.

For clarity and reference within this paper, we refer to this first catalogue as Vesc.

2.2.1 Astrometric Characterization of a HVS

In first approximation, HVSs are travelling away from the Milky Way on radial trajectories. This assumption holds if we consider the contribution given by the stellar disc to be sub-dominant in the total deceleration of the star (Kenyon et al. 2014), and if we neglect deviations from spherical symmetry in the dark matter halo (Gnedin et al. 2005). For a given position in the sky (l, b, r), it is possible to derive the combination of proper motions in Galactic coordinates ($\mu_{l*} \equiv \mu_l \cos b,\ \mu_b$) which is consistent with a star flying away from the GC on a straight line:

$$\mu_{l*}(l,b,r) = \frac{\hat{\mathbf{p}} \cdot \mathbf{v}(l,b,r)}{r} = v(l,b,r)\frac{d_\odot}{r}\frac{\sin l}{r_{\rm GC}(l,b,r)}, \tag{2.2}$$

$$\mu_b(l,b,r) = \frac{\hat{\mathbf{q}} \cdot \mathbf{v}(l,b,r)}{r} = v(l,b,r)\frac{d_\odot}{r}\frac{\cos l \sin b}{r_{\rm GC}(l,b,r)}, \tag{2.3}$$

where $\hat{\mathbf{p}}$ and $\hat{\mathbf{q}}$ are unit basis vectors defining the plane tangential to the celestial sphere, $d_\odot$ is the distance between the Sun and the GC, and $r_{\rm GC}(l,b,r) = \sqrt{r^2 + d_\odot^2 - 2rd_\odot \cos l \cos b}$ is the Galactocentric distance of the star. In the following, we will assume $d_\odot = 8.2$ kpc (Bland-Hawthorn & Gerhard 2016). In order to simulate how these stars will appear in the *Gaia* catalogue, we correct proper motions for the motion of the Sun and for the local standard of rest (LSR) velocity, following Schönrich (2012).

The total velocity v, equal to the escape velocity from the Milky Way in that position, is computed assuming a three component Galactic potential: a Hernquist bulge (Hernquist 1990):

$$\phi_b(r_{\rm GC}) = -\frac{GM_b}{r_{\rm GC} + r_b}, \tag{2.4}$$

a Miyamoto & Nagai disk in cylindrical coordinates $(R_{\rm GC}, z_{\rm GC})$ (Miyamoto & Nagai 1975):

$$\phi_d(R_{\rm GC}, z_{\rm GC}) = -\frac{GM_d}{\sqrt{R_{\rm GC}^2 + \left(a_d + \sqrt{z_{\rm GC}^2 + b_d^2}\right)^2}}, \tag{2.5}$$

Table 2.1: Parameters for the three-components Galactic potential adopted in the paper.

Component	Parameters
Bulge	$M_b = 3.4 \cdot 10^{10}\ \mathrm{M}_\odot$
	$r_b = 0.7$ kpc
Disk	$M_d = 1.0 \cdot 10^{11}\ \mathrm{M}_\odot$
	$a_d = 6.5$ kpc
	$b_d = 0.26$ kpc
Halo	$M_h = 7.6 \cdot 10^{11}\ \mathrm{M}_\odot$
	$r_s = 24.8$ kpc

and a Navarro-Frenk-White (NFW) halo profile (Navarro et al. 1996):

$$\phi(r_{\mathrm{GC}}) = -\frac{GM_h}{r_{\mathrm{GC}}} \ln\Big(1 + \frac{r_{\mathrm{GC}}}{r_s}\Big). \tag{2.6}$$

The adopted values for the potential parameters M_b, r_b, M_d, a_d, b_d, M_h, and r_s are summarized in Table 2.1. The mass and radius characteristic parameters for the bulge and the disk are taken from Johnston et al. (1995); Price-Whelan et al. (2014); Hawkins et al. (2015), while the NFW parameters are the best-fit values obtained in Rossi et al. (2017). This choice of Galactic potential has been shown to reproduce the main features of the Galactic rotation curve up to 100 kpc (Huang et al. (2016), see Fig. A1 in Rossi et al. 2017).

As a result of *Gaia* scanning strategy, the total number of observations per object depends on the ecliptic latitude of the star β, which we determine as (Jordi et al. 2010):

$$\sin\beta = 0.4971 \sin b + 0.8677 \cos b \sin(l - 6.38^\circ). \tag{2.7}$$

To complete the determination of the astrometric parameters, we simply compute parallax as $\varpi = 1/r$, where ϖ is expressed in arcsec and r in parsec.

2.2.2 Photometric Characterization of a HVS

Knowing the position and the velocity of a HVS in the Galaxy, we now want to characterize it from a photometric point of view, since *Gaia* errors on the astrometry depend on the brightness of the source in the *Gaia* passbands.

To compute the apparent magnitudes in different bands, we need to know the age of the HVS at the given celestial location at the moment of its observation. This is required in order to correctly estimate its stellar

parameters (radius, luminosity, and effective temperature) and the corresponding spectrum. We estimate the *flight time* $t_{\rm f}$, the time needed to travel from the ejection region in the GC to the observed position, as:

$$t_{\rm f}(l, b, r) = \frac{r_{\rm GC}(l, b, r)}{v_0(l, b, r)}, \tag{2.8}$$

where $v_0(r, l, b)$ is the velocity needed for a star in the GC to reach the observed position (r, l, b) with zero velocity. We compute v_0 using energy conservation, evaluating the potential in the GC at $r = 3$ pc, the radius of influence of the MBH (Genzel et al. 2010). Since HVSs are decelerated by the Galactic potential, $t_{\rm f}$ is a lower limit on the actual flight time needed to travel from 3 pc to the observed position. We then compare this time to the total main sequence (MS) lifetime $t_{\rm MS}(M)$, which we compute using analytic formulae presented in Hurley et al. (2000)[1], assuming a solar metallicity value (Brown 2015). If $t_{\rm f} > t_{\rm MS}$ we exclude the star from the catalogue: its lifetime is not long enough to reach the corresponding position. On the other hand, if $t_{\rm f} < t_{\rm MS}$, we estimate the age of the star as:

$$t(M, l, b, r) = \varepsilon\big(t_{\rm MS}(M) - t_{\rm f}(l, b, r)\big), \tag{2.9}$$

where ε is a random number, uniformly distributed in $[0, 1]$.

We evolve the star along its MS up to its age t using analytic formulae presented in Hurley et al. (2000), which are functions of the mass and metallicity of the star. We are then able to get the radius of the star $R(t)$, the effective temperature $T_{\rm eff}(t)$, and the surface gravity $\log g(t)$. Chi-squared minimization of the stellar parameters $T_{\rm eff}(t)$ and $\log g(t)$ is then used to find the corresponding best-fitting stellar spectrum, and therefore the stellar flux, from the BaSeL SED Library 3.1 (Westera & Buser 2003), assuming a mixing length of 0 and a an atmospheric micro-turbulence velocity of 2 km s^{-1}.

At each point of the sky we estimate the visual extinction $A_{\rm V}$ using the three-dimensional Galactic dust map MWDUST[2] (Bovy et al. 2016). The visual extinction is then used to derive the extinction at other frequencies A_λ using the analytical formulae in Cardelli et al. (1989), assuming $R_{\rm V} = 3.1$.

Given the flux $F(\lambda)$ of the HVS and the reddening we can then compute the magnitudes in the *Gaia G* band, integrating the flux in the *Gaia*

[1] We assume the MS lifetime to be equal to the total lifetime of a star.

[2] `https://github.com/jobovy/mwdust`

passband $S(\lambda)$ (Jordi et al. 2010):

$$G = -2.5 \log\left(\frac{\int d\lambda \; F(\lambda) \; 10^{-0.4 A_\lambda} \; S(\lambda)}{\int d\lambda \; F^{\mathrm{Vega}}(\lambda) \; S(\lambda)} \right) + G^{\mathrm{Vega}}. \tag{2.10}$$

The zero magnitude for a Vega-like star is taken from Jordi et al. (2010). Similarly, integrating the flux over the Johnson-Cousins V and I_C filters, we can compute the colour index $V - I_C$ (Bessell 1990). We then compute the magnitude in the *Gaia* G_{RVS} band using polynomial fits in Jordi et al. (2010).

2.2.3 *Gaia* Error Estimates

We use the Python toolkit PyGaia[3] to estimate post-commission, end-of-mission *Gaia* errors on the astrometry of our mock HVSs. Measurement uncertainties depend on the ecliptic latitude, *Gaia G* band magnitude, and the $V - I_C$ colour of the star, which we all derived in the previous sections. We can therefore reconstruct *Gaia* precision in measuring the astrometric properties of each HVS, which we quantify as the (uncorrelated) relative errors in total proper motion $z_\mu \equiv \sigma_\mu / \mu$, and in parallax $z_\varpi \equiv \sigma_\varpi / \varpi$.

2.2.4 Number Density of HVSs

In order to determine how many HVSs *Gaia* is going to observe with a given precision, we need to model their intrinsic number density. We assume a continuous and isotropic ejection from the GC at a rate $\dot{N}$. Indicating with $\rho(r_{\mathrm{GC}}, M)$ the number density of HVSs with mass M at a Galactocentric distance r_{GC}, we can simply write the total number of HVSs with mass M within r_{GC} as:

$$N(< r_{\mathrm{GC}}, \; M) = \int_0^{r_{\mathrm{GC}}} 4\pi r'^2 \rho(r', \; M) dr'. \tag{2.11}$$

We assume HVSs to travel for a time $t_{\mathrm{F}} = r_{\mathrm{GC}} / v_{\mathrm{F}}$ to reach the observed position, where $v_{\mathrm{F}} = 1000$ km s^{-1} is an effective average travel velocity. We also neglect the stellar lifetime after its MS, which could only extend by $\sim 10\%$ the travel time. Current observations seem to suggest that the ejection of a HVS occurs at a random moment of its lifetime: $t_{\mathrm{ej}} = \eta t_{\mathrm{MS}}$

[3] https://github.com/agabrown/PyGaia

(Brown et al. 2014), with η being a random number uniformly distributed in $[0, 1]$. We can then only observe a HVS at a distance $r_{\rm GC}$ if $t_{\rm F}$ satisfies:

$$t_{\rm F} = \frac{r_{\rm GC}}{v_{\rm F}} < t_{\rm MS} - t_{\rm ej} = t_{\rm MS}(1-\eta). \tag{2.12}$$

We can then write the total number of HVSs of mass M within $r_{\rm GC}$ as:

$$N(< r_{\rm GC},\ M) = \phi(M)\dot{N}\frac{r_{\rm GC}}{v_{\rm F}}\int_0^1 \theta\left(t_{\rm MS}(1-\eta) - \frac{r_{\rm GC}}{v_{\rm F}}\right) d\eta, \tag{2.13}$$

where $\phi(M)$ is the mass function of HVSs, and $\theta(x)$ is the Heaviside step function. Differentiating this expression, we get:

$$\begin{aligned}\frac{\partial N(< r_{\rm GC},\ M)}{\partial r_{\rm GC}} =& \phi(M)\frac{\dot{N}}{v_{\rm F}}\int_0^1 \left[\theta\left(t_{\rm MS}(1-\eta) - \frac{r_{\rm GC}}{v_{\rm F}}\right) + \right. \\ &\left. - \delta\left(t_{\rm MS}(1-\eta) - \frac{r_{\rm GC}}{v_{\rm F}}\right)\frac{r_{\rm GC}}{v_{\rm F}}\right] d\eta,\end{aligned} \tag{2.14}$$

where $\delta(x)$ is the Dirac delta function. Evaluating the integral and comparing this equation with the one obtained by differentiating equation (2.11) with respect to $r_{\rm GC}$, we can express the number density of HVSs within a given Galactocentric distance $r_{\rm GC}$ and with a given mass M as:

$$\begin{aligned}\rho(r_{\rm GC}, M) =& \theta\left(t_{MS}(M) - \frac{r_{\rm GC}}{v_{\rm f}}\right)\phi(M)\cdot\left(\frac{\dot{N}}{4\pi v_{\rm f}\, r_{\rm GC}^2} + \right. \\ &\left. - \frac{\dot{N}}{2\pi r_{\rm GC}\, t_{MS}(M) v_{\rm f}^2}\right).\end{aligned} \tag{2.15}$$

Brown et al. (2014), taking into account selection effects in the MMT HVS Survey, estimated a total of $\simeq 300$ HVSs in the mass range $[2.5,\ 4]\ M_\odot$ over the entire sky within 100 kpc from the GC, that is:

$$N\Big(r_{\rm GC} < 100\ {\rm kpc},\ M \in [2.5, 4]\ M_\odot\Big) = \varepsilon_{\rm f}\dot{N}\frac{100{\rm kpc}}{v_{\rm f}} = 300. \tag{2.16}$$

In this equation, $\varepsilon_{\rm f}$ is the mass fraction of HVSs in the $[2.5, 4]\ M_\odot$ mass range, taking into account the finite lifetime of a star:

$$\varepsilon_{\rm f} = \varepsilon_0 \int_{2.5M_\odot}^{4M_\odot} \phi(M) dM \int_0^1 \theta\left(t_{\rm M}(1-\eta) - \frac{100{\rm kpc}}{v_{\rm f}}\right) d\eta. \tag{2.17}$$

Assuming a particular mass function we can therefore estimate the ejection rate $\dot{N}$ needed to match observations using equation (2.16) and (2.17). In the following we will assume a Kroupa IMF (Kroupa 2001), for which we get $\dot{N} \simeq 2.8 \cdot 10^{-4}$ year^{-1}. This estimate is consistent with other observational and theoretical estimates (Hills 1988; Perets et al. 2007; Zhang et al. 2013; Brown et al. 2014).

For each object in the mock catalogue we can then compute the intrinsic number density of HVSs in that given volume $dV dM$ using equation (2.15). With a coordinate transformation to the heliocentric coordinate system, the corresponding number of HVSs in the volume element $dV\ dM$ is:

$$\begin{aligned} N(l, b, r, M) &= \rho(r_{\rm GC}, M) dV dM \\ &= \rho(l, b, r, M) r^2 \cos b\ dl\ db\ dr\ dM . \end{aligned} \tag{2.18}$$

2.2.5 *"Vesc"* Catalogue: Number Estimates of HVSs in *Gaia*

We sample the space (l, b, r, M) with a resolution of $\sim 6^\circ$ in l, $\sim 3^\circ$ in b, ~ 0.7 kpc in r and $\sim 0.15\ M_\odot$ in M. For each point we count how many HVSs lay in the volume element $dV\ dM$ using equation (2.18). We want to stress that the results refer to the end-of-mission performance of the *Gaia* satellite.

Fig. 2.1 shows the cumulative radial distribution of HVSs within 40 kpc: stars which will be detectable by *Gaia* with a relative error on total proper motion below 10% (1%) are shown with a blue (purple) line, and those with a relative error on parallax below 20% with a red line. The total number of HVSs with a relative error on total proper motion below 10% (1%) is 709 (241). The total number of HVSs with a relative error on parallax below 20% is 40. We have chosen a relative error threshold of 0.2 in parallax because, for such stars, it is possible to make a reasonable distance estimate by simply inverting the parallax, without the need of implementing a full Bayesian approach (Bailer-Jones 2015; Astraatmadja & Bailer-Jones 2016a,b). This is a great advantage, because uncertainties due to the distance determination dominate the errorbars in total velocity (Marchetti et al. 2017). In all cases we can see that almost all detectable HVSs will be within 10 kpc from us.

Fig. 2.2 shows the total number of HVSs expected to be found in the *Gaia* catalogue as a function of the chosen relative error threshold in total proper motion (solid) and parallax (dashed). We see that there is a total of ~ 1000 (~ 60) HVSs with a relative error on total proper motion (parallax)

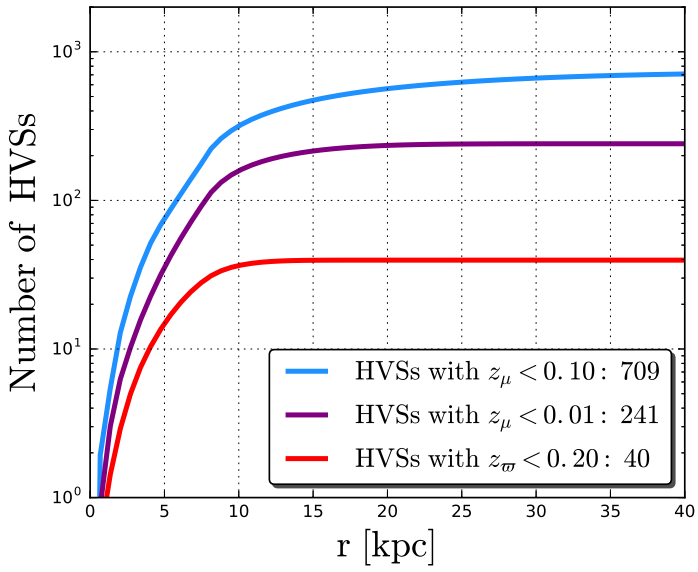


Figure 2.1: VESC catalogue: cumulative radial distributions of HVSs: the total number of HVSs within a heliocentric radius r. The blue (purple) line shows the cumulative radial distribution for HVSs which will be observable by *Gaia* with a relative error on total proper motion below 10% (1%). The red line refers to those stars with a relative error on parallax below 20%.

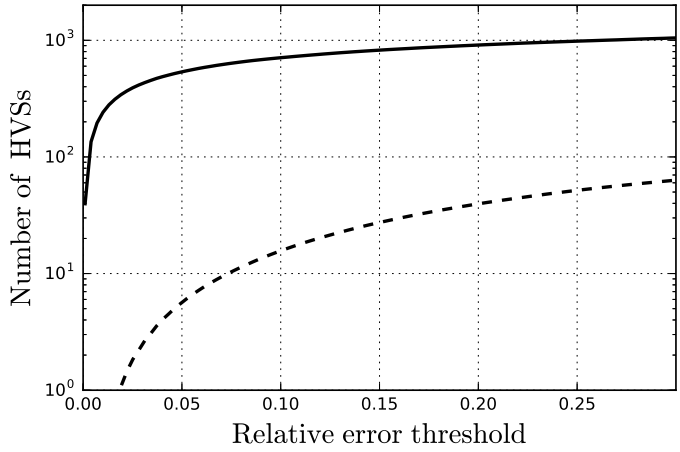


Figure 2.2: VESC catalogue: cumulative number of HVSs in the *Gaia* catalogue for a relative error on total proper motion (solid line) and parallax (dashed line) within a given relative error threshold.

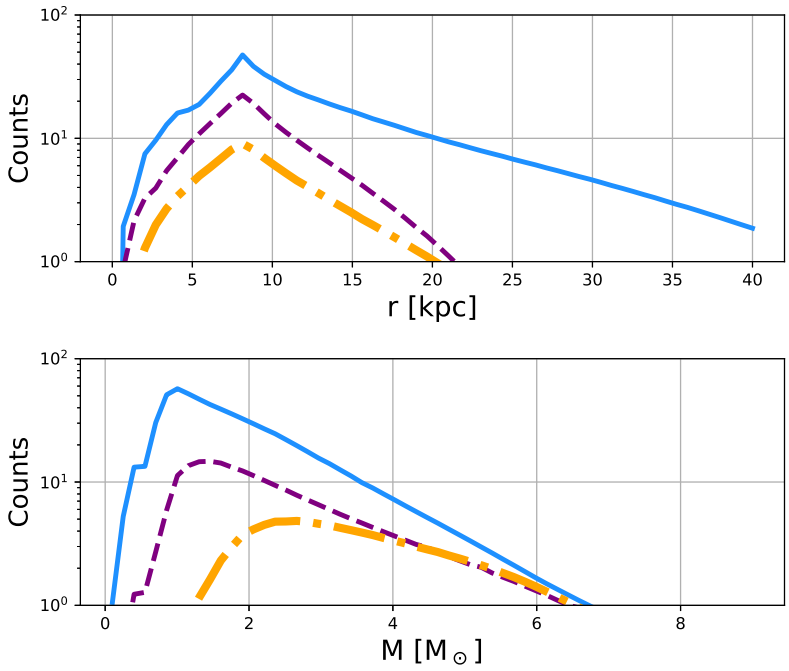


Figure 2.3: VESC catalogue: heliocentric distance (upper panel) and mass (lower panel) distribution for HVSs detectable by *Gaia* with a relative error on total proper motion below 10% (solid), 1% (dashed), and for the golden sample of HVSs with a three-dimensional velocity by *Gaia* alone (dot-dashed).

below 30%. This imbalance reflects the lower precision with which *Gaia* is going to measure parallaxes compared to proper motions.

Since proper motions are the most precise astrometric quantities, we quantify the radial and mass distribution of these precisely-measured HVSs in Fig. 2.3. The solid and dashed curves refer, respectively, to stars detectable with a relative error on total proper motion below 10% and 1%. Most HVSs with precise proper motions measurement will be at $r \simeq 8.5$ kpc, but the high-distance tail of the distribution extends up to ~ 40 kpc for HVSs with $z_\mu < 10\%$. The most precise proper motions will be available for stars within ~ 20 kpc from us. Also the mass distribution has a very well-defined peak which occurs at $M_{\rm peak} \simeq 1\ M_\odot$, consistent with observational results in Marchetti et al. (2017). This is due to two main factors. The chosen IMF predicts many more low-mass than high mass stars, therefore we would expect a higher contribution from low-mass stars, but on the other hand low-mass stars tend to be fainter, and therefore will be detectable by

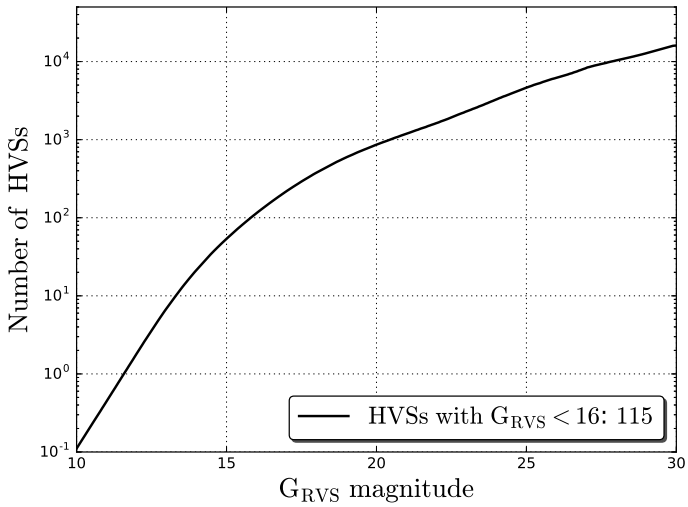


Figure 2.4: VESC catalogue: cumulative distribution of HVSs in the *Gaia* G_{RVS} passband (the golden sample). We estimate a total of 115 HVSs brighter than the 16th magnitude in this filter.

Gaia with a larger relative error. These two main contributions shape the expected mass function of HVSs in the catalogue.

Thanks to our mock populations and mock *Gaia* observations, we can also determine for how many HVSs *Gaia* will provide a radial velocity measurement. We refer to this sample as the *golden sample* of HVSs, since these stars will have a direct total velocity determination by *Gaia*. To address this point we compute the cumulative distribution of magnitudes in the G_{RVS} passband, as shown in Fig. 2.4. There is a total of 115 HVSs which satisfy the condition $G_{RVS} < 16$, required for the Radial Velocity Spectrometer to provide radial velocities. The dot-dashed line in Fig. 2.3 shows the distance and mass distribution for the golden sample of HVSs. The radial distribution is similar to the one shown in Fig.2.3, with a peak at $r \simeq 8.5$ kpc. The mass distribution instead has a mean value $\simeq 3.6\ M_\odot$ and a high-mass tail which extends up to $\simeq 6\ M_\odot$.

Fig. 2.5 shows the cumulative distribution function of stars in the golden sample with a relative error on proper motion (solid) and on parallax (dashed) below a given threshold. This plot shows that proper motions will be detected with great accuracy for all of the stars: $z_\mu \lesssim 0.4\%$ over the whole mass range. 39 of these stars (34% of the whole golden sample) will have $z_\varpi < 20\%$, and therefore it will be trivial to determine a distance for these

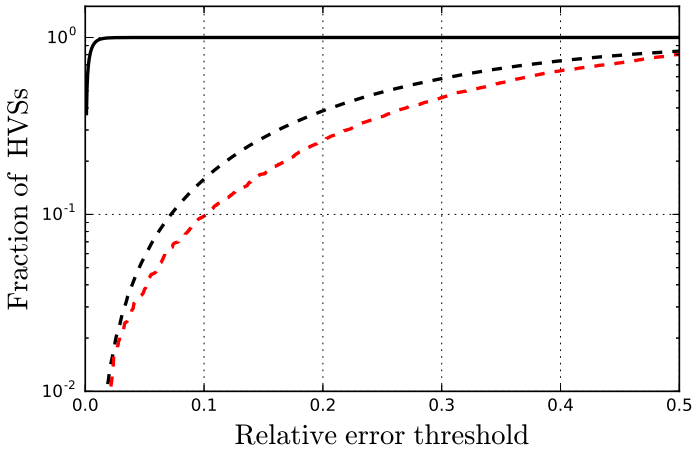


Figure 2.5: Cumulative fraction of HVSs in the golden sample within a certain threshold for relative errors in total proper motion (solid) and parallax (dashed). The black curves refer to the VESC catalogue, while the red dashed one to the HILLS catalogue (refer to Section 2.3). The red solid line overlaps with the black one, therefore it is not shown in the plot. The two curves for the MBHB catalogue coincide with the ones for the VESC catalogue, and thus are not shown.

stars, by simply inverting the parallax.

Estimates in *Gaia* DR1/TGAS and DR2

On September 14th 2016, *Gaia* DR1 provided positions and G magnitudes for all sources with acceptable errors on position (1142679769 sources), and the full five-parameters solution (α, δ, ϖ, $\mu_{\alpha*}$, μ_δ) for stars in common between *Gaia* and the *Tycho*-2 catalogue (2057050 sources, the TGAS catalogue) (Gaia Collaboration et al. 2016b,a; Lindegren et al. 2016).

To estimate the number of HVSs expected to be found in the TGAS subset of the first data release, we repeat the analysis of Section 2.2.5 considering the principal characteristics of the *Tycho*-2 star catalogue (Høg et al. 2000). We employ a $V < 11$ magnitude cut, corresponding to the $\sim 99\%$ completeness of the *Tycho*-2 catalogue (Høg et al. 2000). We find a total of 0.46 HVSs surviving this magnitude cut. This result is consistent with results in Marchetti et al. (2017), which find only one star with both a predicted probability $> 50\%$ of being unbound from the Galaxy and a trajectory consistent with coming from the GC.

Gaia data release 2, planned for April 2018, will be the first release pro-

viding radial velocities. It will consists of the five-parameter astrometric solution for the full billion star catalogue, and radial velocity will be provided for stars brighter than $G_{\rm RVS} = 12$. We find a total of 2 HVSs to survive the $G_{\rm RVS} < 12$ magnitude cut.

2.3 The "Hills" Catalogue

In the previous analysis we derived model independent estimates for unbound stars, by assuming that the total velocity of a HVS in a given point is equal to the local escape velocity from the Milky Way. In this and the next section, we instead employ a physically motivated velocity distribution. In this section we adopt the Hills mechanism (Hills 1988), the most successful ejection mechanism for explaining current observations (Brown 2015). In this case we will have a population of bound HVSs, in addition to the unbound ones (see discussion in Section 2.1). We call this catalogue Hills, to differentiate it from the simpler Vesc catalogue introduced and discussed in Section 2.2.

2.3.1 Velocity Distribution of HVSs

We start by creating a synthetic population of binaries in the GC, following and expanding the method outlined in Rossi et al. (2017) and Marchetti et al. (2017). We identify three parameters to describe binary stars: the mass of the primary m_p (the more massive star), the mass ratio between the primary and the secondary $q < 1$, and the semi-major axis of the orbit a. For the primary mass, we assume a Kroupa initial mass function in the range $[0.1, 100]$ $\mathrm{M}_\odot$, which has been found to be consistent with the initial mass function of stellar populations in the GC (Bartko et al. 2010). We assume power-laws for the distributions of mass ratios and semi-major axes: $f_q \propto q^\gamma$, $f_a \propto a^\alpha$, with $\gamma = -1$, $\alpha = -3.5$. This combination is consistent with observations of B-type binaries in the 30 Doradus star forming region of the LMC (Dunstall et al. 2015), and provides a good fit to the known HVS candidates from the HVS survey for reasonable choices of the Galactic potential (Rossi et al. 2017). The lower limit for a is set by the Roche lobe overflow: $a_{\rm min} = 2.5 \max(R_p, R_s)$, where R_p and R_s are, respectively, the radius of the primary and secondary star. The radius is approximated using the simple scaling relation $R_i \propto m_i$, with $i = p, s$. We arbitrarily set the upper limit of a to 2000 $\mathrm{R}_\odot$.

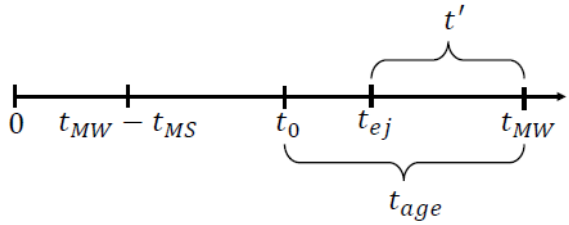


Figure 2.6: Sequence of events in the life of a HVS with a total lifetime $t_{\rm MS}(M) < t_{\rm MW}$. The instant 0 corresponds to the time when the MW was formed, while $t_{\rm MW}$ is today, when we observe the HVS in the sky. The time t_0 ($t_{\rm ej}$) is the age of the Galaxy when the HVS was born (ejected). The time t' is the flight time of the HVS, while $t_{\rm age}$ is its present age.

Kobayashi et al. (2012) showed that, for a binary approaching the MBH on a parabolic orbit, there is an equal probability of ejecting either the primary or the secondary star in the binary. We then randomly label one star per binary as HVS (mass M) and the other one as the bound companion (mass m_c). Following Sari et al. (2010); Kobayashi et al. (2012); Rossi et al. (2014) we then sample velocities from an ejection distribution which depends analytically on the properties of the binary approaching the MBH:

$$v_{\rm ej} = \sqrt{\frac{2Gm_{\rm c}}{a}}\left(\frac{M_\bullet}{m_{\rm t}}\right)^{1/6}, \tag{2.19}$$

where $M_\bullet = 4.3 \cdot 10^6\ M_\odot$ is the mass of the MBH in our Galaxy (Ghez et al. 2008; Gillessen et al. 2009; Meyer et al. 2012), $m_{\rm t} = M + m_{\rm c}$ is the total mass of the binary, and G is the gravitational constant. This equation represents the resulting ejection velocity after the disruption of the binary for a star at infinity with respect only to the MBH potential. Rigorously, there should be a numerical factor depending on the geometry of the three-body encounter in front of the square root, but it has been shown to be of the order of unity when averaged over the binary phase, and not to influence the overall velocity distribution (Sari et al. 2010; Rossi et al. 2014).

2.3.2 Flight Time Distribution of HVSs

Following the discussion in Section 2.2.2, the flight time t' of a HVS is defined as the time between its ejection from the GC and its observation. We assume the total lifetime of a star of mass M to be equal to its main sequence lifetime $t_{\rm MS}(M)$, and we also assume $t_{\rm MW} = 13.8$ Gyr to be the current age of the MW (Planck Collaboration et al. 2016). We compute the average fight

time for stars to which the condition $t_{\rm MS}(M) < t_{\rm MW}$ applies. We call t_0 and $t_{\rm ej}$, respectively, the age of the Galaxy at the instant when a HVS visible today is born and when the star is ejected. We assume t_0 to be distributed uniformly between $t_{\rm MW} - t_{\rm MS}(M)$ and $t_{\rm MW}$:

$$t_0(M) = t_{\rm MW} - t_{\rm MS}(M)(1 - \epsilon_1), \tag{2.20}$$

and $t_{\rm ej}$ to be distributed uniformly between $t_0(M)$ and $t_{\rm MW}$:

$$t_{\rm ej}(M) = t_0(M) + \epsilon_2(t_{\rm MW} - t_0(M)). \tag{2.21}$$

In the above expressions, ϵ_1 and ϵ_2 are two random numbers uniformly distributed in $[0, 1]$. Finally, we can express the flight time of a HVS as:

$$t'(M) = t_{\rm MW} - t_{\rm ej}(M) = \varepsilon_1 \varepsilon_2 t_{\rm MS}(M), \tag{2.22}$$

where $\varepsilon_1 \equiv (1 - \epsilon_1)$ and $\varepsilon_2 \equiv (1 - \epsilon_2)$ are two random numbers uniformly distributed in $[0, 1]$. Figure 2.6 visually presents the relevant time intervals. The probability density function for t' is then:

$$f(t', M) = -\frac{1}{t_{\rm MS}(M)} \log \frac{t'(M)}{t_{\rm MS}(M)}. \tag{2.23}$$

We can then write the survival function $g(t', M)$, the fraction of HVSs alive at a time t' after the ejection, as:

$$g(t', M) = 1 - \int_0^{t'} f(\tau, M) d\tau = 1 + \frac{t'(M)}{t_{\rm MS}(M)} \left(\log \frac{t'(M)}{t_{\rm MS}(M)} - 1 \right). \tag{2.24}$$

We can express the age of a HVS at the moment of its observation as:

$$t_{\rm age}(M) = t_{\rm MW} - t_0(M) = \varepsilon_1 t_{\rm MS}(M). \tag{2.25}$$

To take into account low-mass stars with $t_{\rm MS}(M) \geqslant t_{\rm MW}$, we rewrite equations (2.22) and (2.25) as:

$$t'(M) = \begin{cases} \varepsilon_1 \varepsilon_2 t_{\rm MS}(M) & \text{if } t_{\rm MS}(M) < t_{\rm MW} \\ \varepsilon_1 \varepsilon_2 t_{\rm MW} & \text{if } t_{\rm MS}(M) \geqslant t_{\rm MW} \end{cases}, \tag{2.26}$$

$$t_{\rm age}(M) = \begin{cases} \varepsilon_1 t_{\rm MS}(M) & \text{if } t_{\rm MS}(M) < t_{\rm MW} \\ \varepsilon_1 t_{\rm MW} & \text{if } t_{\rm MS}(M) \geqslant t_{\rm MW} \end{cases}. \tag{2.27}$$

2.3.3 Initial Conditions and Orbit Integration

The ejection velocity for the Hills mechanism, given by equation (2.19), is the asymptotic velocity of a HVS at an infinite distance from the MBH. In practice, we model this distance as the radius of the gravitational sphere of the influence of the black hole, which is constrained to be of the order of $\bar{r}_0 = 3$ pc (Genzel et al. 2010). We then initialize the position of each star at a distance of $\bar{r}_0$, with random angles (latitude, longitude) drawn from uniform spherical distributions. Velocities are drawn according to equation (2.19), and the velocity vector is chosen is such a way to point radially away from the GC at the given initial position, so that the angular momentum of the ejected star is zero.

The following step is to propagate the star in the Galactic potential up to its position (l, b, r) after a time t' from the ejection. We do that assuming the potential model introduced in Section 2.2.1. The orbits are integrated using the publicly available Python package galpy[4] (Bovy 2015a) using a Dormand-Prince integrator (Dormand & Prince 1980). The time resolution is kept fixed at 0.015 Myr. We check for energy conservation as a test for the accuracy of the orbit integration.

We therefore obtain for each star its total velocity v in the observed position, and we build a mock catalogue of HVSs with relative errors on astrometric properties, following the procedure outlined in Sections 2.2.1 to 2.2.3.

2.3.4 "Hills" Catalogue: Number Estimates of HVSs in *Gaia*

We start by estimating the number of HVSs currently present in our Galaxy. We call $\frac{dn}{dM}(M)$ the normalized probability density function of masses upon ejection. We note that this is not a Kroupa function, because the HVS is not always the primary star of the binary, and the secondary star is drawn according the mass ratio distribution $f_q \propto q^{-3.5}$. Assuming that HVSs have been created at a constant rate η for the entire Milky Way's lifetime $t_{\rm MW}$, the present Galactic population of HVSs in the mass range $[0.5, 9]$ $\mathrm{M}_\odot$ is:

$$N = \eta \int_o^{t_{\rm MW}} dt' \int_{0.5\mathrm{M}_\odot}^{9\mathrm{M}_\odot} dM \frac{dn}{dM}(M) g(t', M). \tag{2.28}$$

We choose to restrict ourselves to the mass range $[0.5, 9]$ $\mathrm{M}_\odot$ because stars with higher or lower masses are, respectively, very rare given our chosen

[4] `https://github.com/jobovy/galpy`

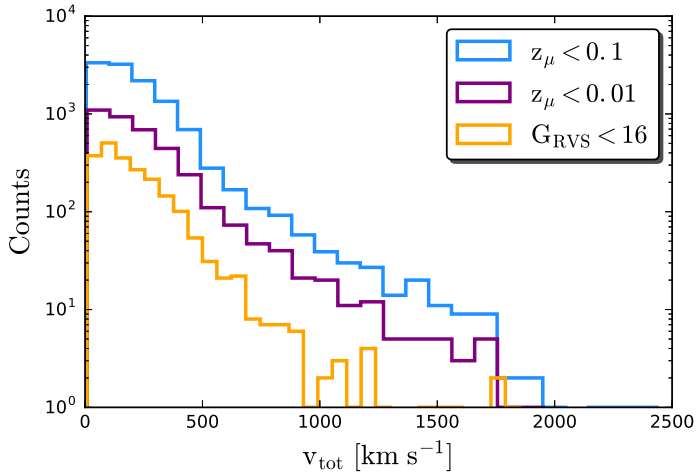


Figure 2.7: HILLS catalogue: distribution of total velocities in the Galactocentric rest frame for the HVSs with a relative error on total proper motion below 10% (blue), 1% (purple), and with a radial velocity measurement (yellow).

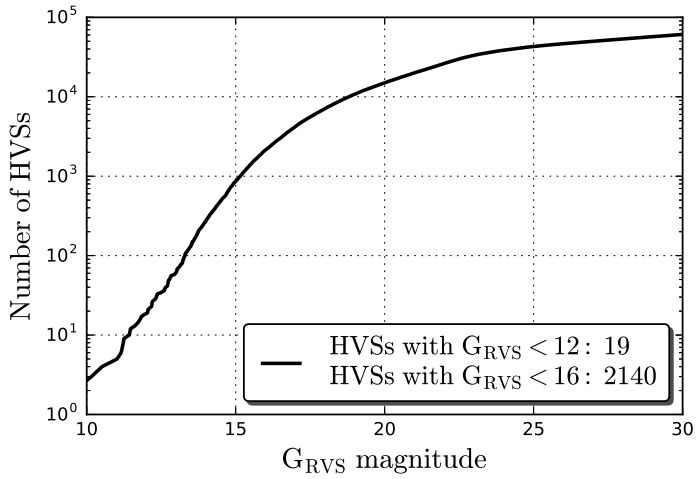


Figure 2.8: HILLS catalogue: cumulative distribution of HVSs in the *Gaia* G_{RVS} passband. We estimate a total of 2140 HVSs brighter than the 16th magnitude in this filter, and 19 HVSs brighter than the 12th magnitude.

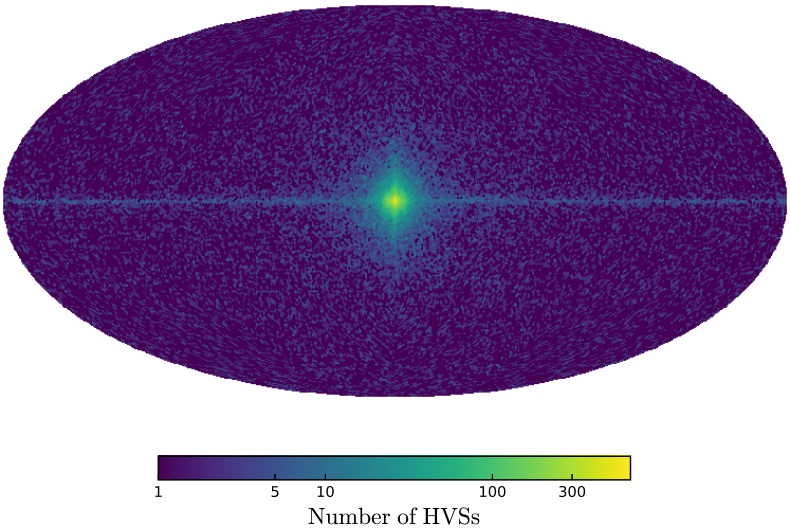


Figure 2.9: HILLS catalogue: sky distribution in Galactic coordinates of the current population of HVSs in our Galaxy (10^5 stars).

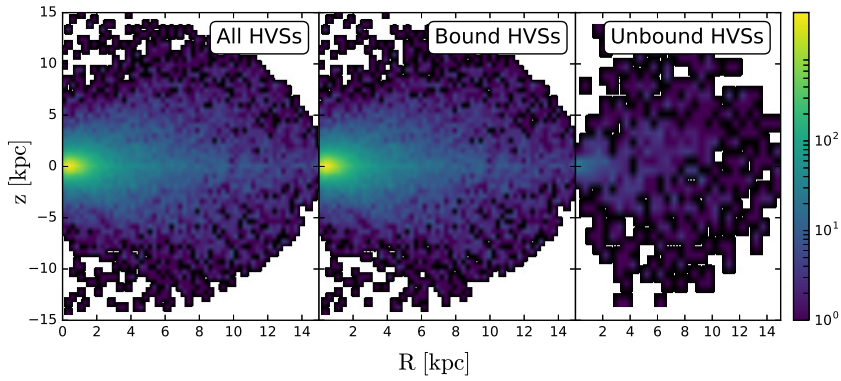


Figure 2.10: HILLS catalogue: distribution in Galactocentric cylindrical coordinates (R, z) of all HVSs (left), bound HVSs (centre), and unbound HVSs (right) within 15 kpc from the Galactic Centre.

IMF or not bright enough to be detectable by *Gaia* with good precision. Assuming the value $\eta = 2.8 \cdot 10^{-4}$ yr^{-1} derived in Section 2.2.4, anchored to the current observations of HVSs, we get $N \simeq 10^5$. We thus generate 10^5 HVSs in the GC as explained in the previous sections, and we propagate them in the Galaxy.

We can now use this realistic mock catalogue to predict the main properties of the Galactic population of HVSs. We find:

- 52% of the total number of stars travel along unbound orbits. Note that this does *not* imply that most of the HVSs will be detected with high velocities: given our choice of the Galactic potential, the escape velocity curve decreases to a few hundreds of km s^{-1} at large distances from the GC ($\gtrsim 100$ kpc). Therefore a large number of HVSs is classified as unbound even if velocities are relatively low. In particular, we find 5% (6%) of the stars with $z_\mu < 0.1$ ($z_\mu < 0.01$) to be unbound from the MW. The distribution of total velocities in the Galactic rest frame is shown in Fig. 2.7, where we can see that the distribution peaks at $v < 500$ km s^{-1}. The blue (purple) curve refers to HVSs that will be detected by *Gaia* with a relative error on total proper motion below 10% (1%), while the yellow curve is the distribution of HVSs with a radial velocity measurement. We can see that majority of stars with extremely high velocities ($v \gtrsim 1000$ km s^{-1}) will not be brighter than $G_{\rm RVS} = 16$, but few of them will be included in the catalogue, becoming the fastest known HVSs. The majority of stars, having low velocities, could easily be mistaken for disc, halo, or runaway stars, based on the module of the total velocity only (refer to discussion in Section 2.6).

- 2.1% of the HVSs will have $G_{\rm RVS} < 16$ with *Gaia* radial velocities. This amounts to 2140 stars. The proper motion and parallax error distributions for this golden sample of HVSs are shown in Fig. 2.5. The cumulative distribution function of $G_{\rm RVS}$ magnitudes for all stars in the mock catalogue is shown in Fig. 2.8. 68 of the $G_{\rm RVS} < 16$ stars are unbound. 165 of the $G_{\rm RVS} < 16$ have total velocity above 450 km s^{-1}.

- From Fig. 2.8 we can see that 19 stars are brighter than the 12th magnitude in the $G_{\rm RVS}$ band, so there will be direct *Gaia* radial velocity measurements already in *Gaia* DR2. We find 0 of these stars to be unbound from the MW. Proper motion error estimates for *Gaia*

DR2 can be obtained rescaling the errors from PyGaia by a factor[5] $(60/21)^{1.5} \sim 4.8$. We find all the 19 stars to have relative errors in total proper motion $\lesssim 0.01\%$, and in parallax $\lesssim 20\%$.

- 250 unbound HVSs with masses in $[2.5, 4]$ $M_\odot$ are within 100 kpc from the GC. This number is consistent with the observational estimate in Brown et al. (2014).

Fig. 2.9 shows the distribution in Galactic coordinates of the population of 10^5 HVSs, while Fig. 2.10 shows the distribution in Galactocentric cylindrical coordinates of the HVSs within 15 kpc from the Galactic Centre. In all cases we can see that most HVSs lie in the direction of the GC: $(l, b) = (0, 0)$. This is due to the presence of the population of bound HVSs, whose velocity is not high enough to fly away from the Milky Way, and therefore they spend their lifetime in the central region of the Galaxy on periodic orbits. Fig. 2.10 also shows how the majority of HVSs in the inner part of the Galaxy are travelling on bound orbits.

The distance distribution of the HVS sample is shown in the top panel of Fig. 2.11 for three samples: stars with a relative error on total proper motion below 10% (blue), below 1% (purple), and with a three-dimensional velocity determination (yellow). We can see that most stars lie within few tens of kpc from us, with only a few objects at distances ~ 50 kpc. We also note the substantial overlap between the purple and the yellow histogram, suggesting again that HVS with a radial velocity measurement will have an accurate *total* velocity by *Gaia*. The peak in the distributions, below 10 kpc, well agrees with the one shown in Fig. 2.3.

We show the mass distribution of the sample of HVSs in the bottom panel of Fig. 2.11. The colour code is the same as before. As expected, massive stars are brighter, and will therefore be measured by *Gaia* with a higher precision. This reflects in the fact that the distribution peaks to higher masses for lower relative error thresholds (brighter stars). In any case, we see that the shape of the curves resembles the ones obtained with the simple approach described in Section 2.2 (see Fig. 2.3).

We can compare our estimates with results from Marchetti et al. (2017), who data-mined *Gaia* DR1/TGAS searching for HVSs. In the Hills catalogue we find a total of 5 HVSs with a magnitude in the V band lower than 11, the $\sim 99\%$ completeness of the *Tycho*-2 catalogue (Høg et al. 2000).

[5]This numerical factor is derived considering that *Gaia* DR2 uses 21 months of input data, and that the error on proper motion scales as $t^{-1.5}$ (taking into account both the photon noise and the limited time baseline).

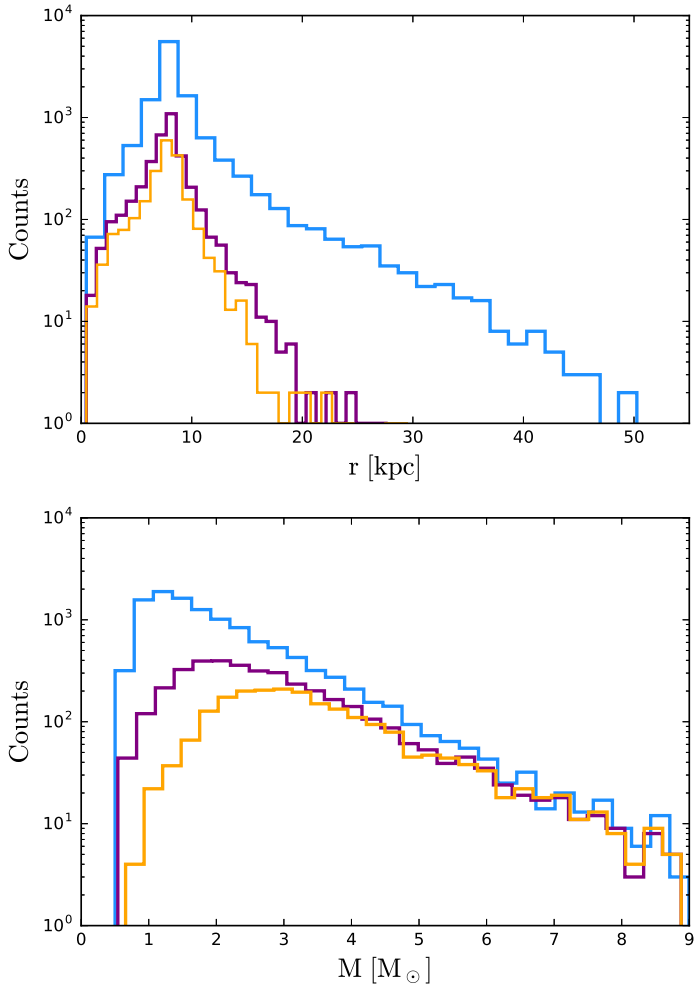


Figure 2.11: HILLS catalogue: heliocentric distance (top) and mass (bottom) distribution of the HVSs with a relative error on total proper motion below 10% (blue), 1% (purple), and with a radial velocity measurement by *Gaia* (yellow).

None of these stars are unbound, and the typical velocities are < 400 km s^{-1}.

2.4 The "MBHB" Catalogue

In this section, we explain how we create a mock population of HVSs ejected by a hypothetical massive black hole binary in the GC. We rely on results from full three-body scattering experiments presented in Sesana et al. (2006). In the following we will assume a massive black hole companion to Sagittarius A* with a mass $M_c = 5 \cdot 10^3$ $M_\odot$, which can not be ruled out by the latest observational results of S stars in the Galactic Centre (Gillessen et al. 2017). We assume a stellar density in the GC $\rho = 7 \cdot 10^4$ $M_\odot$ pc^{-3} and a velocity dispersion of stars in the GC $\sigma = 100$ km s^{-1} (Sesana et al. 2007). The MBHB, with mass ratio $q \simeq 1.2 \cdot 10^{-3}$, is assumed to be in a circular orbit, with an initial separation $a_0 = 0.01$ pc at a given time t_0 after the Milky Was formed, corresponding to a look-back time $t_{\rm lb}$. Using the results presented in Sesana et al. (2006), we adopt the best-fit parameters for the lowest mass ratio explored in their simulation, i.e. $q = 1/243$. This choice is motivated by noticing that the authors' results do not vary appreciably when comparing results obtained for different mass ratios (see Fig.3 and 5 in Sesana et al. 2006). In the following we will assume that the orbit of the MBHB remains circular as the binary shrinks.

2.4.1 Ejection of HVSs by the MBHB

We create a grid of 100 semi-major axes evenly spaced on a logarithmic scale, from a minimum value equal to 0.01 a_h, to a maximum value of a_0. The value a_h defines the minimum separation of a hard binary (Quinlan 1996):

$$a_h = \frac{GM_c}{4\sigma^2} \simeq 110 \text{ au}. \tag{2.29}$$

The total stellar mass ejected by the binary in each bin is computed as Sesana et al. (2006):

$$\Delta M_{\rm ej} = J(M_\bullet + M_c)\Delta \ln\left(\frac{a_h}{a}\right), \tag{2.30}$$

where a is the semi-major axis of the MBHB, and the mass ejection rate $J = J(a)$ is computed using the fitting function presented in Sesana et al.

(2006), with best-fit parameters for a circular orbit with mass ratio $q = 1/243$.

Rates of Orbital Decay

We now compare the rate of orbital decay of the MBHB due to the ejection of HVSs to the one due to the emission of gravitational waves (GWs). We determine the hardening rate of the binary following Quinlan (1996):

$$H = \frac{\sigma}{G\rho}\frac{d}{dt}\left(\frac{1}{a}\right). \tag{2.31}$$

A hard binary ($a < a_h$) hardens at a constant rate H.

The rate of orbital decay due to the ejection of HVSs is then computed as:

$$\left.\frac{da}{dt}\right|_{\rm HVS} = -\frac{G\rho H}{\sigma}a^2, \tag{2.32}$$

where the hardening rate $H = H(a)$ is computed using the numerical fit in Sesana et al. (2006) assuming a circular binary with $q = 1/243$.

The rate of orbital decay due to the emission of gravitational radiation can be approximated by (Peters 1964):

$$\left.\frac{da}{dt}\right|_{\rm GW} = -\frac{64}{5}G^3c^5\frac{(M_\bullet M_c)(M_\bullet + M_c)}{a^3}. \tag{2.33}$$

The two rates of orbital decay are equal for $\bar{a} = 48.4$ au $\sim 0.44a_h$. For $a < \bar{a}$ the orbital evolution is dominated by the emission of gravitational waves, driving the binary to the merging. The binary will start evolve more rapidly, ejecting stars with a lower rate, since the time the binary spends in each bin of a will be dictated by the emission of GWs. For $a < \bar{a}$ we therefore correct equation (2.30) by multiplying it for $T_{\rm GW}/T_{\rm HVS}$, where $T_{\rm GW}$ is the time needed to shrink from a to $a - \Delta a$ because of GWs emission, while $T_{\rm HVS}$ is the time the binary would have taken if it was driven by hardening. The times $T_{\rm HVS}$ and $T_{\rm GW}$ are computed, respectively, integrating equations (2.32) and (2.33).

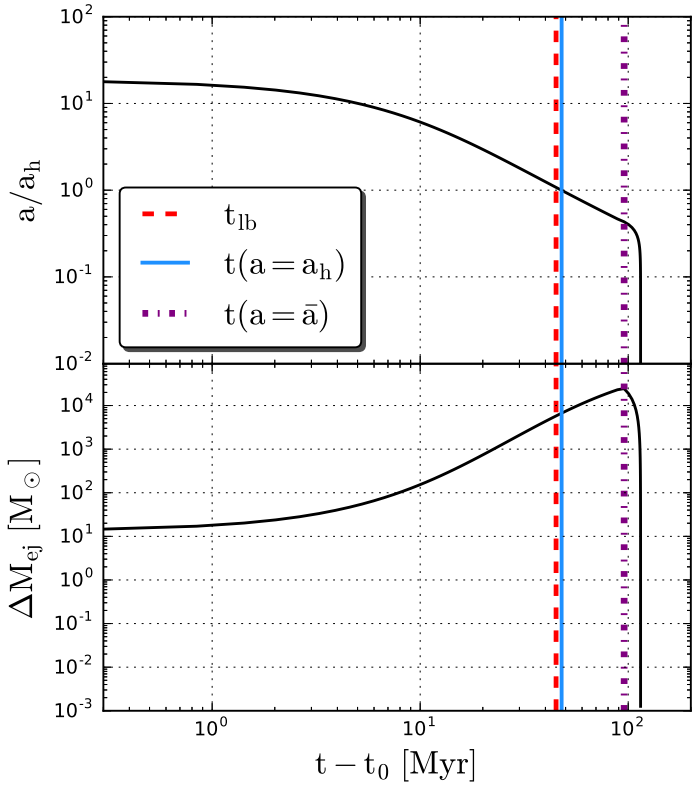


Figure 2.12: Time evolution of the MBHB binary separation (in units of a_h, top panel), computed integrating equations (2.32) and (2.33), and of the ejected stellar mass (bottom panel), computed using equation (2.30).

Creating the Mock Catalogue

For each ejected mass bin $\Delta M_{\rm ej}$, see equation (2.30), we derive the corresponding number of HVSs ΔN as:

$$\Delta N = \frac{\Delta M_{\rm ej}}{\int_{M_{\rm min}}^{M_{\rm max}} M f(M) dM}, \tag{2.34}$$

where $f(M)$ is the stellar mass function in the GC, $M_{\rm min} = 0.1$ M$_\odot$, and $M_{\rm max} = 100$ M$_\odot$. We then draw ΔN stars of mass M from a power-law mass function $f(M)$.

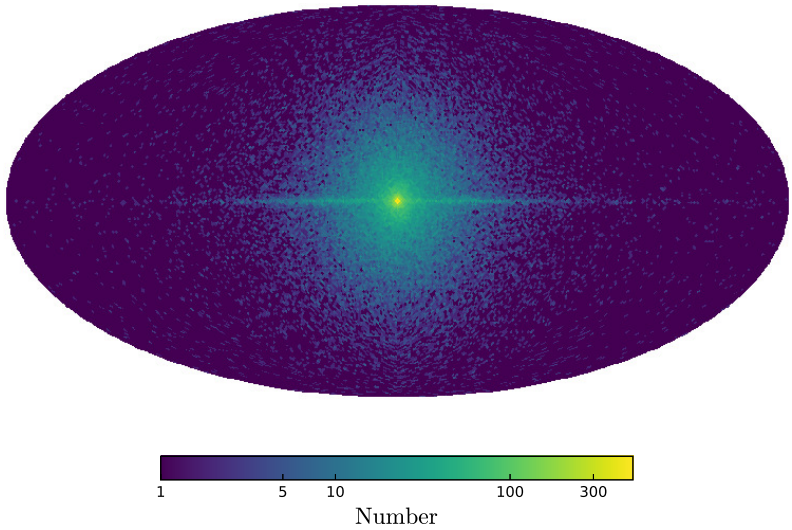


Figure 2.13: MBHB catalogue: sky distribution in Galactic coordinates of the current population of HVSs in our Galaxy (122473 stars).

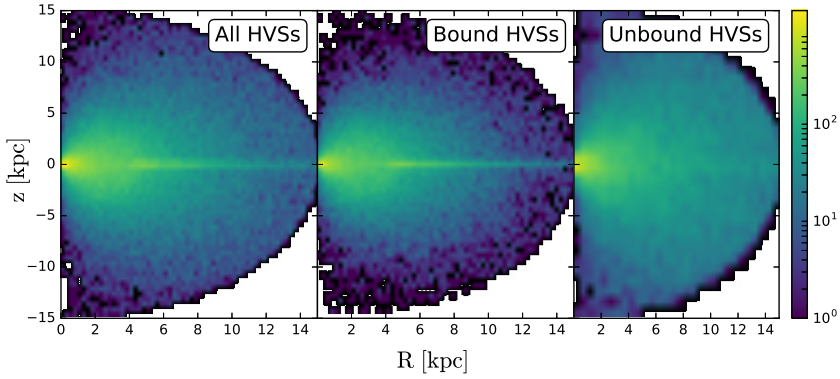


Figure 2.14: MBHB catalogue: distribution in Galactocentric cylindrical coordinates (R, z) of all HVSs (left), bound HVSs (centre), and unbound HVSs (right) within 15 kpc from the Galactic Centre.

We draw velocities from the velocity distribution (Sesana et al. 2006):

$$f(w) = \frac{A}{h}\left(\frac{w}{h}\right)^{\alpha}\left[1+\left(\frac{w}{h}\right)^{\beta}\right]^{\gamma}, \tag{2.35}$$

where $w \equiv v/v_c$, $v_c = \sqrt{G(M_\bullet + M_c)/a}$ is the binary orbital velocity, $h \equiv \sqrt{2q}/(1+q)$, $A = 0.236$, $\alpha = -0.917$, $\beta = 16.365$, and $\gamma = -0.165$ (Sesana et al. 2006). We note that in this scenario the ejection velocity does not depend on the mass of the HVS. We sample this velocity distribution using the MCMC sampler emcee (Foreman-Mackey et al. 2013). Velocities are drawn in the range $[v_{\rm min}, v_{\rm max}]$, $v_{\rm max} = v_c/(1+q)$ (Sesana et al. 2006). We fix $v_{\rm min}$ considering that we are only interested in stars with a velocity high enough to escape from the MW bulge. To be more quantitative, we only consider stars with a velocity v greater then the escape velocity from the radius of influence of the binary, $r_{\rm inf} \equiv 2GM/(2\sigma^2) \sim 1$ pc. Assuming the same bulge profile as discussed in Section 2.2.1, we get $v_{\rm min} = 645$ km s^{-1}, ~ 100 km s^{-1} higher than the one used in Sesana et al. (2006). We note that since a decreases with time, v_c (and therefore $v_{\rm max}$) increase as the binary shrinks: HVSs with the highest velocities will be ejected right before the merger of the two black holes, but the majority of HVSs will be ejected right before the rate of orbital decay is driven by GW emission (see discussion in Section 2.4.1).

For each star, we can compute the corresponding time of ejection after t_0: $\Delta t = t - t_0$, by integrating equation (2.32) (equation (2.33)) for $a > \bar{a}$ ($a < \bar{a}$). The flight time of a star is computed according to $t' = t_{\rm lb} - \Delta t$. The value of $t_{\rm lb}$ is chosen in such a way to match the observational estimate of 300 HVSs in the mass range $[2.5, 4]$ M$_\odot$ within 100 kpc from the GC. We find that we can match this value by assuming that the binary started to eject HVSs $t_{\rm lb} = 45$ Myr ago (see discussion in Section 2.4.2).

We then determine the initial condition of the orbit and we propagate each star in the Galactic potential, with the same procedure outlined in Section 2.3.3. In doing that, we assume for simplicity that the ejection of HVSs by the MBHB is isotropic. Photometry for each star is computed as in Section 2.2.2, using equation (2.27) to determine the age of each star, and *Gaia* errors on astrometry are estimated following Section 2.2.3.

The evolution of the MBHB binary is summarized in Fig. 2.12, where we plot the binary separation (top panel) and the ejected stellar mass (bottom panel) as a function of time. We highlight three key moments in the evolution of the system: the time at which it becomes a hard binary $t(a = a_h)$ (solid line), the time at which its evolution is driven by GW emission

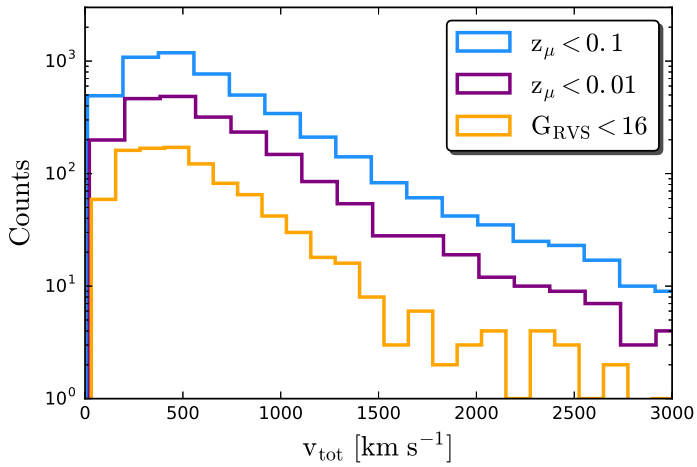


Figure 2.15: MBHB catalogue: total velocity (in the Galactocentric rest frame) of HVSs.

$t(a = \bar{a})$ (dot-dashed line), and the present time t_{lb} (dashed line). We can see that, to reproduce the estimates on the current population of HVSs, we are assuming that the MBHB in the GC has not yet shrunk to the hardening radius a_h, and that its evolution is still driven by dynamical hardening. Once GW emission dominates, the two black holes merge in a few Myr.

2.4.2 "MBHB" Catalogue: Number Estimates of HVSs in *Gaia*

Having created a catalogue of HVSs ejected by the MBHB, we can forecast how many of these HVSs we are expecting to find in the *Gaia* catalogue. We find a total of $N = 122266$ HVSs ejected from the MBHB, corresponding to a total stellar mass $M_{tot} \sim 3.7 \cdot 10^4\ M_\odot$. We note that this number is about of the same order of magnitude than the estimate made using equation (2.28) for the Hills catalogue.

The sky distribution of the population of HVSs is shown in Fig. 2.13. Fig. 2.14 shows the distribution of stars within 15 kpc from the GC in cylindrical coordinates (R, z). We can see that the distribution of unbound HVSs is isotropic, while for bound HVSs the distribution is slightly tilted towards $z = 0$, because of the torque applied by the stellar disc.

We find 59 % of these stars to fly along bound orbits, and the total velocity distribution of the stars is shown in Fig. 2.15 for the subset of stars which will be precisely measured by *Gaia*. Fig. 2.16 shows the cumulative

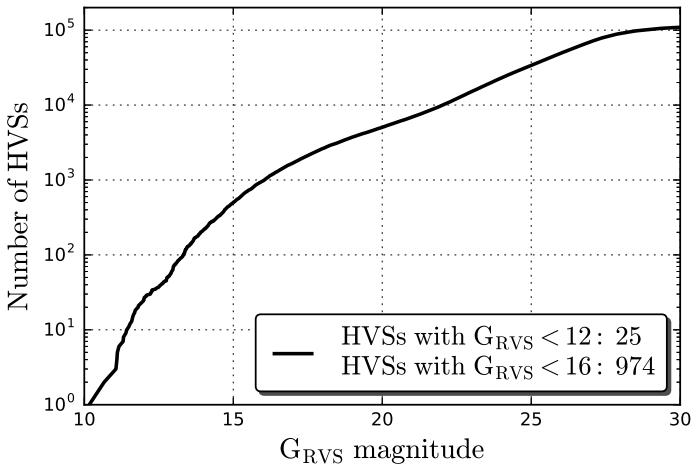


Figure 2.16: MBHB catalogue: cumulative distribution of HVSs magnitudes in the *Gaia* $G_{\rm RVS}$ passband.

distribution of magnitudes in the *Gaia* GRVS filter. A total of 974 (25) stars will be brighter than than the 16th (12th) magnitude, the magnitude limit for the final (second) data release of *Gaia*. If we focus on the $G_{\rm RVS} < 16$ stars, we find that 328 of them are unbound from the Milky Way, and that 527 of them have a total velocity higher than 450 km s^{-1}. We find 257 unbound HVSs with mass between 2.5 and 4 $M_\odot$ within 100 kpc from the GC, which agrees with the 300 HVSs estimated in Brown et al. (2014) and the estimate presented in Section 2.3.4. The distributions of errors in proper motions and parallax for the golden sample of HVSs with a three-dimensional velocity determination by *Gaia* alone is shown in Fig. 2.5.

We predict 12 of the 25 $G_{\rm RVS} < 12$ stars to be unbound from the Galaxy. Their typical relative error in proper motions is $\lesssim 0.01\%$, and in parallax is $\lesssim 40\%$, with 80% of the stars with $z_\varpi \lesssim 0.2$. These numbers have been corrected for the numerical factor introduced in Section 2.3.4.

The heliocentric distance (mass) distribution of HVSs in the catalogue with a precise astrometric determination by *Gaia* is shown in the top (bottom) panel of Fig. 2.17. Comparing these curves with the one obtained for the other mock catalogues, we can see that the shapes and the peak are reasonably similar, since they are shaped by the adopted mass function and stellar evolution model.

We can compare once more our estimates with results in Marchetti et al. (2017) for *Gaia* DR1/TGAS. We find a total of 2 HVSs with $V < 11$. Both of

these stars are unbound from the MW.

2.5 Prospects for the Current Sample of HVSs

In this section we assess the performance of *Gaia* in measuring the astrometric properties of the current observed sample of HVS candidates. Brown et al. (2015) measured proper motions with the *Hubble Space Telescope* (HST) for 16 extreme radial velocity candidates, finding that 13 of them have trajectories consistent with a GC origin within 2σ confidence levels, and 12 of them are unbound to the Milky Way. Proper motion accuracy is essential in constraining the origin of HVSs and is the main source of uncertainty in the orbital traceback, therefore we estimate *Gaia* errors on the total proper motion for this sample of HVS candidates.

For each star we determine the ecliptic latitude using equation (2.7). We find 10 of these 16 stars in *Gaia* DR1, from where we take *Gaia G* band magnitudes. All of the other stars but one (HE 0437-5439 = HVS3, Edelmann et al. 2005) have SDSS magnitudes, and we compute *Gaia G* band magnitudes according to the polynomial fitting coefficients in Jordi et al. (2010). Conversion from SDSS passbands to $(V-I_c)$ Johnson-Cousins color index is done using the fitting formula in Jordi et al. (2005). For HVS3, we estimate the G magnitude and the $(V-I_c)$ color from its B and V magnitude, according to Natali et al. (1994); Jordi et al. (2010). We then use PyGaia to estimate *Gaia* end-of-mission errors on the two proper motions for each star.

Fig. 2.18 shows the comparison between HST proper motions determination and *Gaia* estimates. In both cases we show the quadrature sum of the errors in the two proper motions. Stars with measurements consistent with coming from the GC are shown as red dots, while disk runaways are indicated as black dots, according to the classification presented in Brown et al. (2015). The black dashed line divides stars that will be detectable with a better accuracy than the current one: all stars but three (HVS1, HVS12, and HVS13) will have a better proper motion determination by *Gaia*. This will help reducing in size the errorbars and identifying the ejection location, confirming or rejecting the GC origin hypothesis. This will be crucial to test the alternative ejection model presented in Boubert & Evans (2016); Boubert et al. (2017b), where HVSs originate in the LMC.

We want once more to stress that these estimates refer to the final data release of the *Gaia* satellite, currently planned for 2022. Rescaling proper motion errors for the correcting factor ~ 4.8 introduced in Section 2.3.4,

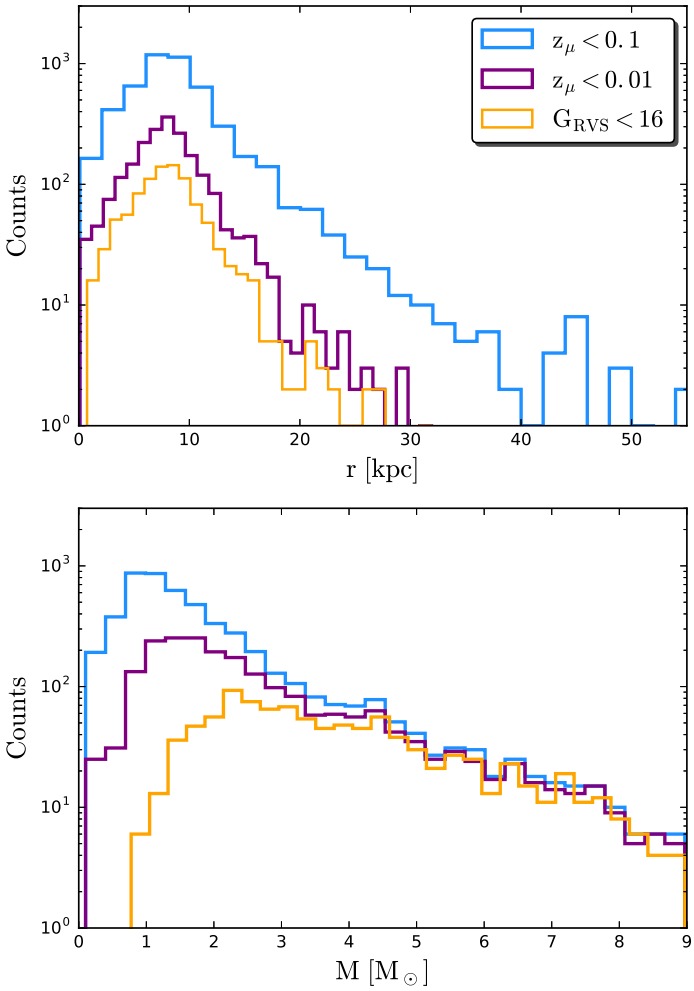


Figure 2.17: MBHB catalogue: heliocentric distance (top) and mass (bottom) distribution of the HVSs with a relative error on total proper motion below 10% (blue), 1% (purple), and with a radial velocity measurement by *Gaia* (yellow).

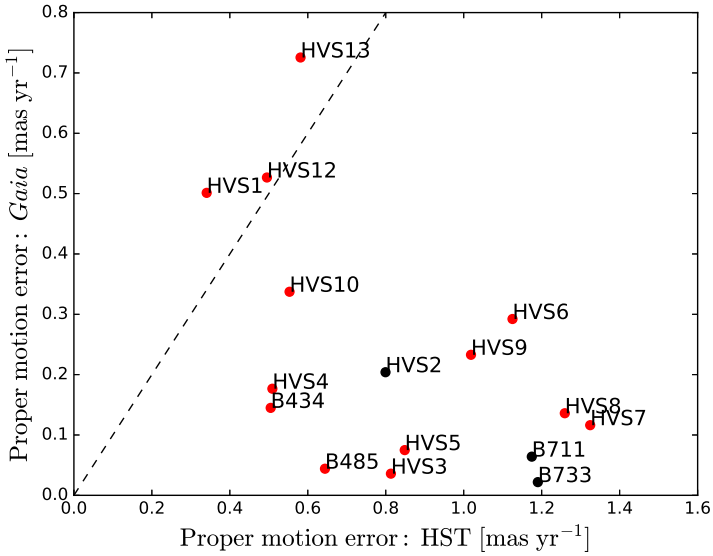


Figure 2.18: Expected performance of *Gaia* in measuring proper motions of the observed sample of candidates in Brown et al. (2015). Red dots correspond to stars with a trajectory consistent with a GC origin, while black dots are disk runaways. On the x axis we report the quadrature sum of the HST proper motion errors (Table 1 in Brown et al. 2015), while on the y axis the estimate obtained with PyGaia. Stars below the dashed line ($y = x$) will have a more precise proper motion determination in the final data release of the *Gaia* mission.

we find that 7 stars (the brightest in the sample) will have a better proper motion determination already in *Gaia* DR2: HVS3, HVS5, HVS7, HVS8, B485, B711, and B733.

2.6 Discussion and Conclusions

In this paper we build mock catalogues of HVSs in order to predict their number in the following data releases of the *Gaia* satellite. In particular, we simulate 3 different catalogues:

1. The Vesc catalogue does not rely on any assumption on the ejection mechanism for HVSs. We populate the Milky Way with stars on radial trajectories away from the Galactic Centre, and with a total velocity equal to the escape velocity from the Galaxy at their position. Therefore we only rely on the definition of HVSs as unbound stars, and we do not make any assumption on the physical process causing their acceleration. We then spatially distribute these stars assuming a continuous and isotropic ejection from the GC.

2. The Hills catalogue focuses on the Hills mechanism, the leading mechanism for explaining the origin of HVSs. Assuming a parametrization of the ejection velocity distribution of stars from the GC, we numerically integrate each star's orbit, and we self consistently populate the Galaxy with HVSs.

3. The MBHB catalogue assumes that HVSs are the result of the interaction of single stars with a massive black hole binary, constituted by Sagittarius A* and a companion black hole with a mass of $5 \cdot 10^3$ M$_\odot$. In this and in the previous catalogue there are bound HVSs: stars that escape the GC with a velocity which is not high enough to escape from the whole Galaxy. These are the result of modelling a broad ejection velocity distribution.

We characterize each star in each catalogue from both the astrometric and photometric point of view. We then derive the star magnitude in the *Gaia* passband filters and the *Gaia* measurement errors in its astrometric parameters. The aim is to assess the size and quality of the *Gaia* HVS sample.

As a summary and for quick consultation, our results for the size of three mock catalogues discussed in the paper are summarised in Table 2.2

Table 2.2: Number estimates of HVSs in the *final* data release of *Gaia*, for the three implemented catalogues of HVSs: VESC, HILLS, and MBHB. N_{tot} is the total number of HVSs in the Galaxy, $N(z_\mu < 0.1)$ ($N(z_\mu < 0.01)$) is the number of HVSs which will be detected by *Gaia* with a relative error on total proper motion below 10% (1%), $N(z_\varpi < 0.2)$ is the number of HVSs with a relative error on parallax below 20%, and N_{vrad} is the number of stars bright enough to have a radial velocity measurement. We remind the reader that the VESC catalogue, by construction, only includes unbound objects, while the HILLS and the MBHB catalogues contain both bound and unbound stars.

Catalogue	N_{tot}	$N(z_\mu < 0.1)$	$N(z_\mu < 0.01)$	$N(z_\varpi < 0.2)$	N_{vrad}
Vesc	17074	709	241	40	115
Hills	100000	11661	3765	568	2140
MBHB	122266	5066	2124	364	974

Table 2.3: Same as 2.2, but for predictions of HVSs in the *second* data release of *Gaia*.

Catalogue	N_{tot}	$N(z_\mu < 0.1)$	$N(z_\mu < 0.01)$	$N(z_\varpi < 0.2)$	N_{vrad}
Vesc	17074	357	81	20	2
Hills	100000	5963	781	261	19
MBHB	122266	2892	750	194	25

for the final *Gaia* data release, and in Table 2.3 for the second data release. Regardless of the adopted ejection mechanism, we can conclude that *Gaia* will provide an unprecedented sample of HVSs, with numbers ranging from several hundreds to several thousands. The peak of the mass distribution and the limiting heliocentric distance at which HVSs will be observed by *Gaia* are presented in Table 2.4. We can see that these values differ from the current sample of observed late B-type stars in the outer halo (refer also to Fig. 2.3, 2.11, 2.17). Most HVSs will have precise proper motion measurements, and therefore data mining techniques not involving the radial velocity information need to be developed in order to extract them from the dominant background of other stars in the MW (Marchetti et al. 2017). Stars with precise proper motions will be visible at typical heliocentric distances $r < 50$ kpc, while stars bright enough to have a radial velocity measurement from *Gaia* will typically be observed at $r < 30$ kpc, with a peak in the distribution for $r \sim 10$ kpc.

We estimate the precision with which *Gaia* will measure proper motions for the sample of HVSs candidates presented in Brown et al. (2015). Fig. 2.18 shows that the majority of HVSs will have a better proper motion determination by *Gaia*. This will help determining their ejection location, confirming or rejecting the Galactocentric origin hypothesis.

We now briefly discuss the impact of the assumptions made on the stel-

Table 2.4: Peak mass of the mass distribution and maximum heliocentric distance for the HVSs in the three different mock catalogues. The maximum heliocentric distance is defined as the distance at which we predict a total of 0.5 stars. Due to the small number of HVSs with a three-dimensional velocity in *Gaia* DR2, we choose not to characterize their distributions here.

Catalogue	$z_\mu < 0.1$	$z_\mu < 0.01$	$z_\varpi < 0.2$	vrad
Vesc	(1.0 M$_\odot$, 40 kpc)	(1.5 M$_\odot$, 25 kpc)	(2.5 M$_\odot$, 12 kpc)	(2.7 M$_\odot$, 25 kpc)
Hills	(1.2 M$_\odot$, 48 kpc)	(2.1 M$_\odot$, 20 kpc)	(2.9 M$_\odot$, 10 kpc)	(3.0 M$_\odot$, 18 kpc)
MBHB	(0.8 M$_\odot$, 41 kpc)	(1.4 M$_\odot$, 28 kpc)	(1.5 M$_\odot$, 12 kpc)	(2.3 M$_\odot$, 24 kpc)

lar population in the GC. The Vesc catalogue does not depend on the binary population properties, but only on the choice of the Galactic potential, which we fix to a fiducial model consistent with the latest observational data on the rotation curve of the MW. In the Hills catalogue, our choice for the binary distribution parameters $\alpha = -1$, $\gamma = -3.5$ is motivated by the fit of the sample of unbound late B-type HVSs to the velocity distribution curve modelled using the Hills mechanism (Rossi et al. 2017). We repeat the same analysis presented in Section 2.3 adopting $\gamma = 0$: a flat distribution of binary mass ratios. This choice implies a higher mass for the secondary star in the binary, compared to the steeper value of $\gamma = -3.5$. Given the mass dependency of equation (2.19), this results in high total velocities for binaries in which the HVSs is the primary star. This in turn implies, on average, a larger number of HVSs with higher mass, which will be observed by *Gaia* to higher heliocentric distances with lower relative errors. Nevertheless, the final estimates on the number of HVSs we are expecting to be found in the *Gaia* catalogue are consistent with results presented in Section 2.3. A choice of a top-heavy initial mass function for stars in the GC (e.g. Bartko et al. 2010; Lu et al. 2013) would produce similar results. As a further check, we study the impact of adopting Galactic binary properties, which can be significantly different than in star forming regions, such as 30 Doradus in the LMC or the GC (Duchêne & Kraus 2013; Sana et al. 2013; Kobulnicky et al. 2014). In particular, we choose to change our prescription for solar mass HVSs, which are the majority of stars in our simulations. From equation (2.19), we can see that, for an equal mass binary ($q = 1$) with $M = 1$ M$_\odot$, the maximum initial separation needed in order to attain ejection velocity of 680 km s^{-1}is $a_{\rm max} \sim 100$ R$_\odot$. This choice of ejection velocity, given our adopted model for the Galactic potential, is the minimum velocity needed for a star in the GC to reach the Sun position with zero velocity. This maximum binary separation corresponds to a max-

imum orbital period $P_{\rm max} \sim 90$ days. For solar-type primaries ($m_p < 1.2$ $M_\odot$) in binaries with periods shorter than $P_{\rm max}$, the mass ratio distribution can be approximated as a broken power-law, with indexes $\gamma_{{\rm small}q} = 0.3$ (for $0.1 < q < 0.3$) and $\gamma_{{\rm large}q} = -0.5$ (for $0.3 < q < 1.0$) (Moe & Di Stefano 2017). The period distribution is flat with very good approximation in this restricted period range (see Figure 37 in Moe & Di Stefano 2017). Moreover, solar mass stars are single twice as often as *B*-type stars (Moe & Di Stefano 2017), therefore, when we draw primary masses from the Kroupa mass function, we select stars with $m_p < 1.2$ $M_\odot$ only 50% of the times. With these prescriptions, using equation (2.28) with this updated dN/dM we again obtain $N_{\rm tot} \simeq 1 \cdot 10^5$. Because of the lower number of solar mass stars in binary systems, we now find the mass distribution to peak around $1.5\,M_\odot$ for stars with precise proper motions by *Gaia*. Apart from this, number estimates agree extremely well with results presented in Section 2.3.4. Constructing the MBHB catalogue it is also worth exploring different values for the mass of the secondary black hole, which we fixed to $5 \cdot 10^3$ $M_\odot$. Higher (lower) masses result in a larger (smaller) total mass ejected by the binary (see equation (2.30)). Tuning the value of $t_{\rm lb}$, the loockback time at which the MBHB started ejecting HVSs, it is then possible to find different values of the secondary mass which are consistent with the observational estimate given by Brown et al. (2014). Regardless of $t_{\rm lb}$, we find $M_c = 1000$ $M_\odot$ to be a lower limit on the black hole mass to be able to observe 300 HVSs in the observed mass range $[2.4, 5]$ $M_\odot$, within 300 kpc from the GC. The possibility of considering multiple merging events, and/or a full parameter space exploration to break the degeneracy between M_c and $t_{\rm lb}$ are beyond the scope of this paper. An improvement over this catalogue would consist in modelling the ejection angles of HVSs as a function of the decreasing binary separation.

Although a full investigation of the detection strategy of HVSs is beyond the scope of this paper, it is interesting to qualitatively compare our findings with the expected major sources of sample contamination. HVSs may be confused with *runaway* stars: stars ejected with high velocities by dynamical encounters in dense stellar systems (Poveda et al. 1967; Portegies Zwart 2000) or by the explosion of a supernova in a binary star (Blaauw 1961; Tauris & Takens 1998). These stars are produced in star forming regions in the stellar disk of the Milky Way, but, given their high velocity, they can travel to the stellar halo (Silva & Napiwotzki 2011). The *Gaia* catalogue will contain $\sim 10^9$ disk stars (Robin et al. 2012). Assuming rates and the velocity distribution in Silva & Napiwotzki (2011), we can estimate a

total of $N_{\rm RS} \sim 10^5$ runaway stars in the *Gaia* catalogue with $v > 400$ km s^{-1}, two order of magnitudes more than the predicted number of HVSs. Nevertheless, the rate of ejection of unbound objects is estimated to be approximately one for every 100 HVSs (Brown 2015), with velocities that can reach up to ~ 1300 km s^{-1}for companion stars in a binary disrupted via an asymmetric supernova explosion (Tauris 2015). Precise proper motions and radial velocities provided by *Gaia* will help discriminating these stars, by tracing back their orbits to determine the ejection location (GC or stellar disk). High velocity halo stars on radial orbits could also be easily mistaken for bound HVSs because of their similar trajectories. To estimate the contamination of such stars to the sample of bound HVSs, we start considering that we are expecting $\sim 10^7$ halo stars in the *Gaia* catalogue (Robin et al. 2012). We estimate a total of $\sim 10^5$ halo stars with a total velocity vector pointing inside the solid angle subtending a cone with base radius of 500 pc around the GC when traced back in time. Given the typical velocity dispersion of stars in the stellar halo $\sim \sqrt{3}\cdot 150$ km s^{-1}(Smith et al. 2009; Evans et al. 2016), we expect ~ 2000 halo stars on radial trajectories from the GC with $v > 400$ km s^{-1}. Further stellar properties, such as metallicity, need to be considered in order to correctly classify those stars (e.g. Hawkins et al. 2015; Zhang et al. 2016).

To summarize, the sample of known HVSs will start increasing in number in April 2018 with DR2, with a few tens of stars with a precise three-dimensional velocity by *Gaia* alone. This sample will already be comparable in size with the current tens of HVSs candidates, but the largest improvement in terms of stars with full three-dimensional velocity will come with the final *Gaia* data release, with hundreds of stars unbound from the Milky Way. The majority of HVSs in *Gaia* will not have radial velocities from *Gaia*, therefore dedicated spectroscopic follow-up programs with facilities such as 4MOST (de Jong et al. 2016) and WEAVE (Dalton 2016) will be necessary to derive their total velocity and to clearly identify them as HVSs.

Acknowledgements

We thank Warren Brown for the careful reading of the manuscript and the useful comments. TM and EMR acknowledge support from NWO TOP grant Module 2, project number 614.001.401. AS is supported by the Royal Society. This research made use of Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013). All fig-

ures in the paper were produced using matplotlib (Hunter 2007). This work would not have been possible without the countless hours put in by members of the open-source community all around the world.

3 An artificial neural network to discover hypervelocity stars: candidates in *Gaia* DR1/TGAS

T. Marchetti, *E.M. Rossi, G. Kordopatis, A.G.A. Brown, A. Rimoldi, E. Starkenburg, K. Youakim, R. Ashley* *2017, MNRAS, 470, 1388-1403*

The paucity of hypervelocity stars (HVSs) known to date has severely hampered their potential to investigate the stellar population of the Galactic Centre and the Galactic potential. The first *Gaia* data release (DR1, 2016 September 14) gives an opportunity to increase the current sample. The challenge is the disparity between the expected number of HVSs and that of bound background stars. We have applied a novel data mining algorithm based on machine learning techniques, an artificial neural network, to the Tycho–*Gaia* astrometric solution catalogue. With no pre-selection of data, we could exclude immediately ~ 99 per cent of the stars in the catalogue and find 80 candidates with more than 90 per cent predicted probability to be HVSs, based only on their position, proper motions and parallax. We have cross-checked our findings with other spectroscopic surveys, determining radial velocities for 30 and spectroscopic distances for five candidates. In addition, follow-up observations have been carried out at the Isaac Newton Telescope for 22 stars, for which we obtained radial velocities and distance estimates. We discover 14 stars with a total velocity in the Galactic rest frame > 400 km s^{-1}, and five of these have a probability of > 50 per cent of being unbound from the Milky Way. Tracing back their orbits in different Galactic potential models, we find one possible unbound HVS with $v \sim 520$ km s^{-1}, five bound HVSs and, notably, five runaway stars with median velocity between 400 and 780 km s^{-1}. At the moment, uncertainties in the distance estimates and ages are too large to confirm the nature of our candidates by narrowing down their ejection location, and we wait for future Gaia releases to validate the quality of our sample. This test successfully demonstrates the feasibility of our new data-mining routine.

3.1 Introduction

Observationally, hypervelocity stars (HVSs) are stars that can reach radial velocities in excess of the Galactic escape speed at their location, and whose trajectories are consistent with a Galactic Centre (GC) origin (Brown et al. 2005). Currently, about ~ 20 unbound stars have been discovered (**?**): most of them are late B-type stars ($\sim 2.5 - 4$ $M_\odot$) detected in the outer halo (but note Zheng et al. 2014) with velocities between $\sim 300 - 700$ km s^{-1}(see Brown 2015, for a review). These stars are in principle unique tools to gather information on the Galactic Centre stellar population and dynamics (Madigan et al. 2014; Zhang et al. 2013, e.g.) and on the Galactic potential (e.g. Gnedin et al. 2005; Yu & Madau 2007; Perets et al. 2009). Using current data, a first proof of principle of how to get joint constraints on both environments was published in Rossi et al. (2017), and attempts to constrain the dark matter halo alone were performed by Sesana et al. (2007) and Fragione & Loeb (2017)[1]. These analyses however are severely hampered by the quality and quantity of the current small and rather biased sample.

So far the most successful observational strategy has been to spectroscopically select late B-type stars in the outer halo. Since the stellar halo is dominated by an old stellar population, young stars likely come from other star-forming regions in the Galaxy, and a late B-type star has a long enough life-time ($\sim 100 - 300$ Myr) to be able to travel to the outer halo from the Galactic Centre if its velocity is hundreds km s^{-1}. Most of the confirmed unbound HVSs have only radial velocity measurements and uncertainties in their photometric distances are large. Proper motions have been acquired with the Hubble Space Telescope for 16 high velocity stars (Brown et al. 2015), but even if the GC origin was confirmed for 13 of these objects, uncertainties are still too large to precisely constrain their origin, and therefore to identify them as HVSs.

Recent years have seen an increasing effort to identify low mass HVSs in the inner Galactic halo. These searches use high proper motion or high radial velocity criteria, as it is not possible to spectroscopically single out these low mass stars in the halo, as is done for B-type HVSs. A few tens of candidates have been reported, but the large majority are bound and/or consistent with Galactic disc origin (e.g. Li et al. 2012; Palladino et al. 2014; Ziegerer et al. 2015; Vickers et al. 2015; Hawkins et al. 2015; Zhang et al.

[1] See also Gnedin et al. (2010), who uses the velocity dispersion of halo stars from the hypervelocity star survey.

2016; Ziegerer et al. 2017). Positive identification is prevented by large distance and proper motion uncertainties.

Major observational advancements in the field are therefore expected from the data taken by the ESA mission *Gaia*, launched on the 19th of December 2013 (Gaia Collaboration et al. 2016b,a). *Gaia* will attain an unparalleled astrometric measurement precision for a total of $\sim 10^9$ stars in the Galaxy. In the end-of-the-mission data release, we anticipate a few hundred (a few thousand) HVSs within 10 kpc from us, in the mass range $\sim 1 - 10$ $M_\odot$, with relative error on total proper motion $< 1\%$ ($< 10\%$), and that radial velocities will be measured for a subsample of these (Marchetti et al. in preparation). For brighter HVSs, accurate *Gaia* parallaxes can eliminate the large distance uncertainties in the existing sample, and for fainter stars calibrated photometric distances may eventually be used.

The first data release (DR1) happened on September 14, 2016, and it contains the five-parameter astrometric solution (positions, parallaxes, and proper motions) for a subset of $\sim 2 \times 10^6$ stars in common between the Tycho-2 Catalogue and *Gaia* (TGAS catalogue, Michalik et al. 2015; Lindegren et al. 2016). Radial velocity information is notably missing. Our expectation is that between $0.1-$ and a few unbound HVSs may be expected to be present in the catalogue, depending on the unknown mass distribution and star formation history in the Galactic Centre (Marchetti et al. in preparation).

In this paper, we report a systematic search for HVSs in DR1. We use an artificial neural network (Section 3.2), which is first applied to the TGAS subset of the *Gaia* catalogue without any prior constraints placed on stellar properties to select HVS candidates (Section 3.3). We then cross check our sample of best candidates with published spectral catalogues to acquire radial velocity and spectroscopic distance information (Section 3.4). We further proceed to describe the radial velocity follow-up observations for candidates with no published radial velocity and observable by the Isaac Newton Telescope (INT) (Section 3.4.2). In Section 3.5 we describe our Bayesian approach to determine distances, and then in Section 3.6 we present our results for HVS candidates in terms of total velocity and ejection location. We sort and characterize candidates in Section 3.7, and discuss their implications in Section 3.8.

3.2 Data Mining algorithm

Hypervelocity stars are rare objects, that occur in the Galaxy at an uncertain rate roughly between $10^{-5} - 10^{-4}$ yr^{-1} (Hills 1988; Perets et al. 2007; Zhang et al. 2013; Brown et al. 2014). Considering the magnitude limit of *Gaia* and different assumptions on the population of binaries in the GC, such a rate implies only $\sim 0.1 - 1$ HVSs for every 10^6 stars in the final *Gaia* catalogue (Marchetti et al. in preparation). In particular for the TGAS catalogue, we expect to find at most a few HVSs (Marchetti et al. 2018b), although a larger number of slower stars generated via the same mechanism (called *"bound HVSs"*) are also expected (Bromley et al. 2006; Kenyon et al. 2008). Thus, *Gaia* can deliver a HVS sample that represents a huge leap in data quality and quantity, but building it requires careful data mining, especially since radial velocity measurements are currently missing.

The TGAS subset of *Gaia* DR1 provides the five-parameter astrometric solution for roughly two million objects, therefore we choose to build a data mining routine based only on the astrometric properties of the stars: position on the sky (α, δ), parallax ϖ, and proper motions $\mu_{\alpha*}, \mu_\delta$. This approach allows us to not make any a priori assumption on the stellar nature of HVSs, avoiding photometric and metallicity cuts which might bias our search towards particular spectral types, and lead to a sample which may not reflect the properties of the binary population in the Galactic Centre. Recent studies have shown indeed how the GC is a complex environment in which different stellar populations coexist and interact, and many properties (mass function, metallicity, binarity) are missing or poorly constrained due to observational limitations (see Genzel et al. (2010) for an exhaustive review). The nuclear star cluster surrounding the central massive black hole has also undergone several star formation episodes throughout its lifetime, which might have changed and influenced the stellar population and mass function (Genzel et al. 2010; Pfuhl et al. 2011).

We have therefore chosen to build a data mining routine based on a machine learning algorithm, an *artificial neural network*. Our chosen approach is a *supervised learning* problem: we present the algorithm with examples and their desired output (*training set*), and we let the algorithm learn the best function mapping inputs into outputs. We decided for a binary classification problem: the desired output of the algorithm is 0 for a "normal" background star, and 1 for a HVS. When we apply the classification rule to a new unlabelled example we can then interpret its output as the probability of that star being a HVS (Saerens et al. 2002).

We now start introducing neural networks, with a brief summary on the main idea behind this algorithm. Next in Section 3.2.2 we discuss how we build our training set, and finally in Section 3.2.3 and Section 3.2.4 how we optimize and determine the performance of the network based on the results on subsets of the data which were not used for the training.

3.2.1 Artificial Neural Networks

Artificial neural networks have been largely used in different branches of science for their ability to provide highly non-linear mapping functions, and for their intrinsic capacity to generalize: to provide reasonable outputs for examples not encountered while training the algorithm (see Haykin 2009, for an exhaustive explanation of neural networks). This latter property is particularly important for our goal, since our training set consists of mock data (see Section 3.2.2), and therefore we want to be flexible enough to find HVSs even if the real population is not perfectly represented by our simulations, which necessarily rely on several hypotheses and assumptions (see Section 3.2.2).

We have developed from scratch an artificial neural network with five input units (the astrometric parameters), two hidden layers of neurons, and a single output neuron for binary classification. Each neuron of the network is a computational unit which outputs a non-linear function[2] $f(v)$, where v is a linear combination of the j-th input M-dimensional vector $\boldsymbol{x}^{(j)}$ with some weight vector $\boldsymbol{\omega}$:

$$v_j(\boldsymbol{x}^{(j)};\boldsymbol{\omega}) = x_0\omega_0 + \sum_{i=1}^{\mathrm{M}} x_i^{(j)}\omega_i, \tag{3.1}$$

where $x_0 \equiv 1$ is referred to as the *bias unit*. In analogy with the brain architecture, the components ω_i are usually referred to as *synaptic weights*. A typical choice for f is a sigmoid function. We choose:

$$f(v) = a\tanh(bv), \tag{3.2}$$

with $a = 1.7159$ and $b = 2/3$. This activation function outputs real numbers in the interval $[-a, a]$, and satisfies several useful properties: it is an odd function of its argument; $f(1) = 1$ and $f(-1) = -1$; its slope at the origin is close to unity; and its second derivative attains its maximum value at

[2] In the following, we will use superscripts in round brackets to refer to a particular vector, and subscripts to specify its components.

$x = 1$. This choice has been shown to yield faster convergence than the usual logistic function, avoiding driving the hidden neurons into saturation (LeCun 1993).

For neurons in the first hidden layer the input $\boldsymbol{x}^{(j)}$ is just the data vector containing the $\mathrm{M} = 5$ astrometric parameters for the j-th training example: $\boldsymbol{x}^{(j)} = (\alpha_j, \delta_j, \varpi_j, \mu_{\alpha*j}, \mu_{\delta j})$, therefore the summation in Equation 3.1 extends over $i = 1, \ldots, 5$. For neurons in the second layer the input $\boldsymbol{x}^{(j)}$ is the M_1-dimensional vector output by the first layer of M_1 neurons, and the summation extends to $\mathrm{M} = \mathrm{M}_1$. Finally, the single neuron in the output layer takes in input a M_2-dimensional vector, with M_2 equal to the number of neurons in the second hidden layer, and in summation $\mathrm{M} = \mathrm{M}_2$. We call $D_j(\boldsymbol{\omega}) \in \mathbb{R}$ the final output of the neural network for the j-th example.

The training process consists in finding the vector of synaptic weights $\boldsymbol{\omega}$ which minimizes the total cost function

$$J(\boldsymbol{\omega}) \propto \sum_{j=1}^{N} (D_j(\boldsymbol{\omega}) - y_j)^2, \tag{3.3}$$

which is just the sum over all the N examples of the squared difference between the output of the neural network $D_j(\boldsymbol{\omega})$ and the desired output y_j of the labelled training example. The value of each synaptic weight is initialized with a random number drawn from a uniform distribution in the interval $[-\sigma_\omega, \sigma_\omega]$, with $\sigma_\omega = m_*^{-1/2}$, where m_* is the number of connections feeding into the corresponding layer of neurons (LeCun et al. 2012). The weights optimization is then performed with an adaptive stochastic (online) gradient descent method, using a specific learning rate η_k for each synaptic weight: the AdaGrad implementation (Duchi et al. 2011). We use the following iterative rule for the t-th update of the k-th weight ω_k (Singh et al. 2015):

$$\Delta\omega_k(t) = -\eta_k(t) g_k(t) = -\frac{\eta_0}{\sqrt{\sum_{i=1}^{t} (g_k(i))^2}} g_k(t), \tag{3.4}$$

where $\eta_0 > 0$ is called the *global learning rate*, $\boldsymbol{g}$ is the gradient of the cost function in Equation 3.3 (derivatives with respect to the weight vector $\boldsymbol{\omega}$), and the denominator is the norm of all the gradients of the previous iterations. The adopted value for η_0 is discussed in Section 3.2.3, while the gradient of the cost function is estimated with a back-propagation algorithm (see LeCun et al. (2012) for tips on an efficient implementation, essential when dealing with large datasets).

3.2.2 Building the Training Set

We train the artificial neural network on a simulated end-of-mission *Gaia* catalogue for the Galaxy: the *Gaia* Universe Model Snapshot (GUMS, Robin et al. 2012), where we inject *mock* HVS data with errors on all astrometric and photometric measurements. A detailed description of how we construct our mock HVS will be the focus of an upcoming paper, and here we only briefly summarize our procedure. In the following we will adopt the Hills mechanism for modelling our mock population of HVSs, involving the disruption of a binary star by the Massive Black Hole (MBH) at the centre of our Galaxy (Hills 1988).

We explore the space (l, b, d, M) to populate each position in Galactic coordinates on the sky (l, b) with stars in a mass range $M \in [0.1 - 9]\ M_\odot$ and in a distance range $d \in [0, 40]$ kpc from us. We adopt a step of $\sim 9^\circ$ in Galactic longitude l, $\sim 4.5^\circ$ in Galactic latitude b, ~ 1 kpc in distance r, and $\sim 0.2\ \mathrm{M}_\odot$ in mass. We draw velocities from an ejection velocity distribution which analytically depends on the properties of the original binary approaching the massive black hole (Sari et al. 2010; Kobayashi et al. 2012; Rossi et al. 2014)[3]:

$$v_{\mathrm{ej}} = \sqrt{\frac{2Gm_c}{a}} \left(\frac{M_\bullet}{m_T} \right)^{\frac{1}{6}}, \tag{3.5}$$

where m_c is the mass of the star that remains bound to the MBH after the binary is disrupted, $m_T = M + m_c$ is the total mass of the disrupted binary, and $M_\bullet = 4.0 \times 10^6\ \mathrm{M}_\odot$ is the mass of the MBH in our Galaxy (Ghez et al. 2008; Gillessen et al. 2009; Meyer et al. 2012). Following Rossi et al. (2014, 2017), we model binary distributions for semi-major axis a and mass ratio q as power-laws: $f_a \propto a^\alpha$, $f_q \propto q^\gamma$, with exponents $\alpha = -1$ (Öpik's law, Öpik 1924) and $\gamma = -3.5$. This combination has been shown to result in a good fit between the observed sample of late B type HVSs in Brown et al. (2014) and the prediction of the Hills mechanism for reasonable choices of Milky Way potentials (Rossi et al. 2017). The total velocity v of the HVS is then computed decelerating the star in a given Galactic potential (refer to Section 3.6.2, Equations 3.12-3.14 for details on the adopted fiducial Milky Way potential).

For each star we compute the combination of proper motions and radial velocity which are consistent with an object moving radially away from

[3] Rigorously, there should be a numerical factor in front of Equation 3.5, depending on the detailed geometry of the three-body encounter. This factor has been shown to be ~ 1 when averaged over the binary's phase (Rossi et al. 2014).

the Galactic Centre, and we correct those values for the motion of the Sun and of the local standard of rest (LSR) (Schönrich 2012). We then roughly estimate the flight time from the GC to the given position in Galactocentric coordinates $r_{\rm GC}$ as $t_F = r_{\rm GC}/v_F$, where v_F is an effective velocity equal to the arithmetic mean between the ejection velocity and the decelerated velocity at the star's position. The age of the star is then computed summing the flight time and the age of the star at its ejection. The latter is computed as a random fraction of its main sequence (MS) lifetime (Brown et al. 2014), and the time spent on the MS is computed using analytic formulae in Hurley et al. (2000). We assume a super-solar metallicity $[M/H] = 0.4$, which corresponds to the mean value of the distribution in the GC (Do et al. 2015). Each star is evolved up to its age using the fast parametric stellar evolution code SeBa (Portegies Zwart & Verbunt 1996; Portegies Zwart et al. 2009) to obtain its radius, effective temperature, and mass, which we use to identify the best-matched stellar spectrum from the BaSeL 3.1 stellar spectral energy distribution (SED) libraries (Westera & Buser 2003) via chi-squared minimization. For each position of the sky we assess dust extinction using a three-dimensional Galactic dust model (Drimmel et al. 2003), and integrating the reddened flux in the respective passbands we estimate the magnitudes in the *Gaia G* band and in the Johnson-Cousins V, I_c bands. We finally use the python toolkit PyGaia[4] to estimate the errors on the astrometry with which *Gaia* would observe these objects. The errors are functions of the magnitude of the star, its color index $V-I_c$, and the ecliptic latitude β, the latter determining the number of observations of the object according to the satellite's scanning strategy.

Parallax and proper motions of each source are then replaced by drawing a random number from a Gaussian distribution centred on the nominal value and with standard deviation equal to the estimated uncertainty. This approach has two main advantages: it allows us to obtain negative parallaxes (which are present in the real *Gaia* catalogue) for faint objects with non-negligible relative errors on parallax; and it helps us mitigate the effect of the spatial grid in distance used for generating mock stars, preventing the algorithm from driving the learning rule towards discrete, fixed values in parallax.

We can therefore build a mock catalogue of HVSs, which we use for the training of the artificial neural network. We combine mock positions, parallaxes and proper motions of HVSs and "normal" background stars randomly picked from the GUMS in a single stellar catalogue, consisting of a

[4] `https://github.com/agabrown/PyGaia`

total of $\sim 2.5 \times 10^6$ objects ($\sim 25\%$ HVSs, label = 1; $\sim 75\%$ *Gaia* stars, label = 0). We randomly split stars of the catalogue into a *training* set ($\sim 60\%$ of the catalogue), a *cross-validation* set ($\sim 20\%$ of the catalogue), and a *test* set ($\sim 20\%$ of the catalogue). The training set consists of the examples the algorithm will learn from, the cross-validation set is used to optimize hyperparameters (see Section 3.2.3), while we use the test set to determine the performance of the neural network (see Section 3.2.4). The use of different examples for performing these tasks is extremely useful to prevent overfitting and to ensure generalization. All features (five parameters) of the complete catalogue have been scaled in such a way to have mean of 0 and variance of 1, to achieve a faster convergence of the stochastic gradient descent algorithm (LeCun et al. 2012).

3.2.3 Optimization of the Algorithm

The effectiveness of a neural network, as the majority of machine learning algorithms, critically depends on the choice of the so-called *hyperparameters*, several parameters that need to be carefully tuned in order to achieve the best compromise between the algorithm performance, the time needed for its training, and its ability to generalize to new input data. We identify three hyperparameters in our algorithm: the number of neurons in the first hidden layer M_1, the number of neurons in the second hidden layer M_2, and the global learning rate η_0 for the adaptive stochastic gradient descent (see Equation 3.4).

A systematic grid search in the hyperparameter space to determine the best combination is not feasible because of time limitations and computational power. We use the pyswarm[5] implementation of a Particle Swarm Optimization (PSO) algorithm (Kennedy & Eberhart 1995) to explore the space (M_1, M_2, η_0) with 20 test particles. The algorithm iteratively adjusts particles' positions towards the minimum value attained by the cost function, with a velocity proportional to the distance from this extremum. Since each iteration involves the full training of the algorithm in order to determine the value of the cost function, we choose to apply PSO to a limited sample of the training set (1000 random training examples), and then we select the combination of parameters which results in the best performance on the full cross-validation set, defined in terms of the Matthews correlation coefficient MCC (Matthews 1975, see next subsection). The PSO algorithm converges to the following values: $M_1 = 119$, $M_2 = 95$, $\eta_0 = 0.071$[6].

[5] `https://github.com/tisimst/pyswarm/`

[6] We initially included the *regularization parameter* λ as a 4th hyperparameter, but due

3.2.4 Performance of the Algorithm

As mentioned before, we choose a stochastic gradient descent optimization to minimize the global cost function. Because of the intrinsic randomness of this algorithm, we train the neural network several times on the complete training set, shuffling the order of the presented example units during each training. Plotting learning curves (the value of the cost function versus the number of training examples presented to the network), we find that 8 complete iterations are enough to reach a minimum in both the training and cross-validation cost functions, again confirming that overfitting is not an issue.

We determine the performance of the algorithm on the test set by computing two different error metrics: the *Matthews correlation coefficient* MCC (Matthews 1975) and the F_1 *score*. Calling TP and TN (FP and FN) respectively the number of true (false) positives and negatives of the confusion matrix on the test set, error metrics are computed as:

$$\mathrm{F_1} \equiv 2\frac{\mathrm{PR}}{\mathrm{P}+\mathrm{R}}, \tag{3.6}$$

$$\mathrm{MCC} \equiv \frac{\mathrm{TP\,TN} - \mathrm{FP\,FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}, \tag{3.7}$$

where P and R are called, respectively, *precision* and *recall*, and they are defined as $\mathrm{P} \equiv \mathrm{TP}/(\mathrm{TP}+\mathrm{FP})$, $\mathrm{R} \equiv \mathrm{TP}/(\mathrm{TP}+\mathrm{FN})$. The $\mathrm{F_1}$ score assumes values in $[0, 1]$ while the MCC in $[-1, 1]$, and in both cases a value of 1 corresponds to a perfect classifier (diagonal confusion matrix). At the end of the training, we obtain the following values on the test set: $\mathrm{F_1} \sim \mathrm{MCC} \simeq 0.95$.

3.3 Application to *Gaia* DR1

Once we have fully trained the neural network on the training set, determining the optimal values for the synaptic weights, we apply the classification rule to real unlabelled data to search for HVS candidates. The application of the neural network to the full TGAS subset of *Gaia* DR1 (2057050 sources) results in 22263 stars with a predicted probability $> 50\%$ of being a HVS, $\sim 1\%$ of the original dataset. The histogram of the output probability D given by the neural network on the full TGAS catalogue is shown

to time limitation with the PSO we decided to discard it, since several tests showed that it always converged to values close to zero. A value $\lambda \sim 0$ is an indication that the algorithm is not overfitting the training set.

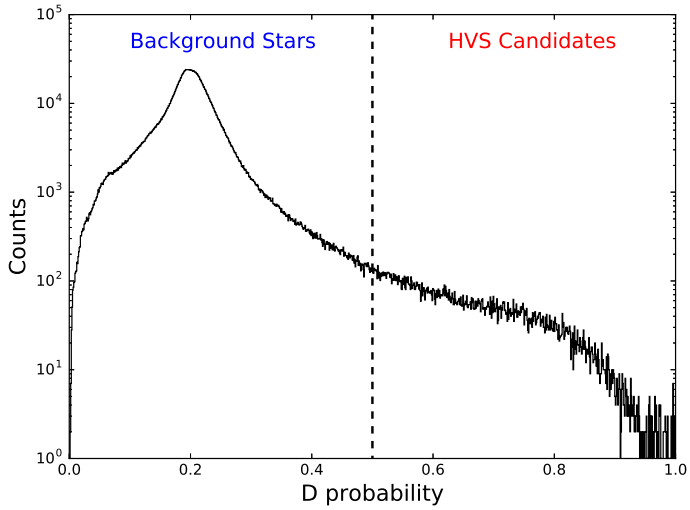


Figure 3.1: Histogram of the probability D of an object of being a HVS (output of the neural network), for all ~ 2 million stars in the TGAS subset of *Gaia* DR1. A dashed vertical line marks the decision boundary $D = 0.5$.

in Figure 3.1. To further reduce the sample of HVS candidates and to have reliable distance determinations, we filter out stars with a relative error on parallax $|\sigma_\varpi/\varpi| > 1$, obtaining a total of 8175 objects ($\sim 0.4\%$ of the original catalogue).

In these first cuts no information on the measured uncertainties is used to determine the probability of a star being a HVS. We subsequently include errors with a Monte Carlo (MC) simulation, randomly drawing one thousand realizations of the astrometry (parallax and proper motions) of each star from a Gaussian distribution centred on the nominal mean value and with a standard deviation equal to the corresponding quoted random uncertainty. This allows us to get for each star in TGAS a probability distribution of the output D of the neural network, which can then be characterized by its mean $\bar{D}$ and standard deviation σ_D. As a final cut, we select only stars with $\bar{D} - \sigma_D > 0.9$, for a total of 80 best HVS candidates, $\sim 0.004\%$ of the original catalogue size.

We stress that all our cuts rely on the astrometry of the objects, without any prior assumption on the spectral type, photometry or more in general stellar properties of the selected best sample, and without any information on radial velocities.

3.4 Acquiring spectral information

To confirm or reject a candidate in our quest for HVSs, a measure of the star *total* velocity is necessary. In the following, we will describe how we obtained reliable heliocentric radial velocities (HRVs) for 47 stars out of the 80 candidates.

3.4.1 Catalogue cross-matching

Our final sample has been cross-matched with several spectroscopic surveys of the Milky Way, covering both the Northern and Southern hemisphere[7]. We find a total of 30 stars in common: a subsample of these (5 stars) have both radial velocity and spectroscopic distance from the RAdial Velocity Experiment (RAVE) DR4 and/or DR5 (Kordopatis et al. 2013a; Kunder et al. 2017).

3.4.2 Follow-up observations with the INT

We successfully applied for director's discretionary time at the Isaac Newton Telescope (INT) in La Palma, Canary Islands, where we followed up spectroscopically 22 HVS candidates on the night of the 5th of October, 2016. We used the Intermediate Dispersion Spectrograph (IDS) with the RED+2 CCD, in combination with the R1200R grating, a 1.35" slit width, and the GG495 sorting order filter. This set-up provided an effective spectral range of $\sim 8000-9150$ Å and a resolution at 7000 Å of 6731 over 2 pixels at the detector. We ensured that all observed spectra had a S/N of at least 50.

Spectra reduction

The spectra were reduced using the Image Reduction and Analysis Facility (IRAF, Tody 1986) software package. The reduction procedure included pre-processing (bias and flat field corrections), spectrum extraction, wavelength calibration, heliocentric radial velocity correction, and continuum normalisation.

[7]RAVE DR4 and DR5 (Kordopatis et al. 2013a; Kunder et al. 2017), *Gaia*-ESO DR2 (Gilmore et al. 2012; Randich et al. 2013), LAMOST DR1 and DR2 (Cui et al. 2012), GALAH (Martell et al. 2017), APOGEE DR13 (Zasowski et al. 2013).

Radial velocities, atmospheric parameters and spectroscopic distance determination

A first pass for radial velocity determination is performed by using the python routine pyasl.crosscorrRV, adopting a Solar template as reference, and errors in radial velocities are obtained following Zucker (2003). In order to obtain the effective temperature, surface gravity and metallicity of the stars, the same pipeline as the one used in RAVE (Kordopatis et al. 2011a, 2013a) has been applied to the spectra. This implies keeping only the wavelength range $\lambda\lambda = [8450.80 - 8746.55]$, removing the cores of the Calcium triplet lines (to avoid a mismatch between the synthetic templates used by the pipeline, computed assuming Local Thermodynamical Equilibrium, and the cores of the lines formed in Non LTE), and convolving the observations to a resolution of $R = 7500$. The output of the pipeline is then calibrated using the formulas presented in Kunder et al. (2017).

Our final radial velocities are obtained through the cross-correlation of a synthetic spectrum of the best-fit parameters to the observed spectrum. This cross-correlation is done with the package fxcor in IRAF (Tody 1986). Both the observed and synthesized spectrum are continuum normalized before cross-correlation and we use a Gaussian fit to all points with a correlation of 0.5 or higher to determine the radial velocity and its corresponding measurement uncertainty. During the observations a sample of 14 radial velocity standard stars from Soubiran et al. (2013) were observed with the same setup and matched closely in sky position to our program targets to check the accuracy of our determined radial velocities. We find that there is a good agreement between the literature values and our radial velocities. A mean offset of ~ 0.1 km s^{-1}assures us that there are no significant systematic effects. However, the rms variance between the literature values and our radial velocity determinations of 2.7 km s^{-1}is significantly larger than the median measurement uncertainty in the cross-correlation alone, which is only 1.1 km s^{-1}. We thus adopt an uncertainty floor of 2.5 km s^{-1}and add this in quadrature to our measurement uncertainties. Although we believe the radial velocities derived in this second iteration to be more precise than the first pass radial velocities due to the use of a synthetic spectrum that fits the stellar parameters, we note that the results presented in this paper are robust to the use of either set of radial velocities.

To obtain the spectroscopic distances of the stars, the calibrated stellar parameters are projected on Padova isochrones spanning ages from 100 Myr to 13.5 Gyr, with a step of 0.1 Gyr and a metallicity range between -2.2 dex and $+0.2$ dex. This allows us to obtain the absolute magnitudes in

several photometric bands as in Kordopatis et al. (2011b, 2013c, 2015), and an estimation of the age of the stars as in Kordopatis et al. (2016); Magrini et al. (2017). The distances are then obtained using the distance modulus in the J band, and assuming $A_J = 0.709\,E(B-V)$ (Schlafly & Finkbeiner 2011), where $E(B-V)$ are the Schlegel extinctions towards each line-of-sight.

Kinematic properties from *Gaia* TGAS, radial velocities and stellar parameters derived from spectra of observed HVS candidates are presented in Table 3.1 and in Table 3.2. For a precise cross-match with future *Gaia* releases and other Milky Way surveys, in Appendix .1 we report the *Gaia* and Hipparcos identifier of all the observed sources. We note that for 4 stars out of 22, the pipeline has not converged (quality flag $F = 1$, see Table 3.2) and therefore are excluded from the following analysis. Furthermore, visual inspection of TYC 2292-1267-1 (quality flag $F = 3$), shows a clear mismatch between the observed spectrum and the fitted template, and therefore was discarded as well.

The metallicity and mass distribution are shown, respectively, in Figure 3.2 and 3.3. The mean metallicity of our sample is -1.2 dex, consistent with the inner Galactic halo distribution, dashed (Chiba & Beers 2000) and dot-dashed (Kordopatis et al. 2013b) lines, but a total of 6 stars have $[M/H] > -0.5$ dex, and one candidate, TYC 3945-1023-1, has $[M/H] = -0.02 \pm 0.12$ dex. Most of the stars have masses slightly below the Solar value, with a peak of the distribution at $M \sim 0.85\ \mathrm{M}_\odot$, and a single star with $M \sim 2\ \mathrm{M}_\odot$: TYC 4032-1542-1. We can see that our sample is very different from the late B-type HVS candidates discovered in Brown et al. (2014). Considering the age estimates in Table 3.2, we note that the peak of the mass distribution is at the main-sequence turn-off of the stellar halo. Stars of this type have been used to trace the stellar halo because of their luminosity (e.g. Cignoni et al. 2007).

3.5 Distance estimation

Most of the stars in *Gaia* DR1 have non-negligible parallax errors. Therefore simply estimating distances as the inverse of parallax leads to biased results due to this highly non-linear transformation (Bailer-Jones 2015; Astraatmadja & Bailer-Jones 2016a). Additionally it can not be applied to negative parallaxes, which are present in our sample. In order to correctly take into account correlations between astrometric parameters supplied by the *Gaia* catalogue (parameter correlations may have an important impact on our results since we are implementing Monte Carlo simulations),

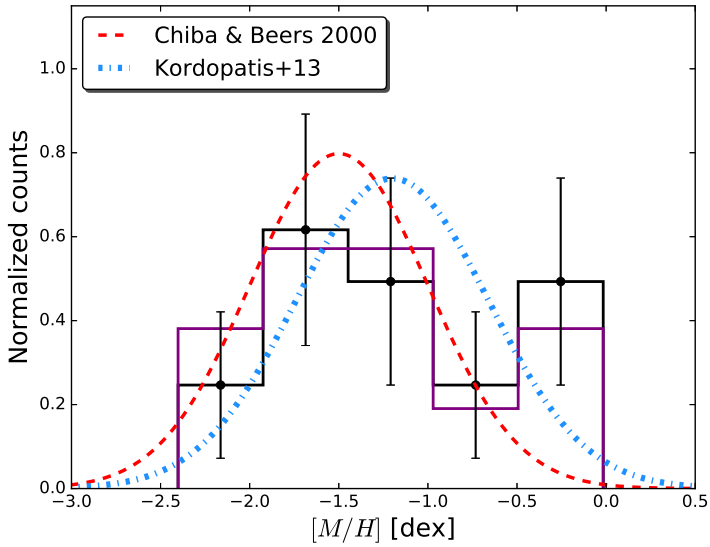


Figure 3.2: Normalized $[M/H]$ distribution for the observed HVS candidates, with error bars computed assuming Poisson noise. For a visual comparison, we overplot with a red dashed (blue dot-dashed) line the inner stellar halo metallicity, modelled as Gaussian with mean and standard deviation from Chiba & Beers (2000) (Kordopatis et al. (2013b)). Purple line shows the normalized $[M/H]$ distribution for high-velocity candidates (see Table 3.3).

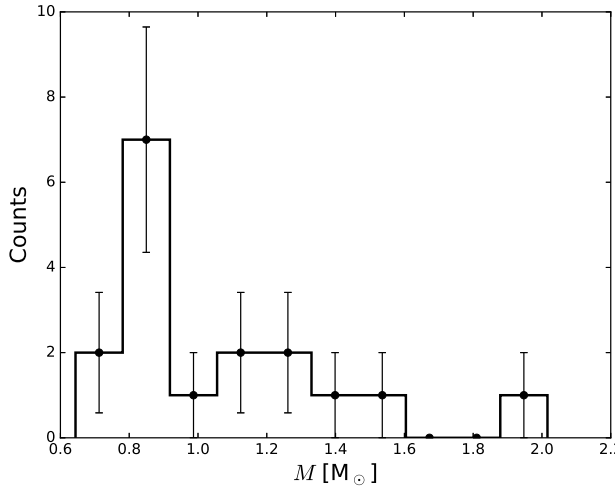


Figure 3.3: Mass distribution for the observed HVS candidates, with error bars computed assuming Poisson noise. The peak of the distribution is ~ 0.85 M$_\odot$.

we choose not to use the distance catalogue presented in Astraatmadja & Bailer-Jones (2016b), but to implement our own Bayesian approach, generalizing their method and considering covariances.

Assuming Gaussian noise for astrometric parameters, we model the likelihood for the triplet $\{\mu_{\alpha*}, \mu_\delta, \varpi\}$ as a multivariate normal distribution with mean vector:

$$\bar{x} = (\mu_{\alpha*},\ \mu_\delta,\ 1/d), \tag{3.8}$$

and with covariance matrix:

$$\Sigma = \begin{pmatrix} \sigma^2_{\mu_{\alpha*}} & \sigma_{\mu_{\alpha*}}\sigma_{\mu_\delta}\rho_{\mu_{\alpha*},\mu_\delta} & \sigma_{\mu_{\alpha*}}\sigma_\varpi\rho_{\mu_{\alpha*},\varpi} \\ \sigma_{\mu_{\alpha*}}\sigma_{\mu_\delta}\rho_{\mu_{\alpha*},\mu_\delta} & \sigma^2_{\mu_\delta} & \sigma_{\mu_\delta}\sigma_\varpi\rho_{\mu_\delta,\varpi} \\ \sigma_{\mu_{\alpha*}}\sigma_\varpi\rho_{\mu_{\alpha*},\varpi} & \sigma_{\mu_\delta}\sigma_\varpi\rho_{\mu_\delta,\mu_\varpi} & \sigma^2_\varpi \end{pmatrix}, \tag{3.9}$$

where $\rho_{i,j}$ is the correlation between the parameters i and j, as given in TGAS. We model the prior probability on distances following the "Milky Way prior" approach presented in Astraatmadja & Bailer-Jones (2016a). We consider a three-dimensional density model for our Galaxy, that takes into account selection effects of the *Gaia* survey:

$$P_{\rm MW}(d, l, b) = d^2 \rho_{\rm MW}(d, l, b)\ p_{\rm obs}(d, l, b). \tag{3.10}$$

Table 3.1: Kinematic properties of 22 HVS candidates spectroscopically followed-up with the INT telescope.

Tycho 2 ID	(RA, dec) (deg)	ϖ (mas)	$\mu_{\alpha*}$ (mas yr^{-1})	μ_δ (mas yr^{-1})	HRV (km s^{-1})
2282-208-1	(16.81855, 33.66159)	2.17 ± 0.31	202.643 ± 1.213	-62.458 ± 0.398	-0.61 ± 1.29
2292-1267-1	(20.86832, 31.78668)	1.78 ± 0.35	90.782 ± 0.969	-15.275 ± 0.644	158.93 ± 5.99
2298-66-1	(25.30039, 33.51859)	2.45 ± 0.34	178.060 ± 1.213	-19.060 ± 0.319	-31.66 ± 2.78
2320-470-1	(31.29, 35.6289)	2.06 ± 0.27	106.443 ± 0.967	6.138 ± 0.290	-43.08 ± 1.32
2376-691-1	(66.43652, 33.59088)	1.17 ± 0.29	62.060 ± 2.077	-9.137 ± 1.547	22.02 ± 1.63
2393-1001-1	(78.45391, 32.03592)	2.21 ± 0.28	121.797 ± 1.710	-46.605 ± 1.158	-106.50 ± 0.94
2818-556-1	(23.79684, 40.43319)	2.56 ± 0.37	147.979 ± 1.369	-41.076 ± 0.468	-92.17 ± 1.42
2822-1194-1	(23.14799, 42.03068)	1.85 ± 0.64	88.644 ± 1.849	2.063 ± 0.496	-23.19 ± 1.87
3163-1181-1	(303.97045, 44.18376)	2.30 ± 0.25	156.232 ± 1.116	67.079 ± 1.026	-194.08 ± 1.61
3263-733-1	(15.00873, 45.13101)	1.83 ± 0.34	95.576 ± 1.290	-3.277 ± 0.425	14.91 ± 1.46
3285-1422-1	(32.53176, 47.41257)	1.10 ± 0.29	75.04 ± 1.682	-31.531 ± 0.505	25.43 ± 1.54
3330-120-1	(56.71171, 48.53692)	2.61 ± 0.30	194.055 ± 0.323	-123.109 ± 0.255	-24.12 ± 1.26
3661-974-1	(4.55758, 57.6662)	3.49 ± 0.651	180.078 ± 1.110	104.039 ± 0.651	-154.53 ± 2.02
3744-1546-1	(67.80849, 58.96855)	1.81 ± 0.42	143.706 ± 1.923	-38.217 ± 1.272	8.72 ± 1.49
3945-1023-1	(304.24414, 56.57186)	-6.07 ± 0.89	-6.097 ± 1.826	-1.265 ± 1.879	-18.79 ± 1.80
3983-1873-1	(338.34366, 52.68866)	1.84 ± 0.23	133.342 ± 0.094	72.34 ± 0.082	-165.28 ± 0.86
4032-1542-1	(26.42901, 60.39286)	0.74 ± 0.40	68.109 ± 0.761	-13.725 ± 0.73	-115.48 ± 7.15
4307-1106-1	(8.16184, 74.08742)	2.31 ± 0.52	72.556 ± 1.141	15.474 ± 1.291	45.88 ± 1.79
4507-1461-1	(33.29978, 82.01739)	2.52 ± 0.31	85.192 ± 0.661	0.366 ± 0.836	-384.65 ± 2.22
4509-1013-1	(58.91556, 75.28116)	2.15 ± 0.24	97.297 ± 0.886	-29.216 ± 0.758	-155.52 ± 1.55
4515-1197-1	(79.71826, 77.83392)	1.28 ± 0.28	96.148 ± 0.892	45.051 ± 1.045	-198.41 ± 1.09
4521-322-1	(55.43942, 81.069)	3.22 ± 0.35	160.469 ± 0.536	1.117 ± 0.768	-129.92 ± 1.19

Notes: Hipparcos and *Gaia* identifiers for these stars are given in Table 4 in Appendix .1. Proper motions and parallaxes are from *Gaia* TGAS, while radial velocities have been derived using the RAVE pipeline. The 2.5 km s^{-1}uncertainty floor is *not* included in the quoted HRV errors, see discussion in §3.4.2.

Table 3.2: Observational properties of 22 HVS candidates spectroscopically followed-up with the INT telescope.

Tycho 2 ID	$T_{\rm eff}$ (K)	$\log g$ (cm s^{-2})	$[M/H]$ (dex)	$d_{\rm spec}$ (pc)	M ($M_\odot$)	$t_{\rm age}$ (Gyr)	F
2282-208-1	5936 ± 136	3.8 ± 0.2	−1.35 ± 0.19	606 ± 152	0.92 ± 0.17	10.4 ± 3.8	1
2292-1267-1	7861 ± 83	4.0 ± 0.2	−0.20 ± 0.12	340 ± 69	1.70 ± 0.14	0.9 ± 0.2	3
2298-66-1	5925 ± 328	3.8 ± 0.5	−2.08 ± 0.26	754 ± 569	0.95 ± 0.23	8.2 ± 4.5	0
2320-470-1	5730 ± 214	3.4 ± 0.5	−3.29 ± 0.27	1240 ± 650	1.00 ± 0.21	6.9 ± 4.0	1
2376-691-1	5260 ± 74	3.5 ± 0.2	−0.67 ± 0.11	249 ± 64	1.22 ± 0.19	4.8 ± 3.8	2
2393-1001-1[8]	4651 ± 1.158	0.6 ± 0.2	−2.40 ± 0.14	3036 ± 462	0.85 ± 0.26	7.6 ± 2.2	0
2818-556-1	5734 ± 63	3.4 ± 0.2	−0.98 ± 0.17	686 ± 153	1.30 ± 0.18	3.4 ± 2.9	2
2822-1194-1	6403 ± 116	4.2 ± 0.2	−0.48 ± 0.12	532 ± 160	1.10 ± 0.09	1.8 ± 2.5	0
3163-1181-1	5570 ± 74	3.4 ± 0.2	−0.30 ± 0.11	463 ± 85	1.59 ± 0.17	2.0 ± 1.1	1
3263-733-1	5425 ± 89	3.8 ± 0.1	−0.81 ± 0.16	517 ± 55	0.88 ± 0.09	12.3 ± 2.2	0
3285-1422-1	5214 ± 89	4.1 ± 0.1	−1.58 ± 0.16	143 ± 87	0.64 ± 0.08	10.9 ± 1.3	2
3330-120-1	5735 ± 89	3.8 ± 0.1	−1.55 ± 0.16	571 ± 30	0.83 ± 0.03	12.5 ± 0.9	0
3661-974-1	6507 ± 100	4.1 ± 0.2	−0.99 ± 0.16	397 ± 83	0.87 ± 0.09	10.2 ± 2.7	1
3744-1546-1	6232 ± 174	4.3 ± 0.3	−1.68 ± 0.20	294 ± 78	0.78 ± 0.05	9.9 ± 4.0	2
3945-1023-1	6239 ± 83	3.8 ± 0.2	−0.02 ± 0.12	1185 ± 150	1.54 ± 0.11	2.3 ± 0.5	0
3983-1873-1	4832 ± 68	2.0 ± 0.2	−1.27 ± 0.14	1096 ± 151	1.06 ± 0.19	5.4 ± 2.5	0
4032-1542-1	7600 ± 83	3.7 ± 0.2	−0.23 ± 0.12	1009 ± 187	2.02 ± 0.16	0.9 ± 0.2	0
4307-1106-1	5517 ± 74	3.5 ± 0.2	−0.45 ± 0.11	844 ± 193	1.41 ± 0.20	3.1 ± 2.4	0
4507-1461-1	6516 ± 100	4.2 ± 0.2	−1.24 ± 0.16	331 ± 30	0.82 ± 0.02	11.8 ± 1.6	0
4509-1013-1	5890 ± 89	3.8 ± 0.1	−1.71 ± 0.16	549 ± 69	0.83 ± 0.08	12.0 ± 1.9	0
4515-1197-1	5398 ± 63	3.4 ± 0.2	−1.63 ± 0.17	902 ± 170	0.88 ± 0.15	11.4 ± 3.5	0
4521-322-1	5872 ± 89	4.0 ± 0.1	−1.38 ± 0.16	428 ± 29	0.83 ± 0.02	12.4 ± 0.6	0

[8] This star has a very low $\log g$, making the position of the isochrones uncertain. Furthermore, its metallicity is outside of the range of our isochrones, therefore distance, mass, and age could be biased or offset.

Notes: Hipparcos and *Gaia* identifiers for these stars are given in Table 4 in Appendix .1. Stellar parameters have been derived using the RAVE pipeline. F = flag for the stellar parameter pipeline: 0 = converged; 1 = not converged; 2 = the pipeline oscillated between two solutions and the mean has been performed; 3 = bookkeping flag, the pipeline has converged.

The stellar number density of the Milky Way $\rho_{\rm MW}(d, l, b)$ is modelled as the sum of three components (see Appendix A in Astraatmadja & Bailer-Jones (2016a) for details), while $p_{\rm obs}(d, l, b)$ describes the fraction of observable stars in a given sky position (Equation (4) in Astraatmadja & Bailer-Jones (2016a)). We choose this prior in our analysis because it gives the best results when comparing distances with a sample of known Cepheids (Astraatmadja & Bailer-Jones 2016b). The impact of assuming different priors on distance is discussed in Appendix .2: except at distances > 800 pc, where errors are large, different priors give similar results. We assume uniform priors on proper motions. By means of Bayes' theorem we draw random samples of proper motions and distances from the resulting posterior distribution with an affine invariant ensemble Markov Chain Monte Carlo (MCMC) sampler (Goodman & Weare 2010), using the *emcee* implementation (Foreman-Mackey et al. 2013). We run the chain with 32 walkers and 4000 steps per walker, for a total of 128000 points drawn from the resulting posterior probability distribution. We check the convergence of the chain in terms of both the mean acceptance fraction and the auto-correlation time.

An example of a cornerplot showing Bayesian posterior distributions and correlations between the astrometric parameters for the candidate TYC 49-1326-1 is shown in Figure 3.4.

For the subset of 22 stars with a spectroscopic distance estimate we simply draw proper motions from a bivariate Gaussian distribution using the 2×2 covariance matrix provided by TGAS, and distances from a Gaussian with standard deviation equal to the estimated random uncertainty on distance.

If parallax-inferred and spectroscopic distance estimates are consistent within the errors, we expect the difference between the two divided by combined uncertainties to be distributed as a Gaussian with mean of zero and standard deviation of one. If we compute a Kolmogorov-Smirnov test to check whether these two distributions are consistent, we find that the null hypothesis cannot be rejected at a 5% level of significance. This is due to large uncertainties in distances, especially when adopting TGAS parallaxes. Since the two estimates can be remarkably different for individual stars, in the following we will present and discuss results assuming both distances.

3.6 Results

Exploiting archival and new data we have assembled a total of 47 candidates with 3D position and velocity. A positive identification of a HVS re-

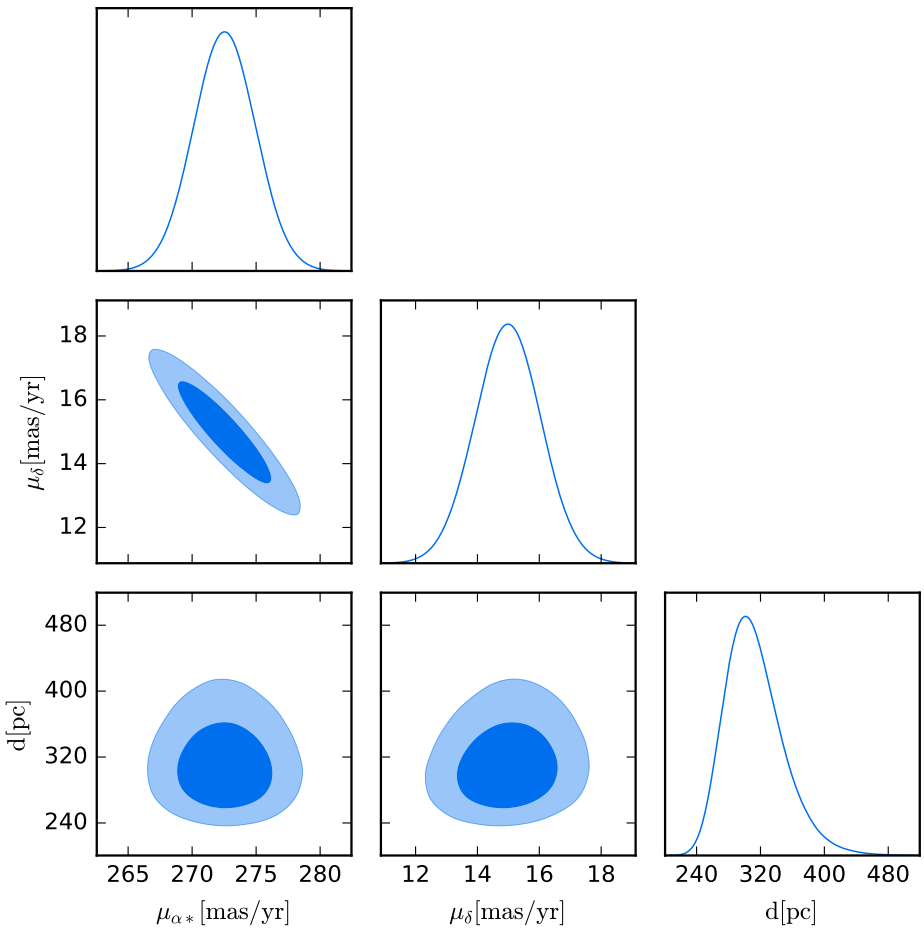


Figure 3.4: Proper motions and distance posterior distributions for the candidate TYC 49-1326-1 as obtained from the MCMC. Correlations from TGAS are $\rho_{\mu_{\alpha*},\mu_\delta} = -0.909$, $\rho_{\mu_{\alpha*},\varpi} = 0.023$, $\rho_{\mu_\delta,\mu_\varpi} = -0.103$. Dark (light) blue regions indicate the extent of the 1σ (2σ) credible intervals.

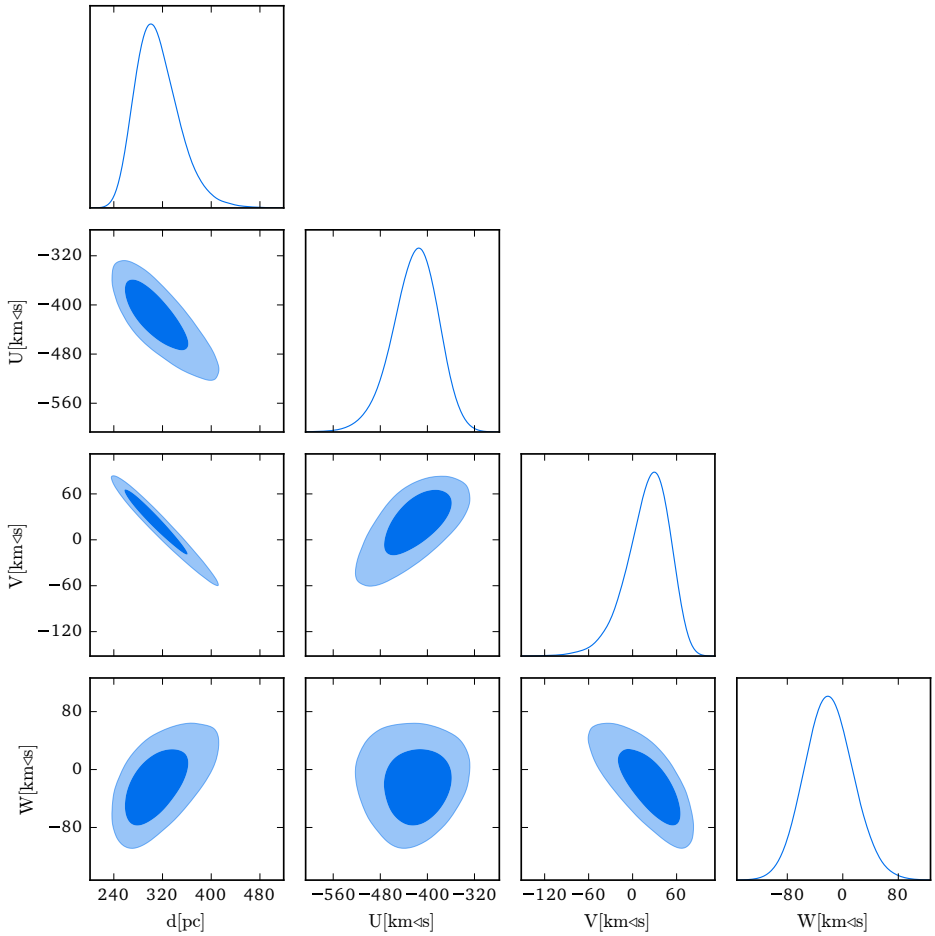


Figure 3.5: Distance and Galactic rectangular velocities U, V, W posterior distributions for TYC 49-1326-1 as obtained from the sampling of the astrometry shown in Figure 3.4. Dark (light) blue regions indicate the extent of the 1σ (2σ) credible intervals. The total galactocentric velocity is $v_{\rm GC} = 419^{+38}_{-35}$ km s^{-1}.

quires both a radial trajectory from the Galactic Centre and a total velocity above the local escape speed. A star with the latter property but a trajectory that originates from the stellar disc will be called an *hyper runaway star*. Finally, *bound HVSs* (BHVSs) have Galactic Centre origin but velocity below the escape speed.

3.6.1 Total Galactocentric velocity

In order to identify HVSs, we compute the total velocity in the Galactic rest frame v_{GC} for the 47 candidates with a reliable radial velocity measurement. We start correcting radial velocities and proper motions for solar and LSR motion, assuming a three-dimensional Sun's velocity vector and LSR velocity (Schönrich 2012). We then calculate Galactic rectangular velocities U, V, and W with the following convention: U is positive if pointing towards the GC, V is positive along the direction of Galactic rotation, and W is positive towards the North Galactic Pole (Johnson & Soderblom 1987). The total velocity in the Galactic rest-frame is then simply computed summing in quadrature these three velocity components. We estimate uncertainties in the velocity vector via MC simulations, using the sampling in proper motions and distance described in Section 3.5. An example of posterior distributions for rectangular velocities is shown in Figure 3.5 for the candidate TYC 49-1326-1, obtained using posterior distributions shown in Figure 3.4.

For each star we draw 10^5 random realizations of its astrometric parameters, and the resulting total velocities are plotted in the first column of Figure 3.6 as a function of Galactocentric distance. We quote our results in terms of the median of the distribution, and errors are derived from the 16th and 84th percentiles. We overplot the median escape speed from the Milky Way derived in Williams et al. (2017) using a dashed line, with corresponding 68% (95%) credible intervals shown as a dark (light) blue region. This shows how the algorithm succeeded in finding high-velocity stars: 45 out of 47 candidates have a median Galactic rest frame velocity > 150 km s^{-1}, which is the typical velocity dispersion of stars in the halo (Smith et al. 2009; Evans et al. 2016). Considering parallax-inferred distances, first row, 11 objects are compatible within their uncertainties to be unbound from the Milky Way. If we use spectroscopic estimates, we find 3 stars with a total velocity consistent with being greater than the median escape speed at their position. Discussion of individual objects is postponed to Section 3.7.

Total velocities and distances are presented in Table 3.3 for the 15 stars with a median Galactic rest-frame velocity > 350 km s^{-1}obtained with at

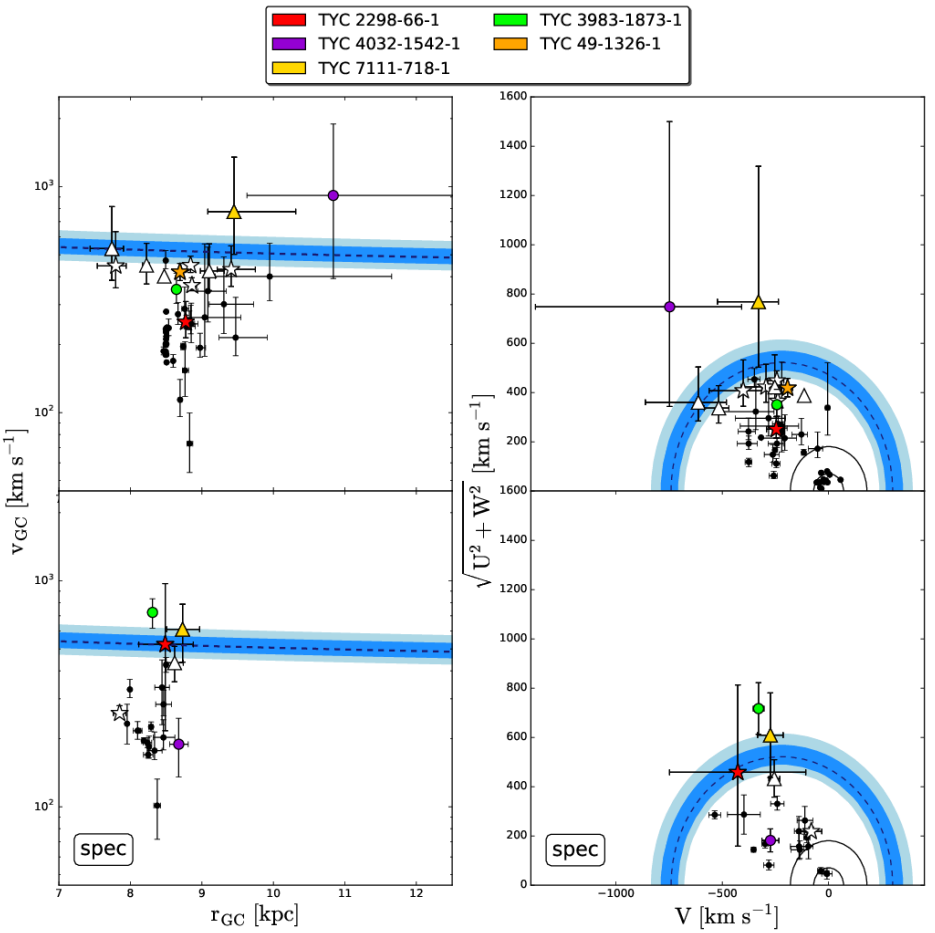


Figure 3.6: *First column*: Total Galactic rest frame velocity versus Galactocentric distance for those HVS candidates with a reliable radial velocity measurement. *Second column*: Toomre diagrams (in the LSR frame) for the same candidates. The two black rings in the bottom-right corner refer to the boundaries of the thin and thick disk, respectively at a constant velocity of 70 and 180 km s^{-1}(Venn et al. 2004). Most of our candidates lie in the kinematic region corresponding to halo stars. *First row*: velocities computed using distances inferred from parallax, using the MW prior. *Second row*: velocities computed using a spectroscopic distance estimate, when available. *All plots*: The dashed line is the median posterior escape speed (as a function of radius in the first column, and the local 521^{+46}_{-30} km s^{-1}in the second one) from Williams et al. (2017) with the 68% (94%) credible interval shown as a dark (light) blue band. Stars mark HVS/BHVS candidates in Table 3.3. Triangles mark runaway star candidates in Table 3.3. 11 objects are consistent with being unbound from the Milky Way in the first row, and 3 if we adopt spectroscopic distances.

least one of the distance estimation methods. The *Gaia* and Hipparcos identifier of these high velocity candidates is presented in Appendix .1. We assign to each star its probability of being unbound from the Galaxy, $P^{\rm u}$. From the posterior probability on distance d, we can compute the escape velocity from the Galaxy in each realization of the star's position using the analytic fit in Williams et al. (2017). We define $P^{\rm u}$ as the fraction of Monte Carlo realizations with $v_{\rm GC}(d) > v_{\rm esc}(d)$.

In the right panels of Figure 3.6 we present Toomre diagrams in the LSR frame for our candidates. In a Toomre's diagram one can identify three regions (separated by two solid black lines), corresponding to stars in the thin, thick disc, and halo (Venn et al. 2004; Hawkins et al. 2015). In the stellar halo kinematic region we report the local escape speed with associated errors (blue stripe, Williams et al. 2017)[9]. The two panels correspond to different distance determinations. Most of our candidates are consistent, from a kinematic point of view, with being halo stars. A total of 12 objects are consistent with being thin/thick disc stars considering parallax-inferred distances, and therefore will not be furthermore discussed.

3.6.2 Orbital traceback

We now proceed to establish the star candidate's origin by tracing back its trajectory in different models for the Galactic potential. We decide to perform the full orbit integration only for the most promising high-velocity stars in our sample, imposing the cut $\max(v_{\rm GC},\, v_{\rm GCspec}) > 350\,{\rm km\,s^{-1}}$, where quoted values denote the median of the distribution. A total of 15 objects passes this cut (see Table 3.3).

We use the publicly available python package *galpy*[10] (Bovy 2015b) to integrate the orbit of each object in the Milky Way. We run 10^5 MC realizations of the star's orbit, using as initial conditions the position, distance, and U, V, W velocities previously randomly sampled from the posterior distributions. We use a four components Galactic potential, and we study the impact of our results depending on the choice of its parameters.

Our fiducial model consists of a point mass black hole potential:

$$\phi_{BH}(r) = -\frac{GM_\bullet}{r}, \tag{3.11}$$

a spherically symmetric bulge modelled as a Hernquist spheroid (Hern-

[9] We choose for simplicity to plot the local value.

[10] `http://github.com/jobovy/galpy`

quist 1990):

$$\phi_b(r) = -\frac{GM_b}{r + r_b}, \tag{3.12}$$

a Miyamoto-Nagai disc in cylindrical coordinates (R, z) (Miyamoto & Nagai 1975):

$$\phi_d(R, z) = -\frac{GM_d}{\sqrt{R^2 + \left(a_d + \sqrt{z^2 + b_d^2}\right)^2}}, \tag{3.13}$$

and a Navarro-Frenk-White (NFW) profile for the dark matter halo (Navarro et al. 1996):

$$\phi_h(r) = -\frac{GM_h}{r} \ln\left(1 + \frac{r}{r_s}\right). \tag{3.14}$$

We adopt the following values for the potential parameters: $M_b = 3.4 \times 10^{10}$ M$_\odot$, $r_b = 0.7$ kpc, $M_d = 1.0 \times 10^{11}$ M$_\odot$, $a_d = 6.5$ kpc, $b_d = 0.26$ kpc (Johnston et al. 1995; Price-Whelan et al. 2014; Hawkins et al. 2015), $M_h = 0.76 \times 10^{12}$ M$_\odot$, $r_s = 24.8$ kpc (Rossi et al. 2017). This potential gives a local escape speed ~ 580 km s^{-1}, in agreement with results in Piffl et al. (2014), and, using data presented in Huang et al. (2016), provides a good fit to the rotation curve of the Milky Way out to ~ 100 kpc (see Appendix A, Figure A1, in Rossi et al. 2017).

For those stars for which we do not have a spectroscopic estimate of the age, we trace the orbit back in time for a fiducial time of 10 Gyr, motivated by the typical age and mass of the observed sample (see Table 3.2 and Figure 3.3). We integrate each orbit with a time resolution of 0.5 Myr, keeping track of each disc crossing (Galactic latitude $b = 0$).

If a star is ejected via the Hills mechanism but it is still gravitationally bound to the Milky Way, after the turn-around (maximum distance from the GC) it might cross multiple time the disc before being observed. This is supported by the fact that INT observations suggest that the majority of our stars have ages much larger than typical flight times from the stellar disc to the observed position, the latter being of the order of hundreds of Myr. An example of such a bound orbit is shown in Figure 3.7. Thus it is not trivial to determine which disc crossing should be assigned in order to understand whether or not our candidates effectively originate from the GC. Zhang et al. (2016), searching for nearby low mass high velocity stars, assume the most-recent disc crossing to be the ejection location of the star in the Galaxy. Given the complexity of bound orbits, we simply check the consistency of the GC origin hypothesis for our candidates by recording the

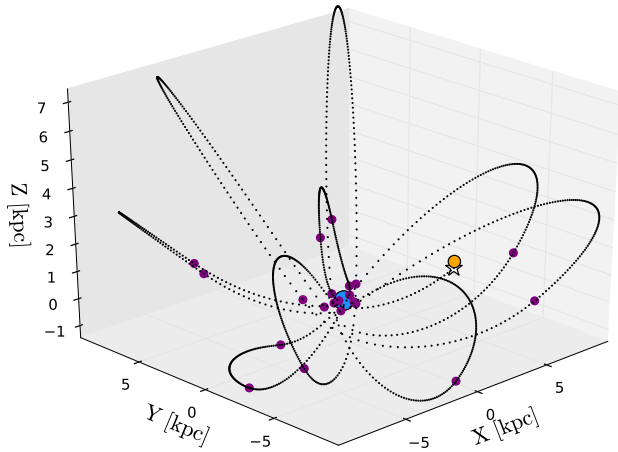


Figure 3.7: Example MC realization of a single bound orbit of TYC 2298-66-1 using the spectroscopic distance estimate. The blue (orange) circle marks the position of the GC (Sun), and the white star corresponds to the observed position of the star. Purple dots mark the disc crossings of the star prior to, and including the one happening closest to the GC. The initial conditions are $d_0 = 1018$ pc, $v_{\rm GC} = 225$ km s^{-1}, the eccentricity is $e \sim 0.96$, and the estimated flight time from the assigned ejection location to the observed position is $t_{\rm f} = 1.3$ Gyr $\ll t_{\rm age} = 8.2$ Gyr. For this particular orbit, the closest disc crossing is at ~ 260 pc from the Galactic Centre.

closest disc crossing to the GC. This approach allows us to directly exclude stars that are not HVSs, since it is a necessary condition for a HVS that this method results in a density contour level containing the GC.

We find 8 stars to have orbits consistent with coming from the Galactic Centre using parallax-inferred distances. Within the sample of stars with spectroscopic distances we find 3 candidates, and all of them originate from the GC also when parallax-inferred distances are used.

We check the robustness of this conclusion integrating trajectories in different Milky Way potentials. Our choice for the mass of the bulge is significantly higher compared to the latest observational constraints (Bland-Hawthorn & Gerhard 2016; McMillan 2017), therefore we integrate each candidate assuming a bulge mass equal to half the previous adopted value: $M_b = 1.7{\times}10^{10}\,{\rm M}_\odot$, keeping fixed all the other parameters. As a second test, we adopt the potential in Kenyon et al. (2014), commonly adopted in HVS papers, which has a less massive bulge and stellar disc (but different scale

parameters). In both cases we find the same candidates to be consistent with coming from the GC. As a final test, we study the impact of assuming a triaxial profile for the bulge, which might influence the orbital traceback in the inner regions of the Galaxy. Results from star counts recently revealed that the Milky Way bulge has a boxy/peanut shape (McWilliam & Zoccali 2010; Wegg & Gerhard 2013), which can be characterized by an axis ratio from top $(b/a) \sim 0.5$, and an edge-on axis ratio $(c/a) \sim 0.26$ (Bland-Hawthorn & Gerhard 2016). Adopting the same mass and scale radius as in our fiducial potential and using a triaxial Hernquist profile to model the bulge, we find the shape of the density contour to change considerably, but the assumption of consistency with coming from the GC is solid.

Figure 3.8 shows example probability density functions of the disc crossing locations in the Galactic plane (rotating anticlockwise) for two candidates which will be further discussed in next sections, assuming our fiducial model for the Galactic potential. TYC 49-1326-1, left panel, is consistent with coming from the GC, while for TYC 3983-1873-1, right panel, the GC origin is excluded.

3.7 Discussion of Individual Candidates

We divide candidates in Table 3.3 in three major classes: HVS and BHVS candidates, runaway star candidates, and "uncertain" objects. To help the discussion, the metallicity distribution of these stars is shown with a purple line in Figure 3.2, where it is compared to typical metallicity distributions of stars in the inner Galactic halo. We will now discuss separately candidates from each class in detail, focusing on the most promising objects and on stars already present in literature. One additional candidate not included in Table 3.3, but known from literature, is discussed in Section 3.7.4.

3.7.1 HVS and BHVS Candidates

In addition to HVSs, the Hills mechanism naturally predicts a population of *bound HVSs*: stars having a velocity high enough to escape from the MBH's gravitational field at their ejection, but not sufficient to be unbound from the whole Milky Way. These stars, being decelerated and deflected by the Galactic potential, can cross the disc multiple times during their life, following a wide variety of highly-non-radial orbits, as previously shown in Figure 3.7. The identification of such objects is observationally particularly difficult.

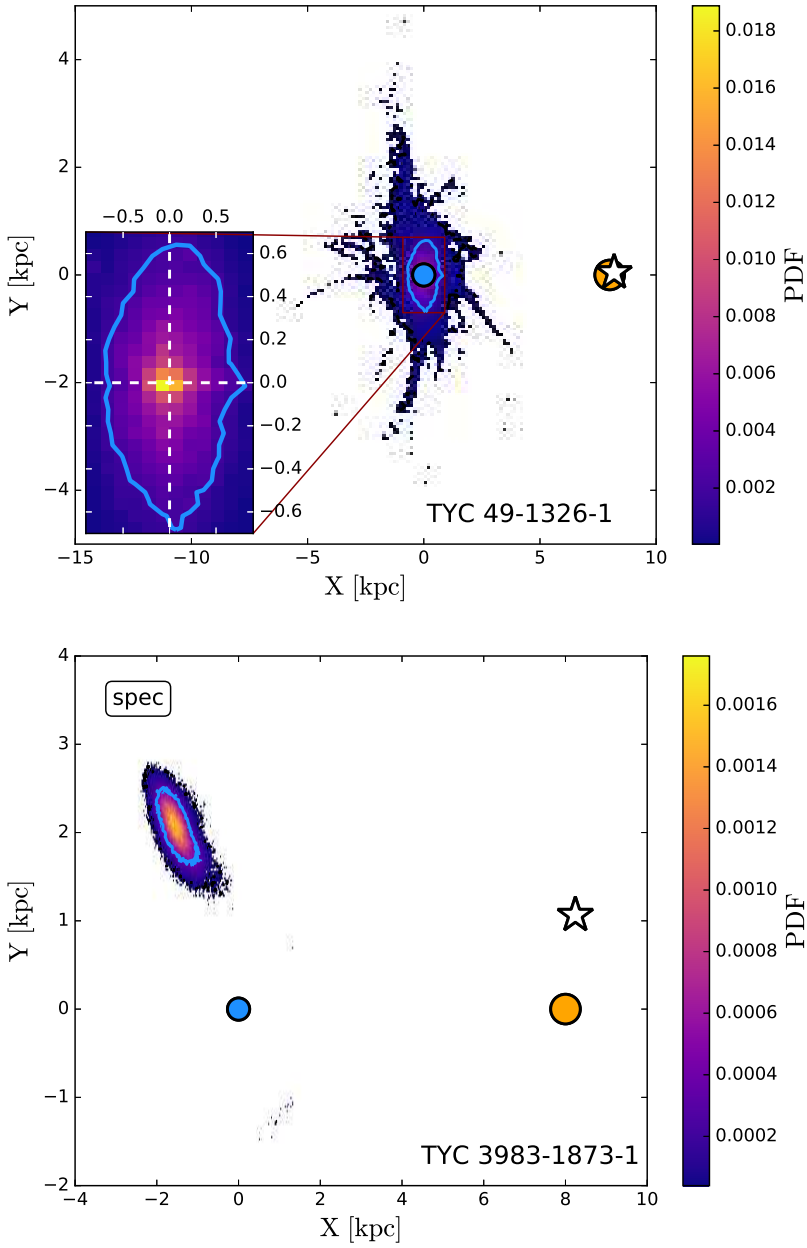


Figure 3.8: Normalised probability distribution function of Galactic disc crossings for the candidates TYC 49-1326-1, assuming the parallax-inferred distance (top panel), and TYC 3983-1873-1, using the spectroscopic distance (bottom panel). The blue line marks the 1σ contour, and the coloured region extends up to the 2σ contour. The MW rotates anticlockwise. The blue (orange) circle marks the position of the GC (Sun), while the white star corresponds to the median observed position of the candidate. The white dashed cross marks the position of the GC in the zoomed inset.

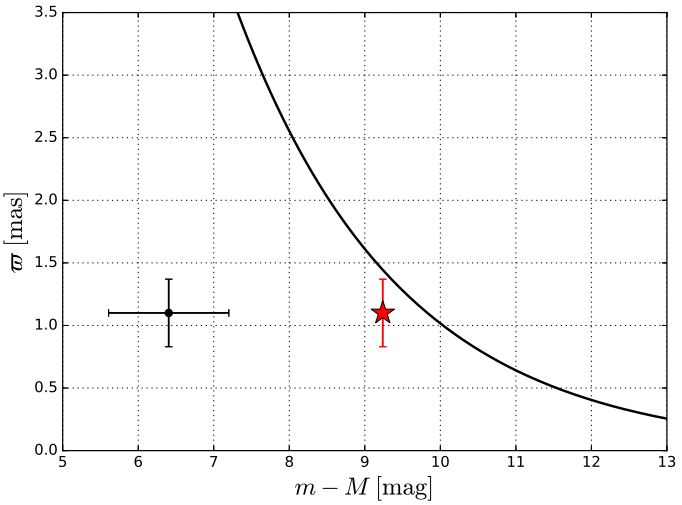


Figure 3.9: Parallax-distance modulus diagram for the RR Lyrae star TYC 8422-875-1 (HD 201484, V Ind), using the parallax from TGAS and the distance modulus from RAVE DR5 (black point). The line shows the analytic prediction assuming the Schlegel extinction towards the line-of-sight. The parallax-inferred distance estimate is clearly favoured. The red star corresponds to adopting the distance modulus obtained using the PLZ relation. Other candidates lie too close to the curve to have a clear preference towards one distance estimate.

Table 3.3: Derived kinematic properties for the 15 HVS candidates with $\max(v_{\rm GC},\ v_{\rm GCspec}) > 350$ km s^{-1}, and interpretation.

Tycho 2 ID	HRV (km s^{-1})	$[M/H]$ (dex)	d (pc)	$d_{\rm spec}$ (pc)	$v_{\rm GC}$ (km s^{-1})	$v_{\rm GCspec}$ (km s^{-1})	$P^{\rm u}$	$P^{\rm u}_{\rm spec}$	Ref
HVS / BHVS candidates									
2298-66-1	-31.66 ± 2.78	-2.08 ± 0.26	431^{+78}_{-55}	754 ± 569	248^{+58}_{-38}	519^{+451}_{-307}	0.1%	50.3%	1
8422-875-1[11]	200.8 ± 0.8	-1.01 ± 0.07	1010^{+400}_{-218}	208 ± 124	446^{+186}_{-89}	259^{+21}_{-7}	29.1%	0.0%	2, 5
2456-2178-1	-243.08 ± 49.53	-2.25 ± 0.24	976^{+358}_{-207}		430^{+117}_{-68}		22.7%		3
2348-333-1	205.26 ± 0.34	-1.26 ± 0.40	407^{+51}_{-40}		448^{+44}_{-32}		7.6%		3, 4
49-1326-1	265.1 ± 37.6		304^{+38}_{-30}		419^{+38}_{-35}		1.2%		2, 5
5890-971-1	348.6 ± 0.8		550^{+93}_{-72}		366^{+29}_{-20}		0.2%		6, 7
Runaway star candidates									
7111-718-1	76.7 ± 1.2	-1.53 ± 0.17	1967^{+1413}_{-683}	1552 ± 430	776^{+576}_{-274}	611^{+176}_{-172}	82.2%	70.7%	2, 5
8374-757-1	71.8 ± 3.7		832^{+338}_{-179}		532^{+284}_{-147}		50.4%		8
1071-404-1	-267.12 ± 0.26	~ -0.5	439^{+91}_{-64}		449^{+113}_{-78}		23.7%		4
4515-1197-1	-198.41 ± 1.09	-1.63 ± 0.17	881^{+292}_{-175}	902 ± 170	423^{+137}_{-76}	433^{+78}_{-76}	23.5%	15.6%	1
9404-1260-1	-94.9 ± 0.6		$67.0^{+1.0}_{-0.9}$		402^{+4}_{-4}		0.0%		9
Uncertain candidates									
3983-1873-1	-165.28 ± 0.86	-1.27 ± 0.14	572^{+88}_{-67}	1096 ± 151	351^{+64}_{-47}	726^{+107}_{-108}	1.5%	97.2%	1
4032-1542-1	-115.48 ± 7.15	-0.23 ± 0.12	3216^{+2918}_{-1574}	1009 ± 187	918^{+979}_{-527}	183^{+59}_{-57}	75.7%	0.0%	1
3945-1023-1	-18.79 ± 1.80	-0.02 ± 0.12	4978^{+2802}_{-1686}	1185 ± 150	399^{+162}_{-87}	215^{+4}_{-4}	24.5%	0.0%	1
3330-120-1	-24.12 ± 1.26	-1.55 ± 0.16	401^{+56}_{-43}	571 ± 30	247^{+58}_{-44}	425^{+32}_{-32}	0.1%	0.3%	1

[11] The parallax-inferred distance d is more likely to be correct for this RR Lyrae star (see Figure 3.9), and is consistent with the value obtained using a PLZ relation (see discussion in §3.7.1).

Notes: Hipparcos and *Gaia* identifiers for these stars are given in Table 4 in Appendix .1. The subscript "spec" refers to quantities computed using the spectroscopic distance (when available). For distances and Galactocentric velocities, results are quoted in terms of the median of the distribution with uncertainties derived from the 16th and 84th percentiles. The 2.5 km s^{-1}uncertainty floor (see discussion in §3.4.2) is *not* included in the quoted HRV errors.

References: (1) This paper, observations at the INT; (2) Kordopatis et al. (2013a); (3) (Cui et al. 2012); (4) Latham et al. (2002); (5) Kunder et al. (2017); (6) Przybylski (1978); (7) Barbier-Brossat et al. (1994); (8) Kharchenko et al. (2007); (9) Holmberg et al. (2007).

The probability of observing a star at a particular moment of its orbit is proportional to the residence time t_r in that orbit element: $p \propto t_r \propto v^{-1}$, therefore we expect most of these stars to be observed when they have low velocities, and they could thus be easily mistaken for halo stars.

Hypervelocity and bound hypervelocity star candidates are marked with a star symbol in Figure 3.6. Stars are classified as HVSs if (i) their velocity is > 350 km s^{-1}with at least one distance estimate, and (ii) if they are consistent with coming from the GC (within 2σ) when traced back in different Galactic potentials. We find a total of 6 stars satisfying both properties within their uncertainties: TYC 2298-66-1, TYC 8422-875-1, TYC 2456-2178-1, TYC 2348-333-1, TYC 49-1326-1, and TYC 5890-971-1. The consistency with the GC origin does not depend on the assumed distance. The further sub-classification as HVSs or BHVSs depends on the value of $P^{\rm u}$. All of these stars are on highly radial orbits, with median eccentricities > 0.9.

- TYC 2298-66-1 (LP 295-632) is a high proper motion metal-poor candidate, identified by a red symbol in Figure 3.6. It is the only star with a probability $> 50\%$ of being unbound from the Galaxy when using the spectroscopic distance estimate ($v \sim 530$ km s^{-1}, even if with large uncertainties), therefore it is a HVS candidate.

- TYC 8422-875-1 (HD 201484, V Ind) is a F0 V variable star of RR Lyrae type (Houk 1978). In the discussion of this candidate, we use Figure 3.9 to help us distinguish which distance estimate is more likely to be correct. This plot compares the position of the star in the parallax-distance modulus diagram to the analytical prediction computed assuming the Schlegel extinction towards the line-of-sight. The distance modulus is taken from RAVE DR5 (Kunder et al. 2017), and the resulting point is shown in black. The total velocity of TYC 8422-875-1 strongly depends on the distance assumption, but from Figure 3.9 we can see that parallax-inferred distance is more likely to be correct. Furthermore, since this star is a RR Lyrae, we can independently determine its distance modulus using a period-luminosity-metallicity (PLZ) relation (Leavitt 1908; Leavitt & Pickering 1912). Period, [Fe/H] metallicity, and mid-infrared $[3.6]$ magnitude are taken from Monson et al. (2017), and we estimate the distance modulus using the PLZ relation in the *WISE W*1 band from Sesar et al. (2017). This results in a distance modulus ~ 9.3, consistent with the parallax measured by *Gaia*, as shown with a red star in Figure 3.9. We then conclude that V Ind is a BHVS candidate, with $v \sim 450$ km s^{-1}and a

probability of $\sim 30\%$ of being unbound.

- TYC 2456-2178-1 is a BHVS candidate, with $v \sim 430$ km s^{-1}and a probability $\gtrsim 20\%$ of being unbound from the Galaxy.

- TYC 2348-333-1 (G 95-11) is a high proper motion and high velocity star which has been previously used to estimate the local Galactic escape speed together with other stars from the $uvby-\beta$ survey of high velocity and metal poor stars (García Cole et al. 1999). With a total velocity around 450 km s^{-1}, this star is most likely a BHVS. We note that our distance estimate is higher than the value ~ 250 pc given in García Cole et al. (1999), resulting in a higher total velocity.

- TYC 49-1326-1 (G 75-29), marked with an orange star in Figure 3.6, is a BHVS candidate with a total velocity particularly well constrained of 419^{+38}_{-35} km s^{-1}.

- TYC 5890-971-1 (HD 27507), even if it has a total velocity lower than the other candidates, is worth mentioning because it is historically the first discovered HVS candidate. Przybylski (1978) discussed the possibility that HD 27507 is a star escaping from our Galaxy given its high velocity, and a following proper motion redetermination confirmed this conclusion (Clements et al. 1980). The authors found a total velocity ~ 360 km s^{-1}, in good agreement with our results, but studies in the past decades substantially increased the value of the local escape speed (see Williams et al. (2017) for the latest constraints), making this star unlikely to be unbound from the Milky Way. Nevertheless, its orbit is consistent with coming from the GC, making TYC 5890-971-1 a bound HVS candidate.

3.7.2 Runaway Star Candidates

Runaway stars (RSs) are high velocity stars ejected in many-body dynamical encounters in dense stellar systems (Poveda et al. 1967; Portegies Zwart 2000) or by the explosion of a supernova in a binary system (Blaauw 1961; Tauris & Takens 1998). Tauris (2015) showed how it is possible to reach Galactic rest frame velocities up to ~ 1280 km s^{-1}for the ejected companion star in a binary disrupted via an asymmetric supernova explosion. These extreme velocities can be achieved by low-mass G/K candidates in very compact presupernova binaries. High velocity runaway stars observed in the halo are most likely produced in the disc (Bromley et al. 2009; Duarte

de Vasconcelos Silva 2012; Kenyon et al. 2014). Since most of our stars have masses slightly below the Solar value, this mechanism can possibly explain the notable velocity of our stars that do not originate from the GC.

With this classification rule we identify as runaway candidates 5 high-velocity stars: TYC 7111-718-1, TYC 8374-757-1, TYC 1071-404-1, TYC 4515-1197-1, and TYC 9404-1260-1. Regardless of the adopted distance, these stars always have median $v_{\rm GC} > 350$ km s^{-1}. In particular, 2 stars have a probability $> 50\%$ of being unbound from the Milky Way, and are therefore classified as *hyper runaway stars* (HRSs). Runaway star candidates are marked with a triangle symbol in Figure 3.6. In the following we discuss them individually.

- TYC 7111-718-1, marked in yellow in Figure 3.6, is a strong hyper-runaway star candidate, with a velocity > 600 km s^{-1}, in excess of the local escape speed regardless of the adopted distance estimate. From a chemical point of view, it is consistent with the inner Galactic halo population.

- TYC 8374-757-1 (HD 176387, MT Tel) is a RR Lyrae variable star. It was previously discovered by Przybylski (1967), which discussed, despite large uncertainties in proper motions, its nature as a high velocity star. Because of large errors in distance we cannot strongly constrain its total velocity, which, with a median value ~ 530 km s^{-1}, is nevertheless consistent with being greater than the escape speed, making MT Tel a hyper-runaway star candidate. We repeat the same approach discussed for TYC 8422-875-1 to determine the distance of MT Tel using the PLZ relation in Sesar et al. (2017) using data from Monson et al. (2017). We find a distance modulus ~ 8.1, consistent with the parallax from *Gaia*, confirming our high-velocity determination.

- TYC 1071-404-1, TYC 4515-1197-1, and TYC 9404-1260-1 are RS candidates most likely bound to the MW, with a remarkably high total velocity $\gtrsim 400$ km s^{-1}.

Another intriguing origin for these stars not originating from the GC is that they come from the Large Magellanic Cloud (LMC), either as runaway stars (Boubert et al. 2017b), or by the extension of the Hills mechanism to a hypothetical MBH at the centre of the LMC (Boubert & Evans 2016). Uncertainties are at the moment too large to pinpoint their ejection location, and we dot not further expand on this possibility in this paper.

3.7.3 Uncertain Candidates

In our final sample (Table 3.3) there are 4 stars with uncertain interpretation: TYC 3983-1873-1, TYC 4032-1542-1, TYC 3945-1023-1, TYC 2393-1001-1, and TYC 3330-120-1. These objects have a debated nature, with velocities and origins highly dependent on the assumed distance indicator. We classify as runaway star (halo star) candidates that are not consistent with coming from the GC, and with a total velocity > 350 km s^{-1}(< 350 km s^{-1}).

- TYC 3983-1873-1 (BD+51 3413) is a high proper motion HVS candidate (green points in Figure 3.6). It is one of the few candidates with a spectroscopic distance higher than the parallax inferred one, which results in a total velocity of ~ 725 km s^{-1}, more than 1σ above the median escape speed. Remarkably, if we assume a spectroscopic distance, this object is not consistent with coming from the GC, and should therefore be classified as a HRS, while it is a BHVS candidate ($v \sim 350$ km s^{-1}) if we adopt the parallax-inferred distance.

- TYC 4032-1542-1, marked in purple in Figure 3.6, suffers from a particularly poor distance determination. The spectroscopic distance gives a relatively low velocity of ~ 190 km s^{-1}, consistent with that of a high velocity halo star. Its velocity increases considerably if we rely on the much more uncertain parallax-inferred distance ($v \sim 900$ km s^{-1}). A point worth mentioning is that the metallicity is considerably higher than the mean value in the inner halo, making this object worth inspecting in order to constrain its nature and origin as kinematic and chemical outlier. Furthermore, TYC 4032-1542-1 is an A type star, more massive compared to the other candidates, therefore it is more difficult to explain its high velocity invoking the disruption of a close binary via supernova explosions (Tauris 2015, and see discussion in Section 3.7.2).

- TYC 3945-1023-1 is a RS ($v \sim 400$ km s^{-1}) or a halo star ($v \sim 200$ km s^{-1}) candidate, if we assume the parallax-inferred or the spectroscopic distance estimate respectively.

- TYC 3330-120-1 is a runaway star candidate ($v \sim 425$ km s^{-1}) if we adopt the spectroscopic distance, but behaves as a typical halo star ($v \sim 250$ km s^{-1}) if we infer distance from parallax.

3.7.4 HD 5223: Most Likely Not a HVS

In this subsection we present one additional star discovered with our data mining algorithm, TYC 1739-1500-1 (HD 5223). Even if it doesn't pass the velocity cut in Table 3.3, this star was previously known and discussed for its high velocity, which we now revisit using *Gaia*'s much more precise data.

HD 5223 is a carbon-enhanced metal-poor star presented in Pereira et al. (2012), which concluded that this object is a hypervelocity star with a total velocity in the Galactic frame of 713 km s^{-1}. Our velocity determination $v = 288^{+72}_{-46}$ km s^{-1}is considerably lower because of a substantial difference in the assumed distance: Pereira et al. (2012) determined $d = 1.2$ kpc, while our computation seems to suggest lower values: $d = 565^{+117}_{-80}$ pc. If our estimate is correct, HD 5223 is bound to the MW, and furthermore we find its orbit not to be consistent with coming from the GC.

3.8 Discussion and Conclusions

We successfully developed a new automatized method to extract high velocity stars, using a data-driven algorithm trained on mock populations of hypervelocity stars. Our data mining routine, an artificial neural network, is optimized for the very unbalanced search of rare objects in a large dataset. This approach avoids a bias towards particular spectral types or stellar properties, making as few assumptions as possible on the stellar nature of stars coming from the Galactic Centre. Applying the algorithm to the TGAS subset of the first release of the *Gaia* satellite, we have identified a total of 80 objects with a predicted probability $> 90\%$ of being a HVS, and for 30 of those we were able to find a radial velocity measurement from literature. We followed up spectroscopically 22 candidates at the Isaac Newton Telescope, for a total of 47 stars with a reliable radial velocity determination. Our stars show a uniform distribution across the sky, showing that the algorithm is not selecting sources in a preferential direction.

With a Bayesian approach we inferred distances from parallax for all our candidates, and total velocities in the Galactic rest frame were computed in order to establish their nature and origin. Without pre-selection of data we were able to recover several objects already noted and discussed in literature because of their remarkably high velocities. We found 45 candidates with a median rest frame velocity > 150 km s^{-1}, 14 of them having $v > 400$ km s^{-1}, and a subset of 5 stars has a probability $> 50\%$ of being unbound from the Milky Way, with median velocities up to ~ 900 km s^{-1}.

Tracing back orbits with Monte Carlo simulations in different Galactic potentials we found:

- 6 stars being consistent with coming from the Galactic Center. One of these stars, with a velocity of ~ 520 km s^{-1}, has a probability $> 50\%$ of being unbound from the Galaxy (HVS), while the others are bound hypervelocity star candidates, with velocities > 360 km s^{-1};
- 5 stars with high velocities but trajectories not consistent with coming from the Galactic Centre: these stars are runaway star candidates. Two of these stars have probabilities $> 50\%$ of being unbound from the Milky Way, and are therefore classified as hyper runaway stars. The explosion of a supernova in a binary system is a plausible mechanism for having accelerated these stars to such high velocities. It is remarkable that a good fraction of our RS candidates have velocities consistent with being higher than the escape velocity from the Galaxy, since these stars are thought to be extremely rare: approximately 1 for every 100 HVSs (Bromley et al. 2009; Perets & Šubr 2012; Kenyon et al. 2014; Brown 2015);
- 4 stars with a velocity and origin highly dependent on the assumed distance estimate. Two of these stars have a high probability of being unbound from the Milky Way.

At the moment, positive identifications are strongly hampered by large uncertainties in distance and limited information on the age and flight time of our sources. The advent of future *Gaia* releases will dramatically increase the number of HVSs we expect to find. The more accurate parallax determination, less affected by systematics, will allow us to decrease error bars and to identify in a clearer way the most interesting objects, narrowing down their ejection location. The brightest stars in the catalogue will also have a radial velocity measurement, allowing us to train the neural network adding this precious information as an extra feature to the astrometric solution.

We are currently working to increase the quality of the training set of mock HVSs, considering not only radial trajectories, but modelling orbits of bound stars and including deviations due to the disc and to a possible triaxiality of the bulge (e.g. McWilliam & Zoccali 2010) and/or the halo (e.g. Bullock 2002; Helmi 2004). Another natural advancement would be to model runaway and halo stars to create mock populations, and then to perform a multiclass classification analysis in order to decrease the number of false positives and achieve a more precise classifier.

Acknowledgements

We thank J. Brinchmann and A. Patruno for useful discussion and comments, U. Bastian and L. Lindegren for suggestions and advice on the use of *Gaia* astrometric data and on distances determination, and T. Astraatmadja and C. Bailer-Jones for the implementation of the Milky Way Prior. We also thank Warren Brown for the careful reading of the manuscript and his useful comments. TM and EMR acknowledge support from NWO TOP grant Module 2, project number 614.001.401. ES and KY gratefully acknowledge funding by the Emmy Noether program from the Deutsche Forschungsgemeinschaft (DFG). This work has made use of data from the European Space Agency (ESA) mission *Gaia* (`http://www.cosmos.esa.int/gaia`), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, `http://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. The Isaac Newton Telescope is operated on the island of La Palma by the Isaac Newton Group of Telescopes in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias. 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).

.1 *Gaia* Identifiers

In Table 4 we present Tycho 2, Hipparcos, and *Gaia* identifiers for the candidates observed at the INT (Table 3.1) and for the stars with $v > 350$ km s^{-1}(Table 3.3).

.2 Assuming Different Priors on Distance

One could argue that assuming a three-components stellar density (bulge + disc + halo) for our Galaxy $\rho_{\rm MW}(\boldsymbol{d})$, as in Equation 3.10, is not appropriate to model the spatial distribution of HVSs, a population of stars that, by definition, is not distributed according to the density profile of the Milky Way. Therefore in this appendix we discuss the implication of assuming different priors on distances $P(\boldsymbol{d})$ in the MCMC sampling described in Section 3.5. In practice we adopt two different priors and we test the impact of these choices on our results: an exponential decreasing prior $P_{\rm exp}(d)$, and a prior

Table 4: Tycho 2, Hipparcos, and *Gaia* identifiers of stars observed at the INT and of high velocity candidates.

Tycho 2 ID	Hipparcos ID	*Gaia* ID
1071-404-1	98492	4299974437593772672
2282-208-1		314799593600582656
2292-1267-1		316401685121779712
2298-66-1		317585859144818688
2320-470-1		329685915888890880
2348-333-1		137859551029399040
2376-691-1		172747742173867904
2393-1001-1		180650104040989568
2456-2178-1		893048667206860800
2818-556-1		347908809291960832
2822-1194-1		348293878879518848
3163-1181-1		2081319505008076416
3263-733-1		377741720849393920
3285-1422-1		353451584846863104
3330-120-1	17648	248695099116287872
3661-974-1		422054582068454016
3744-1546-1		470781741956237696
3945-1023-1		2187713404073484288
3983-1873-1	111334	2000722382112691456
4032-1542-1		509654254003883776
4307-1106-1		539315160710386944
4507-1461-1		569097391651702656
4509-1013-1		550795677011227648
4515-1197-1		552553933541803008
4521-322-1		5681895730004745472
49-1326-1		2503868695508755840
5890-971-1	20214	3172032703298013696
7111-718-1		5590900663125136000
8374-757-1	93476	6662886601414152448
8422-875-1	104613	6483680327939151488
9404-1260-1	46120	5195968559017084160

specifically tailored for HVSs, the *HVS prior* $P_{\rm HVS}(\boldsymbol{d})$, that we introduce in this paper.

Astraatmadja & Bailer-Jones (2016a) show that an exponential decreasing prior

$$P_{\rm exp}(d) \propto d^2 \exp\left(-\frac{d}{L}\right) \tag{15}$$

with $L = 1.35$ kpc gives a better performance in terms of RMS errors compared to the MW prior, when resulting distance estimates are compared with GUMS simulated data. This choice assumes that the disc has the same scale-height as the scale-length, and clearly it is not an accurate description of the MW. We find that this prior overestimates distances for the majority of our candidates, with values well above the spectroscopic ones. This is evident in top panel of Figure 10, where for distances greater than ~ 600 pc we can see that median values obtained with the exponential prior are always higher than the ones derived with the MW prior. This is due to the choice of L, which sets the exponential cut-off of the distribution. Since $L = 1.35$ kpc is higher than the typical distance of stars in the TGAS calatogue, this prior biases our candidates towards greater distances, and thus towards higher total velocities, proper motions and radial velocities being equal.

Assuming a continuous and isotropic ejection of HVSs from the Galactic Centre, the number density of HVSs goes approximately as $1/r^2$, where r is the galactocentric radius (Brown 2015). Following Equation 3.10 we therefore construct the HVS prior as:

$$P_{\rm HVS}(d, l, b) \propto \left(\frac{d}{r(d, l, b)}\right)^2 p_{\rm obs}(d, l, b), \tag{16}$$

with $r(d, l, b) = \sqrt{d^2 + d_\odot^2 - 2dd_\odot \cos(l)\cos(b)}$ and $d_\odot = 8$ kpc. When deriving distances and total velocities with this prior, we find again results to be consistent with the ones derived using the MW prior, but uncertainties are considerably larger, and this prior overestimates distances for further stars, as shown in bottom panel of Figure 10.

In the end, we choose to adopt the MW prior for presenting our results since it allows us to maintain a conservative approach: because of large uncertainties, we only interpret our candidates as HVSs at the end of the kinematic analysis, without biasing our distances and velocities using that assumption.

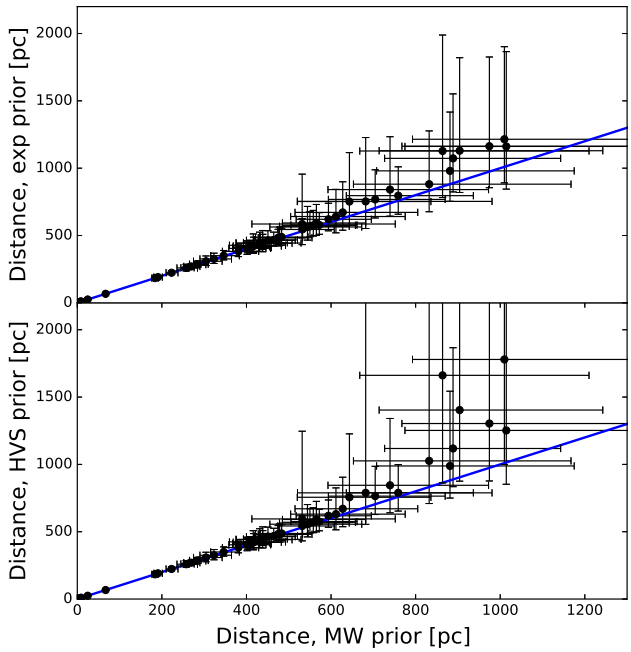


Figure 10: Comparison of distances obtained using the MW prior, on the x-axis, and the exponential decreasing (HVS) prior, y-axis on the top (bottom) panel. The blue line corresponds to equal estimates.

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

T. Marchetti, *E.M. Rossi, A.G.A. Brown 2018, MNRAS*

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 probability greater than 50% of being unbound from the Milky Way. This cut results in a clean sample of 125 sources with reliable astrometric parameters and radial velocities. Of these, 20 stars have probabilities greater than 80% of being unbound from the Galaxy. On this latter sub-sample, we perform orbit integration to characterize the stars' orbital parameter distributions. As expected given the relatively small sample size of bright stars, we find no hypervelocity star candidates, stars that are moving on orbits consistent with coming from the Galactic Centre. Instead, we find 7 hyper-runaway star candidates, coming from the Galactic disk. Surprisingly, the remaining 13 unbound stars cannot be traced back to the Galaxy, including two of the fastest stars (around 700 km s^{-1}). If confirmed, these may constitute the tip of the iceberg of a large extragalactic population or the extreme velocity tail of stellar streams.

4.1 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 velocity 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).

A first class of objects that can be found in the high tail of the total velocity distribution is fast halo stars. Their measured dispersion velocity is around 150 km s^{-1}(Smith et al. 2009; Evans et al. 2016), therefore 3-σ outliers can exceed 450 km s^{-1}, 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 disk 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 ejected 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 a few tens of km s^{-1}, and even if several hundreds of km s^{-1}can be attained for the most extreme systems (Portegies Zwart 2000; Przybilla et al. 2008; Gvaramadze et al. 2009; Gvaramadze & Gualandris 2011; Silva & Napiwotzki 2011), 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 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 were 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 ~ 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 ~ 20 unbound HVSs with velocities ~ 300 - 700 km s^{-1}have been discovered by 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 distances 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 uncertain 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 disk (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. 2019) 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. 2018b). 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. 2017a; Erkal et al. 2019), orbiting the Milky Way with a velocity ~ 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 ~ 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. 2018b).

At the moment, the fastest star discovered in our Galaxy is US 708, traveling away from the Milky Way with a total velocity ~ 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. 2016b, 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. 2019). 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 ≈ 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. 2019).

In this paper we search for the fastest stars in the Milky Way, within the sample of ~ 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 4.2, 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 4.3. In Section 4.4, we focus on the high velocity stars in the sample, and then in Section 4.5 we concentrate on the stars with a probability greater than 80% of being unbound from the Galaxy, discussing individually the most interesting candidates. Finally, we conclude and discuss our results and findings in Section 4.6.

4.2 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 choose to separate the discussion on how we determine distances and total velocities of stars with $0 < f \leqslant 0.1$ (the "*low-f* sample") and of those with either $f > 0.1$ 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.

4.2.1 The "*low-f* Sample"

5393495 out of 7183262 stars ($\sim 75\%$) with radial velocity measurement in *Gaia* DR2 have a relative error in parallax $0 < f \leqslant 0.1$. For this major-

ity 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:

$$\mathbf{m} = [\mu_{\alpha*}, \mu_\delta, \varpi] \tag{4.1}$$

and with covariance matrix:

$$\Sigma = \begin{pmatrix} \sigma^2_{\mu_{\alpha*}} & \sigma_{\mu_{\alpha*}}\sigma_{\mu_\delta}\rho(\mu_{\alpha*}, \mu_\delta) & \sigma_{\mu_{\alpha*}}\sigma_\varpi\rho(\mu_{\alpha*}, \varpi) \\ \sigma_{\mu_{\alpha*}}\sigma_{\mu_\delta}\rho(\mu_{\alpha*}, \mu_\delta) & \sigma^2_{\mu_\delta} & \sigma_{\mu_\delta}\sigma_\varpi\rho(\mu_\delta, \varpi) \\ \sigma_{\mu_{\alpha*}}\sigma_\varpi\rho(\mu_{\alpha*}, \varpi) & \sigma_{\mu_\delta}\sigma_\varpi\rho(\mu_\delta, \mu_\varpi) & \sigma^2_\varpi \end{pmatrix}, \tag{4.2}$$

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 1000 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 peculiar 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). To save computational time, we do not sample within the uncertainties of the Solar position and motion. We verify that this does not considerably affect our results. 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 estimate 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.

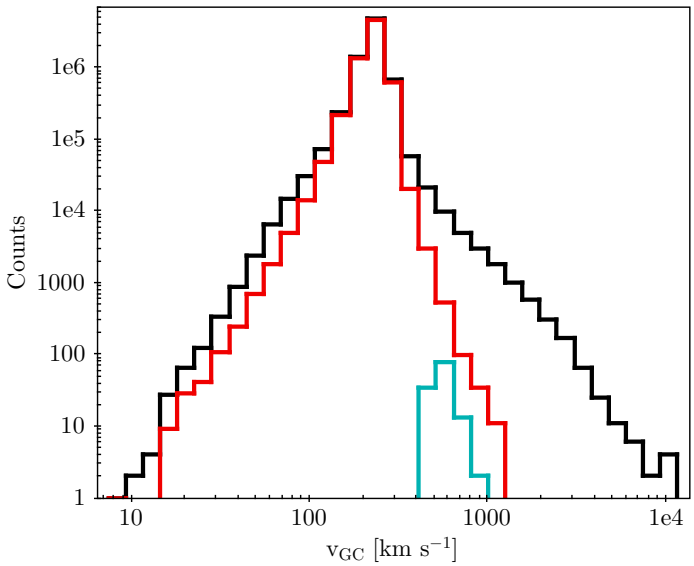


Figure 4.1: Histogram of median total velocities in the Galactic rest frame for all the ~ 7 million stars with three-dimensional velocity by *Gaia* DR2 (black). The red line corresponds to those stars with a relative error on total velocity in the Galactic rest-frame below 30%, while the cyan line refers to our "clean" sample of high velocity stars (see discussion in Section 4.4).

We compute the escape velocity from the Galaxy at each position using the Galactic potential model introduced and discussed in Section 4.4.1.

4.2.2 The "*high-f* Sample"

A more careful analysis is required for 1789767 stars ($\sim 25\%$) with either $f > 0.1$ 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); Bailer-Jones 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*:

$$P(d) \propto d^2 \exp\left(-\frac{d}{L}\right), \tag{4.3}$$

which has been shown to perform best for stars further out than ~ 2 kpc (Astraatmadja & Bailer-Jones 2016b), that is the expected distance of stars with a large relative error on parallax (see Appendix .1). The value of the scale length parameter L is fixed to 2600 pc, and we refer the reader to the discussion in Appendix .1 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:

$$P(d|\varpi, \sigma_\varpi) \propto P(\varpi|d, \sigma_\varpi)P(d), \tag{4.4}$$

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

$$P(\varpi|d, \sigma_\varpi) \propto \exp\left[-\frac{1}{2\sigma_\varpi^2}\left(\varpi - \frac{1}{d}\right)\right]. \tag{4.5}$$

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:

$$\mathbf{m} = [\mu_{\alpha*}, \mu_\delta, 1/d] \tag{4.6}$$

and covariance matrix given by equation (4.2). The prior distribution on distance is given by equation (4.3), 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 (4.4), to achieve a fast convergence of the chain. We run 500 burn-in steps to let the walkers explore the parameter space, and then we use the final positions as initial conditions for the proper MC 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 4.2.1. We check that the mean acceptance fraction (i.e. the fraction of steps accepted for each walker) is between 0.25 and 0.5 as a test for the convergence of each MC chain (Foreman-Mackey et al. 2013).

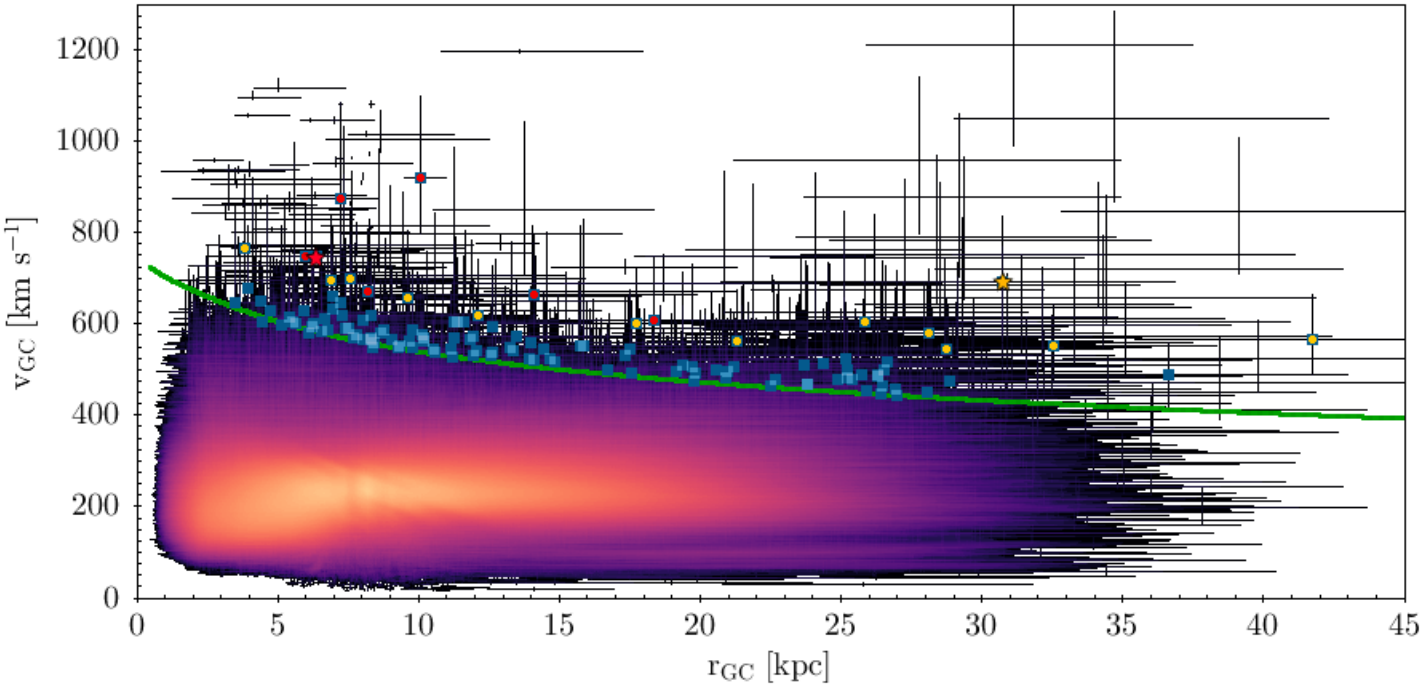


Figure 4.2: Total velocity in the Galactic rest-frame $v_{\rm GC}$ as a function of Galactocentric distance $r_{\rm GC}$ for all the 6884304 stars in *Gaia* DR2 with relative error on total velocity < 0.3. Colour is proportional to the logarithmic number density of stars. The green solid line is the median posterior escape speed from the adopted Galactic potential (Section 4.4.1). We overplot in blue the "clean" high velocity star sample introduced in Section 4.4. Red and yellow points correspond, respectively, to the Galactic and extragalactic candidates discussed in Section 4.5. *Gaia* DR2 5932173855446728064 (*Gaia* DR2 1396963577886583296) is marked with a red (yellow) star.

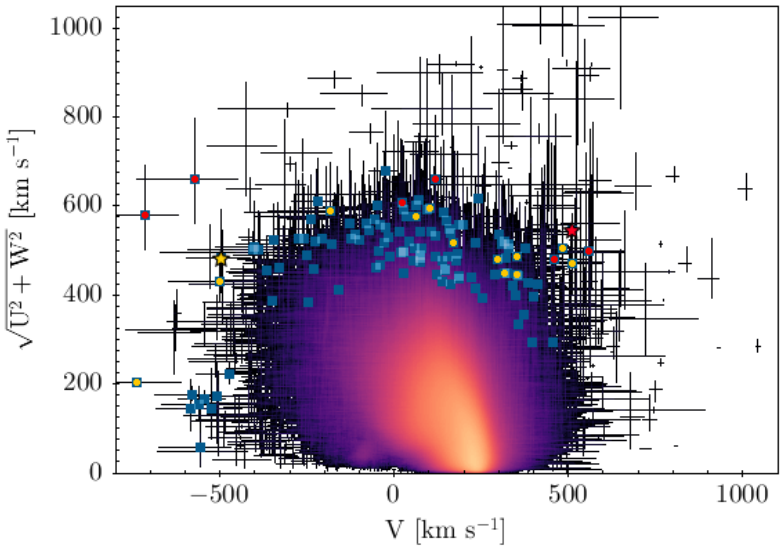


Figure 4.3: Toomre diagram for the same stars plotted in Fig. 4.2.

4.3 The Total Velocity Distribution of Stars in *Gaia* DR2

Using the approach discussed in Section 4.2, 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 .2.

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 $\sigma_{v_{GC}}/v_{GC} < 0.3$, where $\sigma_{v_{GC}}$ is estimated summing in quadrature the lower and upper uncertainty on v_{GC}. This cut results into a total of 6884304 stars, $\sim$ 96% of the original sample of stars. Figure 4.1 shows the total velocity distribution of the median Galactic rest frame total velocity v_{GC} for the original sample of 7183262 stars (black line) and for the stars with a relative error on total velocity below 30% (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 relatively more uncertain measurements by *Gaia*. Nevertheless we note the presence of a high velocity tail extending up to and above $\sim$ 1000 km s^{-1}surviving the cut. We

will now focus only on stars with $\sigma_{v_{\rm GC}}/v_{\rm GC} < 0.3$.

To highlight visually possibly unbound objects, we plot in Figure 4.2 the total velocity for all stars as a function of the Galactocentric distance $r_{\rm GC}$, and we overplot the median escape speed from the Galaxy with a green solid line, computed using the Galactic potential model introduced in Section 4.4.1. Datapoints correspond to the medians 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 510 stars to have probabilities greater than 50% of being unbound from the Galaxy (but note the large errorbars). In particular, 212 (103) stars are more than 1-σ (3-σ) away from the Galactic escape speed.

Figure 4.3 shows the Toomre diagram for all the $\sim$ 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.

4.4 High Velocity Stars in *Gaia* DR2

We now focus our interest towards high velocity stars, which we define as stars with a probability $P_{\rm ub} > 0.5$. Since we are interested in kinematic outliers, we have to pay particular attention not to be contaminated by data processing artifacts and/or 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 $\sigma_{v_{\rm GC}}/v_{\rm GC} < 0.3$ introduced in Section 4.3):

1. astrometric_gof_al < 3;
2. astrometric_excess_noise_sig $\leqslant$ 2;
3. $-0.23 \leqslant$ mean_varpi_factor_al $\leqslant 0.32$;
4. visibility_periods_used > 8;
5. rv_nb_transits > 5.

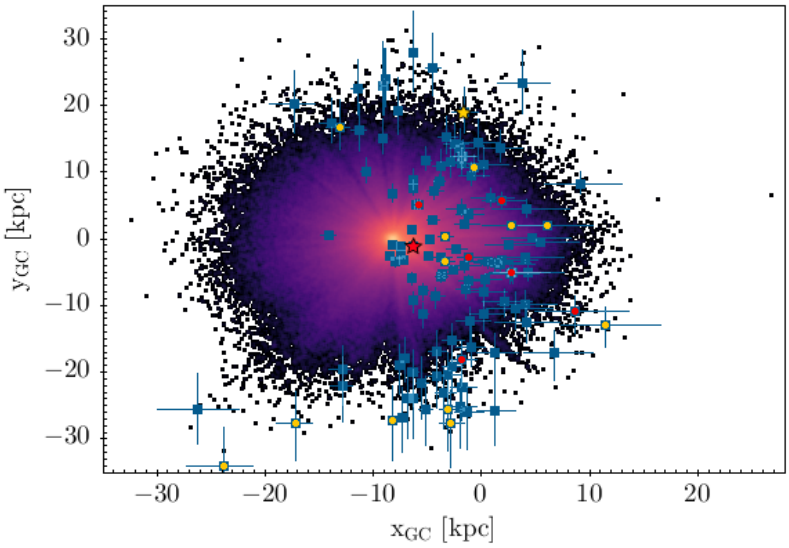


Figure 4.4: Distribution of the ∼ 7 million stars on the Galactic plane. The Sun is located at $(x_{GC}, y_{GC}) = (-8.2, 0)$ kpc. Colours are the same as in Fig. 4.2.

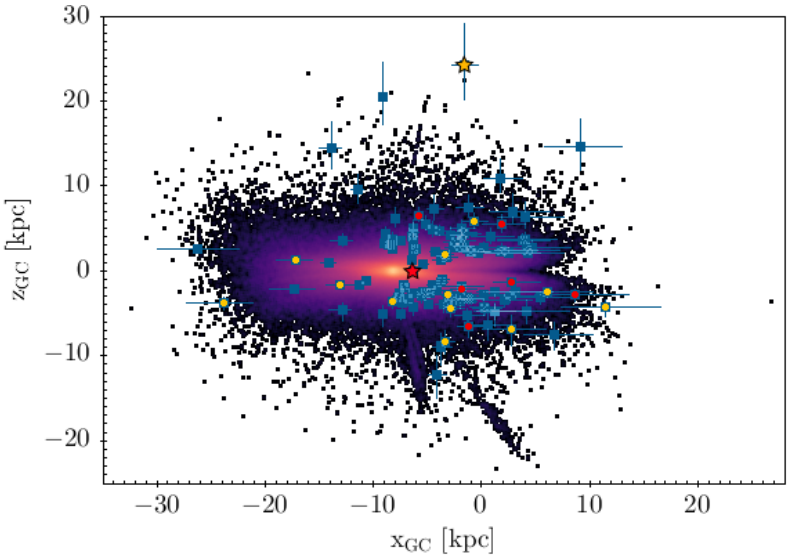


Figure 4.5: Same as Fig. 4.4, but showing the distribution of the stars in the (x_{GC}, z_{GC}) plane. The Sun is located at $(x_{GC}, z_{GC}) = (-8200, 25)$ pc. Colors are the same as in Fig. 4.2.

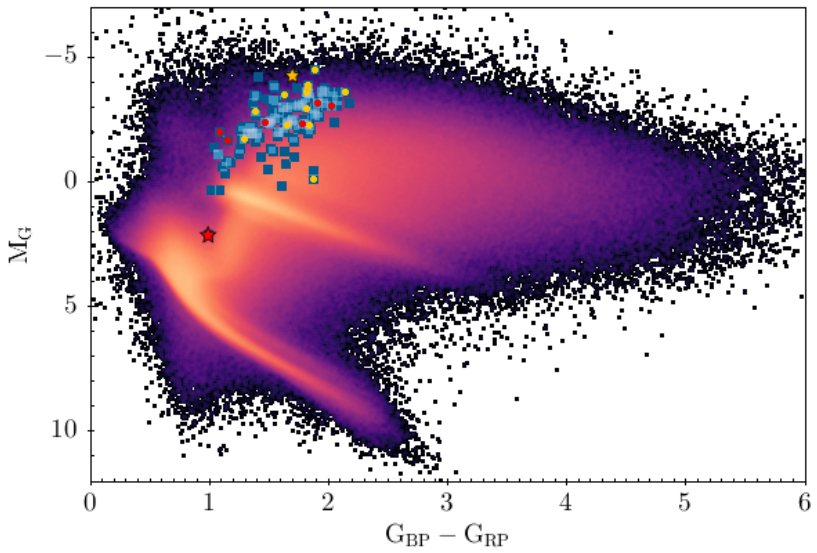


Figure 4.6: HR diagram for all the ∼ 7 million stars in *Gaia* DR2 with a radial velocity measurement. Colours are the same as in Fig. 4.2.

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 a reasonable number of times (5) by *Gaia* to determine its radial velocity. Further details on the parameters used to filter out possible contaminants and the reasons behind the adopted threshold values can be found in the *Gaia* data model[1]. Applying these cuts and with the further constrain on the unbound probability $P_{\rm ub} > 0.5$, we are left with a clean final sample of 125 high velocity stars. We also verify that the quality cuts C.1 and C.2 introduced in Appendix C of Lindegren et al. (2018b), 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. (2018b) does not select any of our candidates. Looking at Fig. 4.2, where this clean sample of 125 stars is highlighted with blue squares, we can see how these cuts filter out most of the stars with exceptionally high velocities, which are therefore

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

likely to be instrumental artifacts. This is also evident in Fig. 4.1, where the Galactic rest-frame total velocity distribution of the 125 high velocity stars is shown with a cyan line.

We present distances, total velocities, and probability of being unbound for all the 105 stars wih $0.5 < P_{\rm ub} \leqslant 0.8$ in Appendidx .3, Table 5. Stars with $P_{\rm ub} > 0.8$ are presented and discussed in detail in Section 4.5.

The spatial distribution of these 125 high velocity stars in our Galaxy is shown in Fig. 4.4, where we overplot the position on the Galactic plane of this subset of stars with blue markers 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 $\lesssim 15$ kpc from the GC. 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 corresponds to the Sun's position, correlating with the underlying distribution of all the stars. In Fig. 4.5, we plot the same but in the $(x_{\rm GC}, z_{\rm GC})$ plane. Most of our high velocity stars lie away from the stellar disk.

Fig. 4.6 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 $G_{\rm BP} - G_{\rm RP}$, while on the y-axis we plot the absolute magnitude in the *Gaia* G band M_G, computed assuming the median of the posterior distance distribution. 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). We can see that the great majority of our stars are giants stars. This is consistent with recent findings of Hattori et al. (2018a); Hawkins & Wyse (2018), which confirm some of these candidates as being old (> 1 Gyr), metal-poor giants ($2 \leqslant$ [Fe/H] $\leqslant 1$).

4.4.1 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 4.2. We integrate each orbit back in time for a total time of 1 Gyr, with a fixed time-step of 0.1 Myr, using

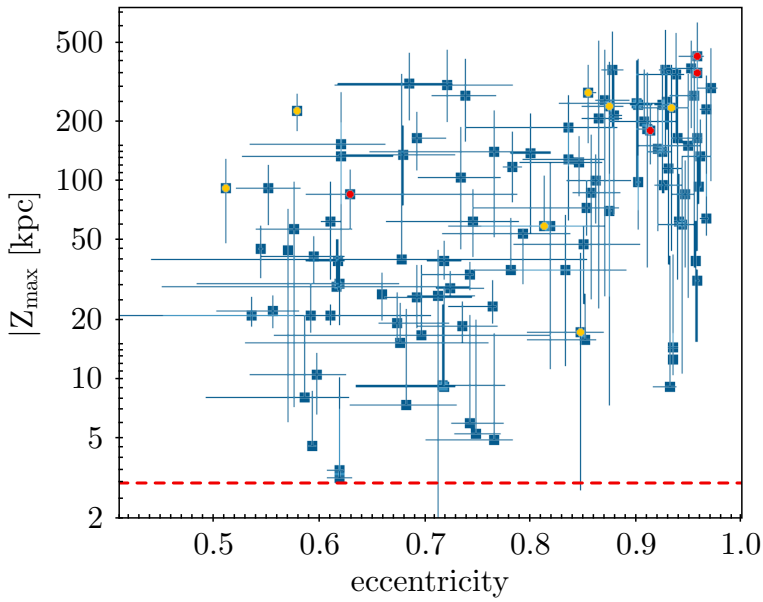


Figure 4.7: Absolute value of the maximum height above the Galactic plane $|Z_{\rm max}|$ as a function of eccentricity for the high velocity sample of stars. The yellow horizontal dashed line corresponds to $Z_{\rm max} = 3$ kpc, the edge of the thick disk (Carollo et al. 2010). Colours are the same as in Fig. 4.2.

Table 4.1: Parameters for the GALA potential *MilkyWayPotential*.

Component	Parameters
Bulge	$M_b = 5.00 \cdot 10^9\ \mathrm{M}_\odot$
	$r_b = 1.00$ kpc
Nucleus	$M_n = 1.71 \cdot 10^9\ \mathrm{M}_\odot$
	$r_n = 0.07$ kpc
Disk	$M_d = 6.80 \cdot 10^{10}\ \mathrm{M}_\odot$
	$a_d = 3.00$ kpc
	$b_d = 0.28$ kpc
Halo	$M_h = 5.40 \cdot 10^{11}\ \mathrm{M}_\odot$
	$r_s = 15.62$ kpc

the gala potential *MilkyWayPotential*. This is a four components Galactic potential model consisting of a Hernquist bulge and nucleus (Hernquist 1990):

$$\phi_b(r_{\mathrm{GC}}) = -\frac{GM_i}{r_{\mathrm{GC}} + r_i}, \tag{4.7}$$

where $i = b, n$ for the bulge and the nucleus, respectively, a Miyamoto-Nagai disk (Miyamoto & Nagai 1975):

$$\phi_d(R_{\mathrm{GC}}, z_{\mathrm{GC}}) = -\frac{GM_d}{\sqrt{R_{\mathrm{GC}}^2 + \left(a_d + \sqrt{z_{\mathrm{GC}}^2 + b_d^2}\right)^2}}, \tag{4.8}$$

and a Navarro-Frenk-White halo (Navarro et al. 1996):

$$\phi_h(r_{\mathrm{GC}}) = -\frac{GM_h}{r_{\mathrm{GC}}} \ln\left(1 + \frac{r_{\mathrm{GC}}}{r_s}\right). \tag{4.9}$$

The parameters are chosen to fit the enclosed mass profile of the Milky Way (Bovy 2015a), and are summarized in Table 4.1. We then derive the pericenter distance and, for bound MC realizations, the apocenter distance and the eccentricity of the 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. 4.7, 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 $Z_{\mathrm{max}} = 3$

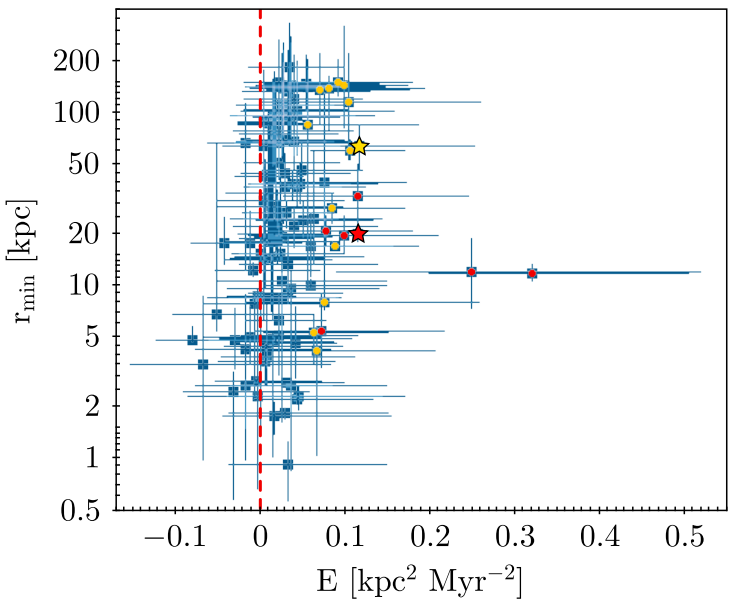


Figure 4.8: Minimum crossing radius $r_{\rm min}$ versus energy E for the 125 high velocity stars. The vertical dashed line separates unbound ($E > 0$) from bound ($E < 0$) orbits. Colors are the same as in Fig. 4.2.

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 ~ 0.8. Most of these stars span a large range of $Z_{\rm max}$, 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 $z_{\rm GC} = 0$) in the orbital traceback of our high velocity star sample. For each MC realization, we then define the crossing radius r_c as:

$$r_c = \sqrt{x_c^2 + y_c^2}, \tag{4.10}$$

where x_c and y_c are the Galactocentric coordinates of the orbit $(x_{\rm GC}, y_{\rm GC})$ at the instant when $z_{\rm GC} = 0$. In the case of multiple disk crossings during the orbital trace-back, we define $r_{\rm min}$ 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 $v_{\rm ej}$: the velocity of the star at the minimum crossing radius, and the flight time $t_{\rm f}$: the time needed to travel from the observed position to the disk crossing happening closest to the GC.

In Fig. 4.8, we plot $r_{\rm min}$ as a function of the orbital energy E. The red dashed line coincides with the separation region between bound and unbound orbits. The majority of candidates are traveling on unbound orbits ($E > 0$), and we can see a few stars with remarkably high values of the energy: 25 stars are unbound at more than 1 sigma significance, and 1 star (*Gaia* DR2 5932173855446728064) is unbound at more than 3 sigma significance.

4.5 Unbound Stars: Hypervelocity and Hyper Runaway Star Candidates

We now focus our search on possible unbound stars, defined as the subsample of clean high velocity stars with $P_{\rm ub} > 80\%$. This amounts to a total of 20 objects. Observed properties from *Gaia* DR2, distances, and total velocities for these stars are summarized in Table 2. Fig. 4.9 shows the position in Galactocentric cylindrical coordinates of these high velocity star candidates. The length of the arrows is proportional to the total velocity of each star in the Galactic rest frame. We note that for most of our candidates (18 out of 20 stars) the parallax uncertainty is smaller than the quoted parallax

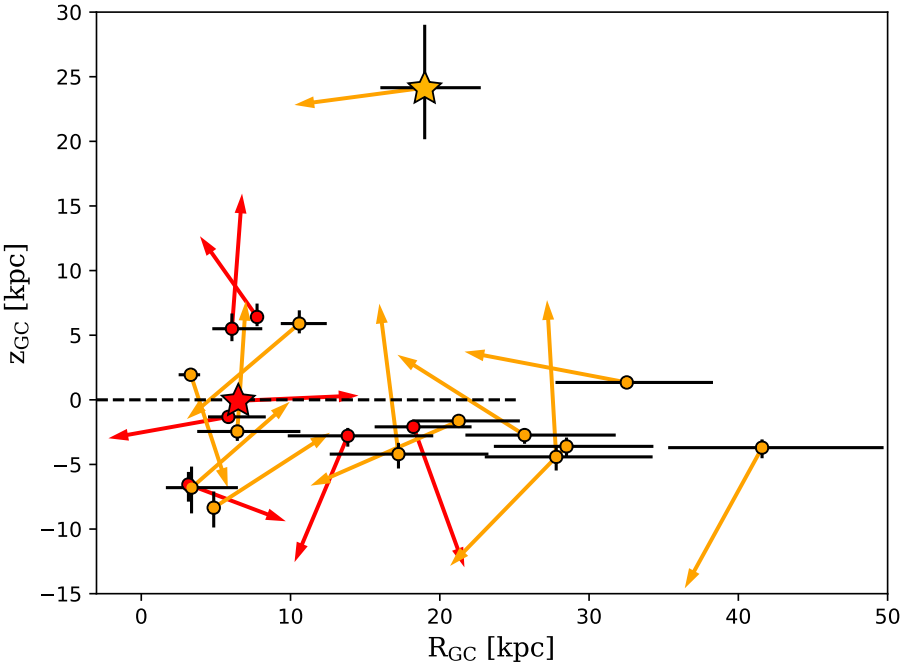


Figure 4.9: Position of the 20 high velocity stars with $P_{\rm ub} > 80\%$ in Galactocentric cylindrical coordinates $(R_{\rm GC}, z_{\rm GC})$. Arrows point to the direction of the velocity vector of the stars in this coordinate system, and the arrow's length is proportional to the total velocity of the star in the Galactic rest-frame. Red (yellow) points and arrows mark the 7 (13) Galactic (extragalactic) candidates with $P_{\rm MW} > 0.5$ ($P_{\rm MW} < 0.5$). *Gaia* DR2 5932173855446728064 (*Gaia* DR2 1396963577886583296) is marked with a red (yellow) star. The Sun is located at $(R_{\rm GC}, z_{\rm GC}) = (8200, 25)$ pc. The horizontal dashed line denotes the position of the Galactic plane, and extends up to the edge of the stellar disk, which we take to be at 25 kpc (Xu et al. 2015).

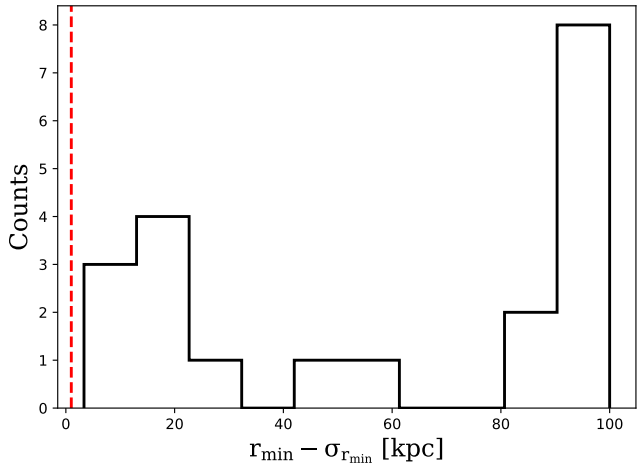


Figure 4.10: Histogram of the median minimum crossing radius $r_{\rm min}$ minus the correspondent lower uncertainty $\sigma_{r_{\rm min,l}}$ for the sample of 20 high velocity stars with $P_{\rm ub} > 0.8$. The vertical dashed line corresponds to $(r_{\rm min} - \sigma_{r_{\rm min,l}}) = 1$ kpc, our boundary condition for not rejecting the GC origin hypothesis for the HVS candidates (see discussion in Section 4.5). $(r_{\rm min} - \sigma_{r_{\rm min,l}}) > 1$ kpc for all the 20 stars, therefore there are no HVS candidates.

zeropoint of -0.029 mas, as estimated by *Gaia*'s observations of quasars (Lindegren et al. 2018b). We discuss the impact of considering this negative offset in the analysis of our stars in Appendix .4. We further discuss the impact of systematic errors for our sample of 20 unbound candidates in Appendix .5.

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 $r_{\rm min}$ 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 $P_{\rm MW}$ for a star to come from the Milky Way as the fraction of MC realizations resulting in a minimum crossing radius within the edge of the stellar disk: $r_{\rm min} < r_{\rm disk}$, where $r_{\rm disk} = 25$ kpc (Xu et al. 2015). This probability is useful to flag candidates of possible extragalactic origin, which we define as those stars with $P_{\rm MW} < 0.5$. This subset of 13 stars, if their high velocity is confirmed, could either originate as RS/HRS/HVS from the LMC (Boubert & Evans 2016; Boubert et al. 2017a; Erkal et al. 2019), or could be the result of the tidal disruption of a dwarf galaxy interacting with the Milky Way (Abadi et al. 2009). Stars with a Galactic and extragalactic origin

are marked in Fig. 4.9 with red and yellow points, respectively. Stars with a Galactic origin have trajectories pointing away from the stellar disk. On the other hand, extragalactic stars are pointing either towards the disk, or are consistent with coming from regions of no current active star formation (i.e. the outer halo).

4.5.1 Galactic Stars

7 of the 20 possible unbound stars have $P_{\rm MW} > 0.5$, and therefore are consistent with being ejected from the stellar disk of the Milky Way. These stars, given their extremely high velocities, could be either HVS or HRS candidates.

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 $r_{\rm min} - \sigma_{r_{\rm min,l}} < 1$ kpc ($r_{\rm min} - \sigma_{r_{\rm min,l}} > 1$ kpc), where $r_{\rm min}$ denotes the median of the distribution, and $\sigma_{r_{\rm min,l}}$ 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 4.10 shows the histogram of the median minimum disk crossing $r_{\rm min}$ minus the lower uncertainty $\sigma_{r_{\rm min,l}}$ for all the 20 stars with $P_{\rm ub} > 0.8$. A vertical red dashed line corresponds to the value 1 kpc, which we use to define HVS candidates.

We find that all of these 7 stars have orbits that, when integrated back in time, are not consistent with coming from the GC. Therefore, according to our classification criterion, there are no stars classified as HVS candidates. The absence of HVS candidates in the subset of *Gaia* DR2 with radial velocities was anticipated by predictions by Marchetti et al. (2018b), analyzing the Hills mock catalogue of HVSs. This is due to the fact that the expected number density of HVSs generated via the Hills' mechanism is expected to increase linearly with increasing galactocentric distance (Brown 2015), and the majority of HVSs in the Milky Way are too faint to have a radial velocity measurement from *Gaia* DR2. We cannot exclude the presence of *bound* HVSs in the subset of ~ 7 million stars considered in this work, but their identification is not trivial because of their complex orbits and lower velocities. About 20 BHVSs are expected to have radial velocities from *Gaia* DR2 (Marchetti et al. 2018b), but their identification is beyond the scope of this manuscript.

All the 7 Galactic stars are therefore HRS candidates (red circles in Fig. 4.2 and following plots). One particular HRS candidate that is worth mentioning is *Gaia* DR2 5932173855446728064 (marked with a red star in Fig.

4.2 and following). This star has an exceptionally well constrained total velocity[2], $v_{\rm GC} = 747^{+2}_{-3}$ km s^{-1}, which results in a probability of being unbound ≈ 1. This star most likely was ejected in the thin disc of the Milky Way.

We note that 5 of the 7 HRS candidates with a Galactic origin have $P_{\rm ub} > 90\%$. 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). This is correct in the context of the Milky Way as a whole. In this study we only focus on bright sources ($G_{\rm RVS} < 12$), therefore we maximize the probability of observing stars ejected from the stellar disk. The HVS population is instead expected to be much fainter than this magnitude cut (Marchetti et al. 2018b). Since estimates on the expected HRS population in *Gaia* are currently missing, at the moment it is not clear whether this tension is real, and/or if other ejection mechanisms are needed (e.g. Irrgang et al. 2018).

4.5.2 Extragalactic Stars

13 of the 20 $P_{\rm ub} > 80\%$ stars have probabilities $< 50\%$ of intersecting the Milky Way stellar disk when traced back in time, therefore an extragalactic origin is preferred. A possible 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 (Abadi et al. 2009).

The extragalactic star with a highest probability of being unbound from our Galaxy is Gaia DR2 1396963577886583296, with a total velocity ~ 700 km s^{-1}, resulting in a probability $P_{\rm ub} = 0.98$. We mark this source with a yellow star in Fig. 4.2 and following. This star is at ~ 30 kpc from the GC, with an elevation of ~ 25 kpc above the Galactic plane.

4.6 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

[2]Because of the small uncertainties, we repeat the total velocity determination for *Gaia* DR2 5932173855446728064 sampling within the uncertainties of the Sun position and motion (see discussion in Section 4.2.1). The result is $v_{\rm GC} = (747 \pm 7)$ km s^{-1}, in agreement with the previous estimate.

stars as those stars with an estimated probability of being unbound from the Milky Way $P_{\rm ub} > 50\%$, resulting in a total of 125 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 these 125 stars, we found the following.

1. 20 stars have predicted probabilities $P_{\rm ub} > 80\%$. The observed and derived kinematic properties of these stars are summarized in Table 2, and are discussed in Section 4.5.

2. None of these 20 stars is consistent with coming from the inner 1 kpc, so there are no HVS candidates. This is consistent with estimates presented in Marchetti et al. (2018b).

3. 7 out of the 20 stars with $P_{\rm ub} > 0.8$, when traced back in time in the Galactic potential, originate from the stellar disk of the Milky Way. These stars are HRS candidates.

4. 13 out of the 20 unbound candidates have probabilities $< 50\%$ to originate from the stellar disk 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.

Another possibility that we cannot rule out is that a subset of these 20 stars is actually gravitationally *bound* to the Milky Way. Recent high-resolution spectroscopic followups showed that some of these stars are actually indistinguishable from halo stars from a chemical point of view (see Hawkins & Wyse 2018), therefore if they are actually bound, this would in turn imply a more massive Milky Way (Hattori et al. 2018a; Monari et al. 2018), a possiblity that cannot be ruled out (e.g. Wang et al. 2015). Otherwise, a confirmation of the global parallax zeropoint measured with quasars could lower down their total velocities, resulting in the same effectAs discussed in Appendix .4, including this parallax offset results in 14 (4) stars with an updated $P_{\rm ub} > 50\%$ ($P_{\rm ub} > 80\%$). The choice of not considering the parallax zero point in the main text is therefore a conservative choice, which ensures us that all the high velocity stars in the subset of *Gaia* DR2 with radial velocities are actually included in this work. In Appendix .5 we show how including systematic errors in parallax can significantly lower

the distances and total velocities for our candidates, but we want to stress that the adopted parameters might be too pessimistic for the stars considered in this paper (Lindegren et al. 2018a). Follow-up observations with ground based facilities and/or future data releases of the *Gaia* satellite will help us confirming or rejecting their interpretation as kinematic outliers.

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. (2018b) 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.

Acknowledgements

We thank the anonymous referee for his/her comments, which greatly improved the quality of this manuscript. We also 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 Col-

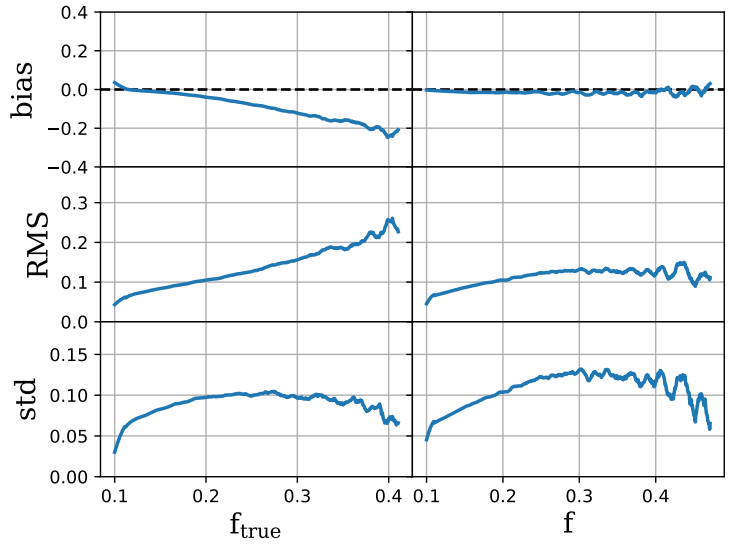


Figure 11: Bias, RMS, and standard deviation of the estimator x_0 as a function of $f_{\rm true} = \sigma_\varpi d_{\rm true}$ (left panel) and $f = \sigma_\varpi/\varpi$ (right panel). The modes of the posterior distributions are estimated using the exponentially decreasing prior with a characteristic scale length $L = 2600$ pc.

laboration 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.

.1 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 4.2). 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

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

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 352010 of the 7 million stars to have $f = \sigma_\varpi/\varpi > 0.1$. We can see that this value is about 5 times smaller than the one found in *Gaia* DR2 (see Section 4.2.2). All these stars are found at distance larger than ~ 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 (4.3). The mode of the posterior distribution in equation (4.4) can be determined by numerically finding the roots of the implicit equation (Bailer-Jones 2015):

$$\frac{d^3}{L} - 2d^2 + \frac{\varpi}{\sigma_\varpi^2} d - \frac{1}{\sigma_\varpi^2} = 0. \tag{11}$$

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):

$$x_i = \frac{d_{\rm Mo,i} - d_{\rm true,i}}{d_{\rm true,i}}, \tag{12}$$

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 ~ 352000 simulated stars. In Fig. 11 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).

Table 2: Observed properties for the 20 "clean" high velocity star candidates with a probability > 80% of being unbound from the Galaxy. Stars are sorted by decreasing $P_{\rm ub}$ (see Table 3).

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)
Galactic						
5932173855446728064	(244.1181, −54.44045)	0.454 ± 0.029	−2.676 ± 0.043	−4.991 ± 0.034	−614.286 ± 2.492	13.81
1383279090527227264	(240.33735, 41.16677)	0.118 ± 0.016	−25.759 ± 0.025	−9.745 ± 0.04	−180.902 ± 2.421	13.01
6456587609813249536	(317.36089, −57.9124)	0.099 ± 0.019	13.002 ± 0.029	−18.263 ± 0.03	−15.851 ± 2.833	13.01
5935868592404029184	(253.90291, −53.29868)	0.074 ± 0.021	5.47 ± 0.032	6.358 ± 0.026	308.412 ± 1.212	13.08
5831614858352694400	(247.45238, −59.96738)	−0.008 ± 0.025	4.405 ± 0.032	1.532 ± 0.03	258.295 ± 1.245	13.37
5239334504523094784	(158.89457, −65.46548)	0.038 ± 0.013	−6.77 ± 0.025	2.544 ± 0.022	22.464 ± 1.891	13.39
4395399303719163904	(258.75009, 8.73145)	0.073 ± 0.019	−9.911 ± 0.029	4.848 ± 0.029	24.364 ± 1.484	13.19
Extragalactic						
1396963577886583296	(237.73164, 44.4357)	−0.017 ± 0.014	−1.649 ± 0.023	−4.966 ± 0.029	−412.464 ± 1.002	13.24
5593107043671135744	(113.26944, −31.3792)	−0.1 ± 0.017	−1.582 ± 0.03	2.113 ± 0.028	104.437 ± 1.511	13.39
5546986344820400512	(125.63998, −32.62)	−0.08 ± 0.022	−1.986 ± 0.028	2.747 ± 0.035	79.255 ± 1.273	13.82
5257182876777912448	(144.73682, −60.53137)	−0.012 ± 0.017	−3.736 ± 0.029	3.444 ± 0.027	22.64 ± 1.723	13.49
4326973843264734208	(248.8923, −14.51844)	0.199 ± 0.031	−20.546 ± 0.05	−33.974 ± 0.033	−220.392 ± 2.052	13.5
5298599521278293504	(140.14259, −62.46243)	−0.053 ± 0.02	−2.373 ± 0.071	3.883 ± 0.055	54.363 ± 1.17	13.39
6700075834174889472	(304.32289, −32.41577)	0.054 ± 0.037	−7.243 ± 0.065	4.955 ± 0.047	22.491 ± 2.057	12.75
4073247619504712192	(280.26863, −26.28806)	0.05 ± 0.024	−3.596 ± 0.046	6.231 ± 0.039	−191.767 ± 2.735	13.58
6492391900301222656	(348.64665, −58.42957)	0.095 ± 0.018	7.502 ± 0.027	−15.822 ± 0.026	−149.856 ± 1.163	13.36
4596514892566325504	(268.57736, 29.12348)	0.064 ± 0.013	−1.086 ± 0.019	−10.512 ± 0.023	−112.792 ± 1.093	13.49
5830109386395388544	(249.9792, −61.90285)	−0.006 ± 0.019	−1.072 ± 0.027	3.932 ± 0.029	143.395 ± 0.633	13.14
1990547230937629696	(344.00637, 53.61551)	0.043 ± 0.017	−4.769 ± 0.028	−2.83 ± 0.027	−83.38 ± 1.158	13.31
5321157479786017280	(128.82063, −53.20458)	−0.023 ± 0.018	−2.518 ± 0.032	3.224 ± 0.034	81.295 ± 0.668	13.59

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 negative bias of $\sim 20\%$ for stars with large values of the *true* relative error $f_{\rm true}$.

The reason why we choose not to use distances from Bailer-Jones et al. (2018) is that the authors fit the values of the scale length L to a full three-dimensional model of the Galaxy[4]. Their values are therefore driven by nearby, bright disk stars, with $f \ll 1$. Such an approach would underestimate distances (and therefore total velocities) to faint distant stars, the ones we are more interested in.

.2 Content of the Distance and Velocity Catalogue

Table 4 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`.

.3 List of High Velocity Stars with $0.5 < P_{\rm ub} \leqslant 0.8$.

In Table 5 we present *Gaia* identifiers, distances, and total velocities for the 105 high velocity stars discussed in Section 4.4, with $0.5 < P_{\rm ub} \leqslant 0.8$.

.4 Global Parallax Offset

In this appendix we discuss the impact of including the -0.029 mas global parallax zeropoint mentioned in Lindegren et al. (2018b), derived from *Gaia*'s observations of distant quasars. Being a negative offset, the net effect is to lower the inferred distances, and therefore the resulting total velocities. We repeat the Bayesian analysis discussed in Section 4.2 to the 20 stars with $P_{\rm ub} > 80\%$. In this case, the likelihood probability is again a multivariate Gaussian distribution, but with mean vector (Bailer-Jones et al. 2018):

$$\mathbf{m} = [\mu_{\alpha*}, \mu_\delta, 1/d + \varpi_{\rm zp}], \tag{13}$$

where $\varpi_{\rm zp} = -0.029$ mas. In Table 6 we report the updated values of the distance, total velocity, and probability of being unbound from the Galaxy

[4]Note that Bailer-Jones et al. (2018) adopt a scale length that varies smoothly with Galactic longitude and latitude.

Table 3: Derived properties for the 20 "clean" high velocity star candidates with a probability $> 80\%$ of being unbound from the Galaxy. Stars are sorted by decreasing $P_{\rm ub}$.

Gaia DR2 ID	d (pc)	$r_{\rm GC}$ (pc)	$v_{\rm GC}$ (km s^{-1})	$P_{\rm MW}$	$P_{\rm ub}$
Galactic					
5932173855446728064	2197^{+162}_{-120}	6397^{+92}_{-123}	747^{+2}_{-3}	1.00	1.00
1383279090527227264	8491^{+1376}_{-951}	10064^{+908}_{-561}	921^{+179}_{-124}	1.00	1.00
6456587609813249536	10021^{+2023}_{-1480}	7222^{+1350}_{-761}	875^{+212}_{-155}	0.98	0.99
5935868592404029184	12150^{+2919}_{-1909}	5985^{+2516}_{-1380}	747^{+110}_{-73}	0.83	0.98
5831614858352694400	20196^{+6006}_{-4394}	14113^{+5781}_{-4061}	664^{+130}_{-93}	0.94	0.92
5239334504523094784	19353^{+4247}_{-2940}	18351^{+3923}_{-2617}	609^{+140}_{-94}	0.77	0.88
4395399303719163904	12848^{+2766}_{-2262}	8194^{+2309}_{-1620}	671^{+136}_{-106}	1.00	0.84
Extragalactic					
1396963577886583296	31374^{+6332}_{-5185}	30720^{+6150}_{-4970}	693^{+145}_{-113}	0.00	0.98
5593107043671135744	37681^{+8295}_{-6444}	41753^{+8183}_{-6322}	567^{+100}_{-76}	0.00	0.97
5546986344820400512	29062^{+5928}_{-4950}	32552^{+5782}_{-4781}	551^{+90}_{-75}	0.00	0.93
5257182876777912448	26140^{+6400}_{-4240}	25824^{+6144}_{-3989}	605^{+148}_{-93}	0.03	0.92
4326973843264734208	5257^{+881}_{-677}	3842^{+450}_{-465}	766^{+163}_{-122}	0.04	0.91
5298599521278293504	28525^{+6774}_{-5110}	28145^{+6545}_{-4850}	579^{+139}_{-104}	0.03	0.88
6700075834174889472	13068^{+3816}_{-3123}	7584^{+3330}_{-2219}	698^{+152}_{-120}	0.10	0.84
4073247619504712192	14653^{+4331}_{-2807}	6884^{+4240}_{-2648}	695^{+139}_{-88}	0.11	0.84
6492391900301222656	10276^{+1878}_{-1541}	9641^{+1335}_{-944}	658^{+149}_{-117}	0.06	0.84
4596514892566325504	14255^{+2485}_{-1839}	12120^{+2106}_{-1453}	617^{+121}_{-90}	0.07	0.84
5830109386395388544	23852^{+6287}_{-4917}	17735^{+6123}_{-4680}	600^{+118}_{-88}	0.08	0.84
1990547230937629696	17543^{+4372}_{-3415}	21331^{+4114}_{-3130}	563^{+112}_{-84}	0.05	0.83
5321157479786017280	27523^{+6086}_{-5176}	28715^{+5877}_{-4914}	545^{+110}_{-95}	0.08	0.83

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

Table 4: 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]

Table 5: Distances and total velocities in the Galactic rest frame for the 105 "clean" high velocity star candidates with $0.5 < P_{\rm ub} \leqslant 0.8$. Sources are sorted by decreasing $P_{\rm ub}$.

Gaia DR2 ID	d (pc)	$v_{\rm GC}$ (km s^{-1})	$P_{\rm ub}$
5718618735518384768	31308^{+6464}_{-6622}	488^{+69}_{-73}	0.79
4532372476587492608	14132^{+3798}_{-2288}	606^{+142}_{-83}	0.78
4366218814874247424	7506^{+1521}_{-955}	678^{+137}_{-86}	0.78
5244448023850619648	16553^{+3638}_{-2756}	552^{+91}_{-66}	0.77
1994938164981988864	22185^{+5526}_{-4751}	516^{+85}_{-70}	0.77
2159020415489897088	7686^{+1651}_{-1293}	603^{+123}_{-97}	0.77
2112308930997657728	6114^{+999}_{-712}	619^{+119}_{-84}	0.77
5802638672467252736	9985^{+1804}_{-1322}	647^{+150}_{-108}	0.76
5996908319666721792	13616^{+3593}_{-2595}	662^{+151}_{-108}	0.75
5316722526615701504	24242^{+6103}_{-4691}	525^{+123}_{-89}	0.74
2095259117723646208	13359^{+2970}_{-2614}	594^{+134}_{-112}	0.73
5839686407534279808	7346^{+1033}_{-839}	633^{+112}_{-92}	0.72
1333199496978208128	20038^{+4062}_{-3076}	543^{+120}_{-86}	0.72
2089995308886282880	13397^{+2700}_{-1874}	573^{+121}_{-81}	0.71
2045752026157687040	11799^{+2705}_{-2004}	604^{+144}_{-106}	0.71
6431596947468407552	11356^{+2099}_{-1531}	590^{+66}_{-47}	0.71
5247579810921207680	27357^{+5878}_{-4547}	499^{+115}_{-85}	0.7
5298494930231856512	23913^{+5493}_{-4057}	510^{+119}_{-85}	0.7
2095397827987170816	14751^{+2839}_{-2301}	574^{+122}_{-98}	0.7
4656931544705794816	24368^{+5597}_{-4637}	514^{+118}_{-95}	0.7
6642234513167197824	6836^{+1252}_{-1037}	649^{+117}_{-91}	0.69
5399966178291369728	10155^{+2090}_{-1430}	566^{+121}_{-81}	0.69
5374177064347894272	6225^{+1109}_{-879}	587^{+97}_{-76}	0.68
2072048770884296704	16139^{+3291}_{-2678}	552^{+118}_{-94}	0.68
6116555426949827200	7741^{+1164}_{-1011}	628^{+118}_{-102}	0.67
6500989806352727936	10407^{+2456}_{-1809}	577^{+128}_{-90}	0.67
5217818333256869376	8642^{+1631}_{-1139}	585^{+118}_{-81}	0.67
2106519830479009920	8213^{+1326}_{-1065}	570^{+85}_{-67}	0.67
6397497209236655872	5802^{+643}_{-487}	587^{+54}_{-41}	0.66

Table 5: - *continued.*

Gaia DR2 ID	d (pc)	$v_{\rm GC}$ (km s^{-1})	$P_{\rm ub}$
2044224735768501760	15167^{+3227}_{-2538}	560^{+124}_{-96}	0.66
5303927273594669056	20331^{+5200}_{-3372}	508^{+118}_{-73}	0.66
1966103266381646720	28232^{+6210}_{-5780}	474^{+88}_{-76}	0.65
6241406793347941504	14098^{+4035}_{-3000}	609^{+139}_{-98}	0.65
5627896072604568960	22754^{+5478}_{-4591}	490^{+101}_{-83}	0.65
5415267600583814912	24505^{+6046}_{-4520}	498^{+115}_{-87}	0.65
5856098302217892352	19735^{+4562}_{-3404}	529^{+127}_{-93}	0.65
6444276683058885248	11413^{+3064}_{-2202}	617^{+147}_{-103}	0.65
2094386346009409280	14643^{+2968}_{-2007}	549^{+125}_{-82}	0.64
5309766504975294592	25956^{+5528}_{-5114}	490^{+106}_{-96}	0.64
3905884598043829504	2709^{+385}_{-289}	580^{+115}_{-86}	0.63
2038012426369296128	16453^{+4086}_{-3062}	543^{+127}_{-88}	0.63
5317203154946837760	18068^{+3537}_{-3079}	510^{+94}_{-80}	0.63
5897201311028035456	17717^{+4423}_{-4116}	543^{+83}_{-70}	0.62
5823425661366917376	15652^{+4759}_{-3695}	568^{+127}_{-97}	0.62
5807202126764572288	14365^{+3602}_{-2776}	563^{+97}_{-74}	0.62
3705761936916676864	3756^{+371}_{-300}	566^{+59}_{-46}	0.62
2183775885439262592	23213^{+5580}_{-4338}	480^{+102}_{-78}	0.62
5317776481532378240	19139^{+4400}_{-3115}	500^{+112}_{-79}	0.62
6077622510498751616	14503^{+3852}_{-2502}	538^{+84}_{-46}	0.62
4531575708618805376	12030^{+2748}_{-1974}	562^{+80}_{-56}	0.62
1956680279930601344	23550^{+6723}_{-4451}	480^{+113}_{-75}	0.62
6010197124582216832	10863^{+3441}_{-1945}	629^{+118}_{-65}	0.62
5232568213032618496	27921^{+5690}_{-4842}	487^{+111}_{-92}	0.61
5249820306388948992	26092^{+6478}_{-4213}	478^{+117}_{-78}	0.61
5779439836114210304	23901^{+5743}_{-4509}	492^{+69}_{-53}	0.6
5247264629041172608	20274^{+3940}_{-3336}	507^{+100}_{-80}	0.6
5912922197004254848	12401^{+3128}_{-2696}	610^{+122}_{-99}	0.6
5247811567357582336	21321^{+4641}_{-3453}	497^{+114}_{-86}	0.59
4489509905555953408	11610^{+2734}_{-2257}	590^{+117}_{-91}	0.59
2121857472227927168	13251^{+2401}_{-1679}	522^{+92}_{-63}	0.59
1989862986804105344	10429^{+2057}_{-1607}	523^{+107}_{-82}	0.58

Table 5: - *continued.*

Gaia DR2 ID	d (pc)	v_{GC} (km s^{-1})	P_{ub}
6677910160794903296	4345^{+554}_{-396}	604^{+106}_{-76}	0.58
6229070238523155328	13987^{+4361}_{-2810}	567^{+142}_{-87}	0.58
4452929978332889216	24168^{+5324}_{-4537}	496^{+108}_{-88}	0.58
5785402796909679744	14723^{+3134}_{-2187}	543^{+132}_{-89}	0.58
5362114562797004544	23461^{+5342}_{-4015}	479^{+113}_{-80}	0.57
1331585993728475264	10902^{+2413}_{-1920}	544^{+115}_{-87}	0.57
6733156428223193856	13978^{+3684}_{-2829}	601^{+122}_{-92}	0.57
6221350429945324032	8878^{+2117}_{-1582}	593^{+141}_{-104}	0.57
3454083549225619712	5943^{+794}_{-627}	522^{+100}_{-77}	0.57
6868478546915992320	14043^{+4460}_{-3582}	576^{+130}_{-101}	0.57
4127621699294858368	13174^{+3602}_{-2904}	615^{+128}_{-98}	0.56
1364548016594914560	10327^{+1989}_{-1642}	531^{+66}_{-50}	0.56
4609875745549298688	10640^{+1380}_{-1204}	544^{+76}_{-66}	0.56
5212817273334550016	3811^{+330}_{-283}	565^{+59}_{-51}	0.56
1268023196461923712	4586^{+500}_{-390}	568^{+79}_{-61}	0.56
1696697285206197248	23235^{+5014}_{-3909}	464^{+111}_{-81}	0.56
6034352158118691072	11013^{+2964}_{-2267}	646^{+157}_{-104}	0.56
2098831980759357696	15685^{+3439}_{-2694}	518^{+119}_{-92}	0.56
5354094037807264384	11683^{+2120}_{-1758}	533^{+111}_{-90}	0.56
4220617568115374848	4978^{+814}_{-677}	603^{+114}_{-92}	0.56
5779919841659989120	10641^{+2101}_{-1505}	568^{+135}_{-95}	0.55
5317675979297751040	27098^{+5311}_{-4561}	451^{+81}_{-70}	0.55
3891412241883772928	7004^{+1531}_{-1150}	539^{+88}_{-65}	0.55
4916199478888664320	5579^{+725}_{-629}	549^{+66}_{-56}	0.55
2255126837089768192	24623^{+4714}_{-4286}	456^{+85}_{-74}	0.55
5511130239834500864	20579^{+5603}_{-3668}	467^{+100}_{-68}	0.55
3784964943489710592	4031^{+733}_{-505}	552^{+92}_{-61}	0.55
2038818952503671424	26358^{+5535}_{-5090}	469^{+105}_{-92}	0.55
1954400884950622464	19455^{+4960}_{-3498}	482^{+108}_{-75}	0.54
5846560382443820032	7054^{+936}_{-629}	585^{+96}_{-64}	0.54
6130863887159694848	9639^{+2070}_{-1335}	550^{+133}_{-85}	0.54
5231000034569444992	18206^{+3340}_{-3437}	501^{+101}_{-105}	0.53

Table 5: - *continued.*

Gaia DR2 ID	d (pc)	$v_{\rm GC}$ (km s^{-1})	$P_{\rm ub}$
2186887606421426816	24376^{+4607}_{-4292}	454^{+74}_{-67}	0.53
5818738237122521344	11884^{+3059}_{-2216}	559^{+136}_{-89}	0.53
5249917441371959040	17540^{+4063}_{-3149}	494^{+116}_{-85}	0.53
6639557580310606976	11135^{+3975}_{-2226}	579^{+108}_{-55}	0.53
4210389120686616832	7886^{+2550}_{-1822}	599^{+143}_{-88}	0.52
1191989287342960640	10798^{+2233}_{-1691}	549^{+131}_{-96}	0.52
6098331056080412416	16089^{+3894}_{-3358}	528^{+89}_{-72}	0.52
2086507417487662976	26304^{+5278}_{-4208}	448^{+90}_{-72}	0.51
5303240216263896192	21972^{+5482}_{-3995}	464^{+111}_{-79}	0.51
2000253135474943616	16537^{+3984}_{-3129}	475^{+89}_{-69}	0.51
6035120957243593600	10873^{+3525}_{-2307}	603^{+124}_{-76}	0.51
1612628419987892096	25402^{+5063}_{-3992}	442^{+104}_{-79}	0.5

for the 20 stars discussed in Section 4.5. We now find 14 candidates (70%) to have an updated $P_{\rm ub} > 50\%$, and 4 stars (20%) to have $P_{\rm ub} > 80\%$.

.5 Systematic Errors in Parallax

Gaia DR2 uncertainties in parallax do not include the contribution from *systematic* errors, which might depend on the magnitude, position, colour, and other property of the source. The mean value of the systematic errors is the global parallax offset $\varpi_{\rm zp}$ already discussed in Appendix .4. In this appendix we discuss the impact of adding this contribution to the quoted values of the parallax uncertainties. To do that, we follow the advice and guidelines presented in Lindegren et al. (2018a). Internal uncertainties published in the *Gaia* DR2 catalogue can be artificially inflated to keep into account systematic errors (e.g. Lindegren et al. 2016):

$$\sigma_{\varpi,\rm ext} = \sqrt{k^2\sigma_\varpi + \sigma_s^2}, \tag{14}$$

where $k \gtrsim 1$ is a correction factor, and σ_s is the variance of the systematic error. These parameters need to be calibrated using external datasets. Lindegren et al. (2018a) suggest adopting $k = 1.08$, $\sigma_s = 0.021$ mas ($k = 1.08$, $\sigma_s = 0.043$ mas) for bright stars with $G \lesssim 13$ (faint stars with $G \gtrsim 13$).

Table 6: Distances and total velocities in the Galactic rest frame for the 20 "clean" high velocity star candidates with $P_{\rm ub} > 0.8$ presented in Table 2, including the -0.029 mas global parallax offset. For comparison, stars are sorted as in Table 2.

Gaia DR2 ID	d (pc)	$v_{\rm GC}$ (km s^{-1})	$P_{\rm ub}$
5932173855446728064	2096^{+130}_{-117}	747^{+3}_{-3}	1.0
1383279090527227264	7144^{+809}_{-782}	745^{+105}_{-102}	0.98
6456587609813249536	7964^{+1297}_{-885}	660^{+135}_{-92}	0.82
5935868592404029184	10010^{+2144}_{-1800}	665^{+81}_{-67}	0.75
5831614858352694400	17160^{+4736}_{-4055}	600^{+101}_{-86}	0.73
5239334504523094784	14426^{+3339}_{-2236}	454^{+105}_{-66}	0.32
4395399303719163904	9934^{+2389}_{-1586}	535^{+112}_{-71}	0.37
1396963577886583296	23038^{+5341}_{-3347}	511^{+112}_{-68}	0.73
5593107043671135744	32604^{+6740}_{-4982}	511^{+79}_{-61}	0.9
5546986344820400512	26048^{+6507}_{-4962}	507^{+99}_{-74}	0.78
5257182876777912448	21973^{+4863}_{-4292}	515^{+106}_{-91}	0.66
4326973843264734208	4718^{+725}_{-580}	670^{+131}_{-104}	0.72
5298599521278293504	24102^{+6820}_{-3800}	489^{+140}_{-74}	0.63
6700075834174889472	11382^{+4021}_{-2622}	631^{+158}_{-98}	0.69
4073247619504712192	11656^{+3234}_{-1949}	601^{+101}_{-61}	0.47
6492391900301222656	7999^{+1457}_{-1042}	487^{+109}_{-73}	0.29
4596514892566325504	10522^{+1717}_{-1145}	436^{+83}_{-53}	0.14
5830109386395388544	19057^{+4550}_{-3307}	514^{+84}_{-59}	0.51
1990547230937629696	13243^{+2851}_{-2563}	456^{+71}_{-62}	0.37
5321157479786017280	22613^{+5272}_{-4543}	456^{+94}_{-81}	0.51

In Table 7 we report the updated values for distances, total velocities, and probability of being unbound from the Galaxy for the sample of 20 stars discusses in Section 4.5. All of the stars but one are classified as *faint* stars. 9 (5) stars out of 20 now have an updated probability $P_{\rm ub} > 0.5$ ($P_{\rm ub} > 0.8$). We want to stress that the adopted value for σ_s is likely overestimated for the typical magnitude of stars in our sample (Lindegren et al. 2018a), therefore this is a conservative approach, which underestimates distances (and therefore total velocities).

Table 7: Distances and total velocities in the Galactic rest frame for the 20 "clean" high velocity star candidates with $P_{\rm ub} > 0.8$ presented in Table 2. Parallax uncertainties are inflated according to equation (14). For comparison, stars are sorted as in Table 2.

Gaia DR2 ID	d (pc)	$v_{\rm GC}$ (km s^{-1})	$P_{\rm ub}$
5932173855446728064	2316^{+306}_{-265}	746^{+3}_{-3}	1.0
1383279090527227264	8577^{+3716}_{-2135}	931^{+484}_{-278}	0.94
6456587609813249536	9370^{+3917}_{-2262}	806^{+414}_{-234}	0.86
5935868592404029184	10744^{+3489}_{-2685}	694^{+131}_{-101}	0.8
5831614858352694400	13924^{+5147}_{-3860}	531^{+109}_{-79}	0.53
5239334504523094784	12051^{+4102}_{-3115}	384^{+124}_{-84}	0.22
4395399303719163904	11019^{+3704}_{-3061}	585^{+179}_{-138}	0.58
1396963577886583296	15707^{+5086}_{-3944}	372^{+92}_{-54}	0.21
5593107043671135744	18643^{+5317}_{-4575}	348^{+62}_{-53}	0.14
5546986344820400512	16803^{+5056}_{-4307}	371^{+77}_{-61}	0.2
5257182876777912448	14545^{+4481}_{-3702}	361^{+91}_{-65}	0.17
4326973843264734208	6032^{+2296}_{-1452}	909^{+426}_{-265}	0.91
5298599521278293504	16316^{+5884}_{-4573}	341^{+109}_{-77}	0.18
6700075834174889472	12278^{+4717}_{-3393}	667^{+187}_{-129}	0.75
4073247619504712192	11462^{+4236}_{-2678}	593^{+135}_{-83}	0.48
6492391900301222656	9897^{+4543}_{-2488}	630^{+359}_{-185}	0.69
4596514892566325504	11421^{+4949}_{-2998}	479^{+242}_{-141}	0.44
5830109386395388544	14312^{+5392}_{-4022}	430^{+97}_{-69}	0.27
1990547230937629696	11614^{+4698}_{-2769}	416^{+116}_{-66}	0.35
5321157479786017280	15167^{+5524}_{-3904}	328^{+95}_{-62}	0.13

5 Joint constraints on the Galactic dark matter halo and Galactic Centre from hypervelocity stars

E.M. Rossi, **T. Marchetti**, *M. Cacciato, M. Kuiack, R. Sari 2017, MNRAS, 467, 1844-1856*

The mass assembly history of the Milky Way can inform both theory of galaxy formation and the underlying cosmological model. Thus, observational constraints on the properties of both its baryonic and dark matter contents are sought. Here, we show that hypervelocity stars (HVSs) can in principle provide such constraints. We model the observed velocity distribution of HVSs, produced by tidal break-up of stellar binaries caused by Sgr A*. Considering a Galactic Centre (GC) binary population consistent with that inferred in more observationally accessible regions, a fit to current HVS data with significance level > 5 per cent can only be obtained if the escape velocity from the GC to 50 kpc is $V_G \lesssim 850$ km s^{-1}, regardless of the enclosed mass distribution. When a Navarro, Frenk and White matter density profile for the dark matter halo is assumed, haloes with $V_G \lesssim 850$ km s^{-1}are in agreement with predictions in the cold dark matter model and a subset of models around $M_{200} \sim 0.5$-1.5×10^{12} M$_\odot$ and $r_s \lesssim 35$ kpc can also reproduce Galactic circular velocity data. HVS data alone cannot currently exclude potentials with $V_G > 850$ km s^{-1}. Finally, specific constraints on the halo mass from HVS data are highly dependent on the assumed baryonic mass potentials. This first attempt to simultaneously constrain GC and dark halo properties is primarily hampered by the paucity and quality of data. It nevertheless demonstrates the potential of our method, that may be fully realized with the ESA *Gaia* mission.

5.1 Introduction

The visible part of galaxies is concentrated in the centre of more extended and more massive dark matter structures, that are termed haloes. In our Galaxy, the baryonic matter makes up a few percent of the total mass, and the halo is ~ 10 times more extended than the Galactic disc. In the current paradigm, galaxies assemble in a hierarchical fashion from smaller structures and the result is due to a combination of merger history, the underlying cosmological model and baryonic physics (e.g. cooling and star formation). Thanks to our vantage point, these fundamental ingredients in galaxy assembly, can be uniquely constrained by observations of the matter content of the Milky Way and its distribution, when analysed in synergy with dedicated cosmological simulations.

Currently, our knowledge of the Galactic dark matter halo is fragmented. Beyond ~ 10 kpc dynamical tracers such as halo field stars and stellar streams become rarer and rarer and astrometric errors significant. In particular, there is a large uncertainty in the matter density profile, global shape, orientation coarseness (e.g. Bullock et al. 2010; Law & Majewski 2010; Vera-Ciro & Helmi 2013; Loebman et al. 2014; Laevens et al. 2015; Williams & Evans 2015) and current estimates of the halo mass differ by approximately a factor of 3 (see fig.1 in Wang et al. 2015, and references therein). This difference is significant as a mass measurement in the upper part of that range together with observations of Milky Way satellites can challenge (Klypin et al. 1999; Moore et al. 1999; Boylan-Kolchin et al. 2011) the current concordance cosmological paradigm: the so-called Λ cold dark matter model (ΛCDM). In particular, the "too big to fail problem" (Boylan-Kolchin et al. 2011) states that, in ΛCDM high mass ($\gtrsim 2\times10^{12} M_{\odot}$) haloes, the most massive subhaloes are too dense to correspond to any of the known satellites of the Milky Way. Therefore, the solution may simply be a lighter Galactic halo of $< 10^{12} M_{\odot}$ (e.g. Vera-Ciro et al. 2013; Gibbons et al. 2014). This is an example of how a robust measurement of the Galactic mass can be instrumental to test cosmological models.

On the other extreme of Galactic scales, the Galactic Centre (GC) has been the focus of intense research since the beginning of the 1990s, and it is regarded as a unique laboratory to understand the interplay between (quiescent) supermassive black holes (SMBHs) and their environment (see Genzel et al. 2010, for a review). Indeed, the GC harbours the best observationally constrained SMBH, called Sgr A*, of mass $\approx 4.0\times10^{6} M_{\odot}$ (Ghez et al. 2008; Gillessen et al. 2009; Meyer et al. 2012). In particular, GC observa-

tions raise issues on the stellar mass assembly, which is intimately related to the SMBH growth history. For example, in the central $r \sim 0.5$ pc the light is dominated by young (~ 6 Myr old) stars (e.g. Paumard et al. 2006; Lu et al. 2013) with a suggested top-heavy initial mass function (IMF Bartko et al. 2010; Lu et al. 2013) and a large spread in metallicity at $r < 1$ pc (Do et al. 2015). The existence of young stars well within the gravitational sphere of influence of Sgr A* challenges our knowledge of how stars form, as molecular clouds should not survive tidal forces there. These stars are part of a larger scale structure called nuclear star cluster with half-light radius around ~ 5 pc (e.g. Schödel et al. 2014b; Fritz et al. 2016): in contrast with the inner region, its IMF may be consistent with a Chabrier/Kroupa IMF and between 2.5 pc $< r < 4$ pc the majority of stars appear to be older than 5 Gyr (e.g. Pfuhl et al. 2011; Fritz et al. 2016). The origin of this nuclear star cluster and its above mentioned features is highly debated, and the leading models consider coalescence of stellar clusters that reach the GC and are tidally disrupted or in situ formation from gas streams (see Böker 2010, for a review on nuclear star cluster). The Hubble Space Telescope imaging surveys have shown that most galaxies contain nuclear clusters in their photometric and dynamical centres (e.g. Carollo et al. 1997; Georgiev & Böker 2014; Carson et al. 2015), but the more observationally accessible and best studied one is the Milky Way's, which once more give us a chance of understanding the formation of galactic nuclei in general. However, to investigate the GC via direct observations, one must cope with observational challenges such as the strong and spatially highly variable interstellar extinction and stellar crowding. A concise review of the current knowledge of the nuclear star cluster at the GC and the observational obstacles and limitations is given in Schödel et al. (2014a).

Remarkably, a single class of objects can potentially address the mass content issue from the GC to the halo: hypervelocity stars (HVSs). These are detected in the outer halo (but note Zheng et al. 2014) with radial velocities exceeding the Galactic escape speed (Brown et al. 2005; see Brown 2015, for a review). So far around 20 HVSs have been discovered with velocities in the range $\sim 300 - 700$ km s^{-1}, and trajectories consistent with coming from the GC. Because of the discovery strategy, they are all B-type stars mostly in the masses range between $2.5-4 M_\odot$ (e.g. Brown et al. 2014). Studying HVSs is thus a complementary way to investigate the GC stellar population, by surveying more accessible parts of the sky. After ejection, HVS dynamics is set by the Galactic gravitational field. Therefore, regardless of their origin, HVS spatial and velocity distributions can in principle

probe the Galactic total matter distribution (Gnedin et al. 2005, 2010; Yu & Madau 2007; Sesana et al. 2007; Perets et al. 2009; Fragione & Loeb 2017).

Retaining hundreds of km s^{-1}in the halo while originating from a deep potential well requires initial velocities in excess of several hundreds of km s^{-1}Kenyon et al. (2008), which are very rarely attained by stellar interaction mechanisms put forward to explain runaway stars (e.g. Blaauw 1961; Aarseth 1974; Eldridge et al. 2011; Perets & Šubr 2012; Tauris 2015; Rimoldi et al. 2016). Velocity and spatial distributions of runaway and HVSs are indeed expected to be different (Kenyon et al. 2014). For example, high velocity runaway stars would almost exclusively come from the Galactic disc (Bromley et al. 2009). Instead, HVS energetics and trajectories strongly support the view that HVSs were ejected in gravitational interactions that tap the gravitational potential of Sgr A*, and, as a consequence of a huge "kick", escaped into the halo. In particular, most observations are consistent with the so called "Hills' mechanism", where a stellar binary is tidally disrupted by Sgr A*. As a consequence, a star can be ejected with a velocity up to thousands km s^{-1}(Hills 1988). Another appealing feature is that the observed B-type stellar population in the inner parsec — whose in situ origin is quite unlikely — is consistent with being HVSs' companions, left bound to Sgr A* by the Hills' mechanism (Zhang et al. 2013; Madigan et al. 2014).

In a series of three papers, we have built up a solid and efficient semi-analytical method that fully reproduces 3-body simulation results for mass ratios between a binary star and a SMBH ($m_t/M \sim 10^{-6}$) expected in the GC. In particular we reproduce star trajectories, energies after the encounter and ejection velocity distributions (see Sari et al. 2010; Kobayashi et al. 2012; Rossi et al. 2014, and section 5.2 in this paper). Here, we will capitalise on that work and apply our method to the modelling of current HVS data, with the primary aim of constraining the Galactic dark matter halo and simultaneously derive consequences for the binary population in the GC. Since star binarity is observed to be very frequent in the Galaxy (around 50%) and the GC seems no exception ($\sim 30\%$ for massive binaries Pfuhl et al. 2014), clues from HVS modelling are a complementary way to understand the stellar population within the inner few parsecs from Sgr A*.

This paper is organised as follows. In Section 5.2, we describe our method to build HVS ejection velocity distributions, based on our previous work on the Hills' mechanism. In Section 5.3 , we present our first approach to predict velocity distributions in the outer Galactic halo and we show our results when comparing them to data in Section 5.3.3. In Section 5.4, we

will specialise to a "Navarro, Frenk and White" (NFW) dark matter profile and present results in Section 5.4.2. In Section 5.5, we discuss our findings, their limitations and implications and then conclude. Finally, in Appendix .1, we describe our analysis of the Galactic circular velocity data, that we combine with HVS constraints.

5.2 Ejection velocity distributions

We here present our calculation of the *ejection* velocity distribution of hypervelocity stars (i.e. the velocity distribution at infinity with respect to the SMBH) via the Hills' mechanism. We denote with M Sgr A*'s mass, fixed to $M = 4.0 \times 10^6 M_\odot$.

Let us consider a stellar binary system with separation a, primary mass m_p, secondary mass m_s, mass ratio $q = m_\mathrm{s}/m_\mathrm{p} \leqslant 1$, total mass $m_\mathrm{s} + m_\mathrm{p} = m_\mathrm{t}$ and period P. If this binary is scattered into the tidal sphere of Sgr A*, the expectation is that its centre of mass is on a nearly parabolic orbit, as its most likely place of origin is the neighbourhood of Sgr A*'s radius of influence. Indeed, this latter is ~ 5 orders of magnitude larger than the tidal radius, and therefore the binary's orbit must be almost radial to hit the tiny Sgr A*'s tidal sphere. On this orbit, the binary star has[1] $\sim 90\%$ probability to undertake an exchange reaction, where a star remains in a binary with the black hole, while the companion is ejected. In addition, we proved that the ejection probability is independent of the stellar mass, when the centre of mass of the binary is on a parabolic orbit. This is different from the case of elliptical or hyperbolic orbits where the primary star, carrying most of the orbital energy, has a greater chance to be respectively captured or ejected (Kobayashi et al. 2012).

The ejected star has a velocity at infinity, *in solely presence of the black hole potential*, equal to

$$v_\mathrm{ej} = \sqrt{\frac{2\,G m_\mathrm{c}}{a}} \left(\frac{M}{m_\mathrm{t}}\right)^{1/6}, \tag{5.1}$$

(Sari et al. 2010) where m_c is the mass of the binary companion star to the HVS and G is the gravitational constant. Rigorously, there is a numerical factor in front of the square root in (eq. 5.1) that depends on the binary-black hole encounter geometry. However, this factor is ~ 1, when averaged

[1] In Sari et al. (2010), we show that a binary star on a parabolic orbit has 80% chance of disruption, when considering prograde and retrograde orbits. Our (unpublished) calculations averaged over all orbital inclinations indicate a high percentage around $\sim 90\%$.

over the binary's phase[2]. Moreover, the velocity distributions obtained with the full numerical integration of a binary's trajectory and those obtained with (eq. 5.1) are almost indistinguishable (Rossi et al. 2014). Given these results and the simplicity of eq. 5.1, it is possible to predict *ejection* velocity distributions, efficiently exploring a large range of the parameter space in Galactic potentials, binary separations and stellar masses. This latter is the main advantage over methods using 3-body (or N-body) simulations.

Since we are only considering binaries with primaries' mass $\gtrsim 3M_{\odot}$, we may consider observations of B-type and O-type binary stars for guidance. Because of the large distance and the extreme optical extinction, observations and studies of binaries in the inner GC are limited to a handful of very massive early-type binary stars (e.g. Ott et al. 1999; Pfuhl et al. 2014) and X-ray binaries (e.g. Muno et al. 2005).

For more reliable statistical inferences, we should turn to observations of more accessible regions in the Galaxy and in the Large Magellanic Cloud (LMC). They suggest that a power-law description of these distributions is reasonable. In the Solar neighbourhood, spectroscopic binaries with primary masses between $1-5M_{\odot}$ have a separation distribution, $f_{\rm a}$, that for short periods can be both approximated by a $f_{\rm a} \propto a^{-1}$ (Öpik's law, i.e. $f(\log_{10} P) \propto (\log_{10} P)^{\eta}$, with $\eta = 0$) and a log normal distribution in period with $\langle P \rangle \simeq 10$ day and a $\sigma_{\rm logP} \simeq 2.3$ (Kouwenhoven et al. 2007; Duchêne & Kraus 2013). However, in the small separation regime, relevant for the production of HVSs, the log normal distribution may also be described by a power-law[3]: $f_{\rm a} \propto a^{0.8}$. For primary masses $> 16M_{\odot}$, Sana et al. (2012) find a relatively higher frequency of short-period binaries in Galactic young clusters, $\eta \approx -0.55$, but a combination of a pick at the smallest periods and a power-law may be necessary to encompass all available observations (see e.g. Duchêne & Kraus 2013). For this range of massive stars ($\sim 20M_{\odot}$), a similar power-law distribution $\eta \approx -0.45$ is also consistent with a statistical description of O-type binaries in the VLT-FLAMES Tarantula Survey of the star forming region 30 Doradus of the LMC (Sana et al. 2013). In the same region, a similar analysis for observed early ($\sim 10M_{\odot}$) B-type binaries recovers instead an Öpik's law (Dunstall et al. 2015).

Mass ratio distributions, $f_{\rm q}$, for Galactic binaries are generally observed

[2]The binary's phase is the angle between the stars' separation and their centre of mass radial distance from Sgr A*, measured, for instance, at the tidal radius or at pericentre.

[3]This fit value does not significantly depends on the total mass assumed for binaries. We do not calculate errors on this fitted index, because our aim is to draw in the $\gamma-\alpha$ parameter space an indicative range of power-law exponents for the separation distribution of B-type binaries in the Solar Neighbourhood (see Figure 5.2).

to be rather flat, regardless of the primary's mass range (e.g. Sana et al. 2012; Kobulnicky et al. 2014; Duchêne & Kraus 2013, see their table 1). Differently, in the 30 Doradus star forming region, the mass ratio distributions appear to be steeper, ($f_{\rm q} \propto q^{\sim(-1)}$ in O-type banaries and $f_{\rm q} \propto q^{\sim(-3)}$ in early B-type ones), suggesting a preference for pairing with lower-mass companions: still a power-law may be fitted to data (Sana et al. 2013; Dunstall et al. 2015).

We therefore assume a binary separation distribution

$$f_{\rm a} \propto a^{\alpha}, \tag{5.2}$$

where the minimum separation is taken to be the Roche-Lobe radius $a_{\rm min} = 2.5 \times \max[R_*, R_{\rm c}]$, where R_* and $R_{\rm c}$ are the HVS's and the companion's radii, respectively. As a binary mass ratio distribution, we assume

$$f_{\rm q} \propto q^{\gamma}, \tag{5.3}$$

for $m_{\rm min} \leqslant m_{\rm s} \leqslant m_{\rm p}$. If not otherwise stated, $m_{\rm min} = 0.1M$.

The mass of the primary star ($m_{\rm p} \gtrsim 3M_{\odot}$) is taken from an initial mass function, that needs to mirror the star formation in the GC in the last $\sim 10^9$ yr. As mentioned in our introduction, the stellar mass function is rather uncertain and may be spatially dependent. Observations of stars with $M > 10M_{\odot}$ within about 0.5 pc from Sgr A* indicate a rather top-heavy mass function with $f_{\rm m} \propto m_{\rm p}^{-1.7}$ (Lu et al. 2013). At larger radii observations of red giants (and the lack of wealth of massive stars observed closer in) may instead point towards a more canonical bottom-heavy mass function (e.g. Pfuhl et al. 2011; Fritz et al. 2016). Given these uncertainties, we explore the consequences of assuming either a Kroupa mass function (Kroupa 2002), $f_{\rm m} \propto m_{\rm p}^{-2.3}$ or top-heavy distribution, $f_{\rm m} \propto m_{\rm p}^{-1.7}$, in the mass range $2.5M_{\odot} \leqslant m_{\rm p} \leqslant 100M_{\odot}$.

Finally, we do not introduce here any specific model for the injection of binaries in the black hole tidal sphere and consequently, we do not explicitly consider any "filter" or modification to the binary "natal" distributions. Likewise, we do not explicitly account for higher order multiplicity (e.g. binary with a third companion, i.e. triples) that may result in disruption of binaries with different distributions than those cited above. On the other hand, a way to interpret our results is to consider that the separation and mass ratio distributions already contain those modifications. We will explore these possibilities in Section 5.5.

5.3 Predicting velocity distributions in the halo: first approach.

In this Section, we first describe how we compute *the halo* velocity distribution with a method that allows us to use a single parameter to describe the Galactic deceleration, without specifying its matter profile (Sec. 5.3.1) . Given the large Galactocentric distances at which the current sample of HVSs is observed, our method is shown to be able to reproduce the correct velocity distribution for the velocity range of interest, *without* the need to calculate the HVS deceleration along the star's entire path from the GC. These features allow us to efficiently explore a large range of the binary population and the dark matter halo parameter space. Then, in Sec. 5.3.2, we describe how we perform our comparison with current selected data and finally we present our results in Sec. 5.3.3.

5.3.1 Velocity distribution in the halo: global description of the potential

Our first approach follows Rossi et al. (2014) and consists in *not* assuming any specific model for the Galactic potential, but rather to globally describe it by the minimum velocity, $V_{\rm G}$, that an object must have at the GC in order to reach 50 kpc with a velocity equal or greater than zero. In other words, the parameter $V_{\rm G}$ is a measure of the net deceleration suffered by a star ejected at the GC into the outer halo, regardless of the mass distribution interior to it. The statement is that Galactic potentials with the same $V_{\rm G}$ produce the same velocity distribution beyond 50 kpc, which is where most HVSs are currently observed[4].

The physical argument that supports this statement is the following. For any reasonable distribution of mass that accounts for the presence of the observed bulge, most of the deceleration occurs well before stars reach the inner halo (e.g. Kenyon et al. 2008) and therefore, any potential with the same escape velocity $V_{\rm G}$ will have the same net effect on an initial ejection velocity:

$$v = \sqrt{v_{\rm ej}^2 - V_{\rm G}^2}. \tag{5.4}$$

Although practically we are interested in the HVS distribution beyond 50 kpc, the method outlined here is valid for any threshold distance as long as

[4]There is one discovered at $\sim$ 12 kpc (Zheng et al. 2014), but we will not include in our analysis because it has a different mass and location than our working sample, and therefore it would need a separate analysis.

the deceleration beyond that is negligible and, as justified below, all stars in the velocity range of interest reach it within their life-time. Therefore in the following, when a specific choice is not needed, we will generically call this threshold distance "$r_{\rm in}$". This, we recall, is also the radius associated to $V_{\rm G}$.

Let us now proceed to calculate the HVS velocity distribution within a given radial range $\Delta r = [r_{\rm out} - r_{\rm in}]$ in spherical symmetry, assuming a time-independent ejection rate $\mathcal{R}$ (typically $\sim 10 - 100$ Myr^{-1}). Given the above premises, HVSs with a velocity around v cross $r_{\rm in}$ at a rate $d\dot{N}/dv$, that can be obtained from the ejection-velocity probability density function (PDF) $P(v_{\rm ej})$ equating bins of corresponding velocity,

$$\frac{d\dot{N}}{dv} dv = \mathcal{R} P(v_{\rm ej}) {\rm d}v_{\rm ej},$$

with the aid of eq.5.4, that gives $v = v(v_{\rm ej})$. Consequently, the halo-velocity PDF (dn/dv) within a given radial range Δr can be simply computed as

$$dn(v, \Delta r) \propto \frac{d\dot{N}}{dv} \times \min[\Delta r/v, \langle t_{\rm life} \rangle]\, dv, \tag{5.5}$$

where $\min[\Delta r/v, \langle t_{\rm life} \rangle]$ is the average residence time in that range of Galactocentric distances of HVSs in a bin dv of velocity around v. This is the minimum between the crossing time $\Delta r/v$ and the average life-time $\langle t_{\rm life} \rangle$ beyond $r_{\rm in}$ of a star in that velocity bin. This latter term accounts for the possibility that stars may evolve out of the main sequence and meet their final stellar stages before they reach the maximum radial distance considered (i.e. $r_{\rm out}$) .

More precisely for a given star $t_{\rm life}$ should be equal to the time left from its main sequence lifetime $t_{\rm MS}$, after it has dwelled for a time $t_{\rm ej}$ in the GC, and subsequently travelled to $r_{\rm in}$ in a flight-time $\tau(r_{\rm in})$: $t_{\rm life} = t_{\rm MS} - (t_{\rm ej} + \tau(r_{\rm in}))$. Observations suggest that a HVS can be ejected at anytime during its lifetime with equal probability and therefore on average $t_{\rm ej} \approx t_{\rm MS}/2$ (**?**). In addition, if $\tau(r_{\rm in}) \ll t_{\rm MS}$, we can write $\langle t_{\rm life} \rangle = \langle t_{\rm MS} \rangle /2$, where $\langle t_{\rm MS} \rangle = \int (dn/dm)\, t_{\rm MS}(m) dm$ is the average main sequence life-time weighted for the star mass distribution dn/dm in a given velocity bin.

In the HVS mass and metallicity range considered here $t_{\rm MS}(m) \approx 200 - 700$ Myr (and $\langle t_{\rm MS} \rangle \approx 300 - 600$ Myr). Consequently our calculations typically show $\tau(r_{\rm in}) < t_{\rm MS}$ for velocities > 150 km s^{-1}, when adopting $r_{\rm in} = 50$ kpc. This means that $\tau(r_{\rm in}) \ll t_{\rm MS}$ in the whole velocity range of interest in this work ($v \geqslant 275$ km s^{-1}, see Section 5.3.2).

In this framework, we construct a Monte Carlo code where 10^7 binaries are drawn from the distributions described in Section 5.2 to build an ejection velocity PDF. This is used to construct the expected PDF in the outer halo (eq.5.5) between $r_{\rm in} = 50$ kpc and $r_{\rm out} = 120$ kpc (the observed radial range), using the formalism detailed above. For each bin of velocity, we calculate the $\langle t_{\rm MS} \rangle$, using the analytical formula by Hurley et al. (2000, see their equation 5). The lifetime for a star in the $2.3 - 4 M_\odot$ range is of a few to several hundred million years, but the exact value depends on metallicity (higher metallicities correspond to longer lifetimes). Until recently, solar metallicity was thought to be the typical value for the GC stellar population. However, more recent works suggest that there is a wider spread in metallicity, with a hint for a super-solar mean value (Do et al. 2015).

In the following, our fiducial model will assume:

- HVSs masses between 2.5 and 4 solar masses;
- A Kroupa ($f_{\rm m} \propto m_{\rm p}^{-2.3}$) IMF for primary stars between 2.5 and 100 solar masses;
- For a given primary mass $m_{\rm p}$, a mass ratio distribution $f_{\rm q} \propto q^\gamma$ in the range $[m_{\rm min}/m_{\rm p}, 1]$, with $m_{\rm min} = 0.1 M_\odot$ and $-10 \leqslant \gamma \leqslant 10$;
- A separation distribution $f_{\rm a} \propto a^\alpha$ between $a_{\rm min} = 2.5 \times \max[R_*, R_{\rm c}]$ and $a_{\rm max} = 10^3 R_\odot$, with $-10 \leqslant \alpha \leqslant 10$;
- A HVS mean metallicity value of $Z = 0.05$ (i.e. super-solar).

We will explore different assumptions in Section 5.5. In particular, we will investigate a top-heavy primary IMF, explore the consequence of a solar metallicity and finally assume a higher value of $m_{\rm min}$, over which we have no observational constraints in the GC. We will find that only the latter, if physically possible, may significantly impact our results and will discuss the consequences.

Examples of velocity distributions in the halo for our fiducial model are shown in Figure 5.1. Our selected data (see the Figure's caption and next Section) are over-plotted with an arbitrary binning (histogram). It is here worth reminding some of the features derived in Rossi et al. (2014). There, we analytically and numerically showed that the HVS halo velocity distribution encodes different physical information in different parts of the distribution. In particular, the peak of the distribution depends on both $V_{\rm G}$ and the binary distributions, and moves towards lower velocity for lower $V_{\rm G}$ (right panel) and higher values of $|\gamma|$ and α (left and central panels).

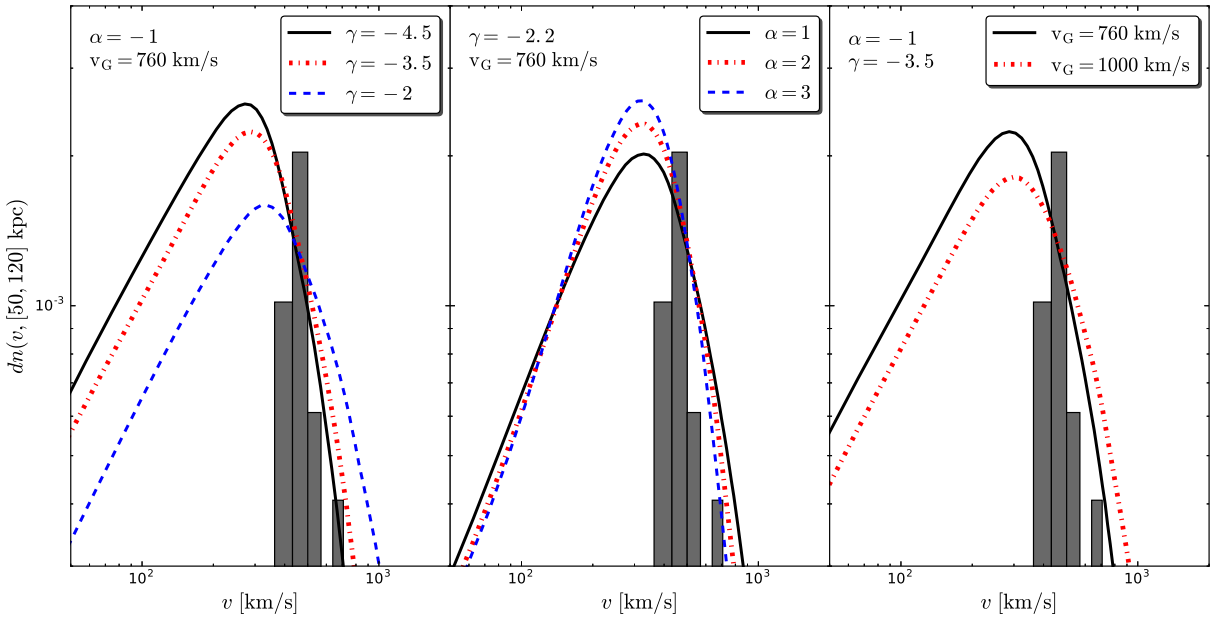


Figure 5.1: Probability density functions for HVS velocities in the outer halo of our Galaxy, between 50 kpc and 120 kpc. They are calculated following the deceleration procedure explained in Section 5.3 and depend on 3 main parameters: γ, α (for the binary mass ratio and semi-major axis distributions) and V_G . In each panel, two parameters are kept fixed while we show how the distribution changes by changing the value of the third parameter. See text for a detailed description. For a visual comparison, we over-plot data from Brown et al. (2014) ("unbound sample" only), with an arbitrary binning.

On the other hand, the high-velocity branch only depends on the binary properties, as the Galactic deceleration is negligible at those velocities. From eq.5.5, one can derive that for $v \gg v_{\rm G}$ the high-velocity branch is independent of the binary semi-major axis distribution (i.e. α) for $\gamma > -(\alpha+2)$ and

$$dn \propto v^{2\gamma} dv.$$

Therefore larger value of $|\gamma|$ result in a steeper distribution at high velocities. This is shown in the left panel of Figure 5.1. Instead in the $v \gg v_{\rm G}$ and $\gamma < -(\alpha+2)$ regime,

$$dn \propto v^{-2(\alpha+2)} dv,$$

independently of the assumed mass ratio distribution and a steeper power-law is obtained for larger α values (central panel). A discussion on the low-velocity tail, that it is *solely* shaped by the deceleration, is postponed to Section 5.4.1.

5.3.2 Comparison with data

Beside the current HVS sample of so-called "unbound" HVSs (velocity in the standard rest frame $\gtrsim 275$ km s^{-1}), there is an equal number of lower velocity "bound" HVSs[5]. Currently, it is unclear if they all share the same origin as the unbound sample, as a large contamination from halo stars cannot be excluded. We will therefore restrict our statistical comparison with data to the unbound sample (see upper part of table 1in Brown et al. 2014). As mentioned earlier, we only select HVS with masses between $2.5-4M_{\odot}$, with Galactocentric distances between 50 kpc and 120 kpc, for a total of 21 stars. These selections in velocity, mass and distance will be also applied to our predicted distributions.

Specifically, we calculate the total PDF as described by eq. 5.5 and we perform a one dimensional Kolmogorov-Smirnov (K-S) test applied to a left-truncated data sample[6]. If we call $n(<v, \Delta r)$ the cumulative probability function (CPF) for HVS velocities in the distance range Δr, then the *actual CPF* that should be compared with data is,

$$n^*(<v, \Delta r) = \frac{n(<v, \Delta r) - n(<275 \text{ km s}^{-1}, \Delta r)}{1 - n(<275 \text{ km s}^{-1}, \Delta r)}. \tag{5.6}$$

[5]Here, we simply follow the nomenclature given in Brown et al. (2014) of the two samples, even if, in fact, a knowledge of the potential is required to determine whether a star is bound and this is what we are after.

[6]See for example: Chernobai, A., Rachev, S. T., and Fabozzi, F. J. (2005). Composite goodness-of-fit tests for left-truncated loss samples. Technical Report, University of California, Santa Barbara

Therefore, the K-S test result is computed as

$$D \equiv \max[|n^*(< v, \Delta r) - n_{\rm d}(< v)|], \tag{5.7}$$

where $n_{\rm d}(< v)$ is the CPF of the actual data The significance level $\bar{\alpha} = 1 - P(D \leqslant \bar{d})$ is the probability of rejecting a fitted distribution $n(< v, \Delta r)$, when in fact it is a good fit. The most commonly used threshold levels for an acceptable fit are $\bar{\alpha} = 0.01$ and $\bar{\alpha} = 0.05$. For 21 data points $\bar{d} = 0.344$ and $\bar{d} = 0.287$ are the critical values below which the null hypothesis that the data are drawn from the model cannot be rejected at a significance level of 1% and 5% respectively.

Note that no HVS is observed with a velocity in excess of $v > 700$ km s^{-1}. Since the HVS discovery method is spectroscopic as opposed to astrometric, there is no obvious observational bias that would have prevented us from observing HVS with $v > 700$ km s^{-1}within 120 kpc and so we do not perform any high-velocity cut to our model[7]. Indeed, the absence of high-velocity HVSs in the current (small) sample suggests that they are rare, and this fact puts strong constraints on the model parameters. From the discussion in the previous section, a suppression of the high-velocity branch can be achieved by either choose a lower $V_{\rm G}$ or choose steeper binary distributions (a larger $|\gamma|$ or α), as we will explicitly show in the next section.

5.3.3 Results

In each panel of Figure 5.2, we explore the parameter space $\alpha - \gamma$ for a fixed global deceleration that brakes stars while travelling to 50 kpc, i.e. for a given $V_{\rm G}$. The contour plots show our K-S test results and models below and at the right of the white dashed line have a significance level higher than 5%: i.e. around and below that line current data are consistent with coming from models with those sets of parameters. Let us first focus on the upper right panel ($V_{\rm G} \approx 700$ km s^{-1}), as it shows clearly a common feature of all our contour plots in this parameter space. There is a stripe of minima that, from left to right, first runs parallel to the α-axis and then to the γ-axis[8].

[7]We remark in addition that our eq. 5.5 takes already into account that faster stars have a shorter residence time by suppressing their number proportionally to v^{-1}

[8]We note that, even if not completely apparent in all our panels, the K-S test values start to increase again moving towards high values of $|\gamma|$ and α: i.e. the stripe of minima has a finite size.

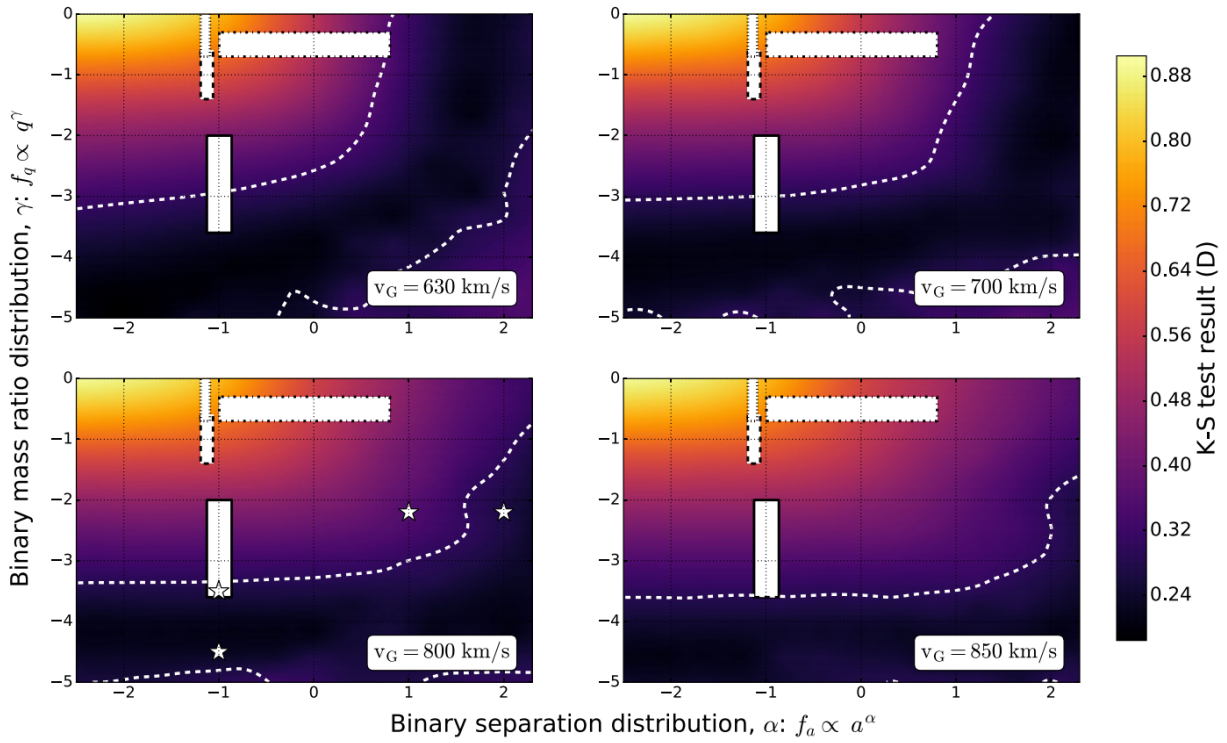


Figure 5.2: Contour plots for K-S test results in the parameter space $\alpha - \gamma$ for 4 different values of V_G (see panels' label). The white dashed line indicates the 5% significance level contours. The white regions correspond to observed properties of B-type or O-type binaries: the region enclosed by a dash-dotted line is for late B-type stars ($2 - 5M_\odot$) in the Solar Neighbourhood (Kouwenhoven et al. 2007; Duchêne & Kraus 2013); results for Galactic O-type binaries are shown within the region marked by a dotted line (Sana et al. 2012); the region enclosed by a solid (dashed) line is for early $\sim 10M_\odot$ B-type (O-type) binaries observed in 30 Doradus (Sana et al. 2013; Dunstall et al. 2015). The four stars mark the points (α, γ) in the parameter space for which the PDF is shown in Figure 5.1 (see also Fig.5.6).

This stripe is the locus of points where the high-velocity tail of the distributions has a similar slope: this happens for values of γ and α related by $\gamma \approx -(\alpha + 2)$ (see discussion of Figure 5.1 in Section 5.3.1). For negative α values (distributions with more tight binaries than wide ones), the high-velocity distribution branch is mainly shaped by the mass ratio distribution and, for example in this panel, a value around $\gamma \approx -4$ gives the best fit. On the other hand, for positive α (i.e. more wider binaries than tight ones), the high-velocity tail is shaped by the separation distribution and a value of around $\alpha \approx 2$ gives the best K-S results.

When increasing the escape velocity (from top left to bottom right) the stripe of minima moves towards the right lower part of the plots and gets further and further from the regions in the $\alpha-\gamma$ parameter space that correspond to observations of B-type binaries, and actually, to our knowledge, of *any* type of binaries currently observed with enough statistics in both star-forming and quiescent regions. We focus on observations of B-type binaries because, although our calculation consider $\sim 3M_{\odot}$ HVSs ejected from binaries with all possible mass combinations, we find that the overall velocity distribution is highly dominated by binaries where HVSs were the primary (more massive) stars, i.e. late B-type binaries[9].

In all panels, but the bottom right one, the white dashed line crosses or grazes the $\alpha - \gamma$ parameter space indicated by a white rectangle within a solid black line. We conclude that within an approximate range $V_{\rm G} \lesssim 850$ km s^{-1}, the current observed HVS velocity distribution can be explained assuming a binary statistical description in the GC that is consistent with the one inferred by Dunstall et al. (2015) for $\sim 10M_{\odot}$ B-type binaries in the star forming region of the Tarantula Nebula. In addition, for $V_{\rm G} \lesssim 630$ km s^{-1}the 5% confidence line *also* crosses the parameter space observed for Galactic B-type binaries (Kouwenhoven et al. 2007). An argument in favour of a similarity between known star forming regions and the inner GC is that, in this latter, Pfuhl et al. (2014) infer a binary fraction close to that in known young clusters of comparable age. However, we warn the reader that the Tarantula Nebula's results are affected by uncertainties beyond those represented by the nominal errors on α and γ reported by Dunstall et al. (2015) and we will discuss those in Section 5.5.

Finally, we comment on our choice to define the $V_{\rm G}$ limit using a 5% significance level threshold. If we relax this assumption and accept mod-

[9]Binaries where the HVS companions are the primary stars just contribute at a percentage level and only to the highest velocity part of the velocity distribution (see eq.5.1) in the whole parameter space explored in this work.

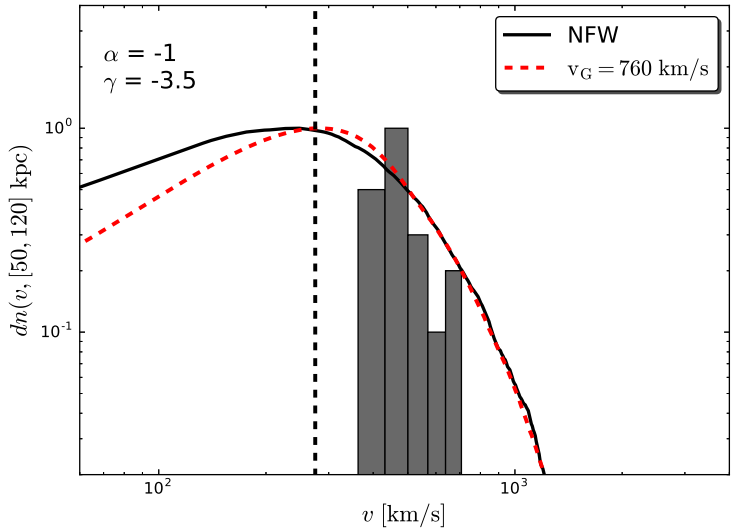


Figure 5.3: Galactic halo velocity distributions between 50 and 120 kpc for a fixed binary statistical description (see parameters in the upper left corner) but with different treatments of the star deceleration: the red dashed line is computed as described in Section 5.3.1 for $V_{\rm G} = 760$ km s^{-1}while the black solid line is our model where stars are continuously decelerated in a potential whose halo is described by a NFW profile with mass $M_{\rm h} = 0.5 \times 10^{12} M_{\odot}$ and scale radius $r_{\rm s} = 31$ kpc (see Section 5.4). This potential requires an initial velocity to escape from the GC to 50 kpc of $V_{\rm G} \approx 760$ km s^{-1}(see eq. 5.12). Unlike Figure 5.1, both model distributions and data are normalised at the peak for an easier visual comparison. The vertical dashed line marks the selection threshold ($v = 275$ km s^{-1}) of the Brown et al. unbound sample. This comparison shows that for $v \gtrsim 250$ km s^{-1}the two distributions are similar, as confirmed by the results from the K-S test ($D = 0.25$ for the black solid line and $D = 0.26$ for the red dashed line).

els with significance level $> 1\%$ (another commonly used threshold) the $V_{\rm G}$ limit moves up to $V_{\rm G} \approx 930$ km s^{-1}. On the other hand, models with $> 10\%$ significance level have $V_{\rm G} \lesssim 800$ km s^{-1}. Therefore, as a representative value, we cite here and thereafter the intermediate one of 850 km s^{-1}, corresponding to the 5% threshold.

5.4 Second approach: assuming a Galactic Potential model

We now choose a specific model to describe the Galactic potential, in order to cast our results in terms of dark matter mass and its spatial distribution.

We represent the dark matter halo of our Galaxy with a Navarro Frank and White (NFW) profile,

$$\phi(r)_{\rm NFW} = -GM_{\rm h}\left(\frac{\ln(1+r/r_{\rm s})}{r}\right), \tag{5.8}$$

(Navarro et al. 1996). In this spherical representation there are only two parameters: the halo mass $M_{\rm h}$ and the scale radius $r_{\rm s}$, where the radial dependence changes. Eq.5.8 assumes an infinite potential (no outer radius truncation) which is justified in our case since we consider Galactocentric distances smaller than the halo virial radius (~ 200 kpc).

The baryonic mass components of the Galactic potential can be described by a Hernquist's spheroid for the bulge (Hernquist 1990),

$$\phi(r)_{\rm b} = -\frac{GM_{\rm b}}{r+r_{\rm b}}, \tag{5.9}$$

(in spherical coordinates) plus a Miyamoto-Nagai disc (Miyamoto & Nagai 1975, in cylindrical coordinates, where $r^2 = R^2 + z^2$),

$$\phi_{\rm d}(R,z) = -\frac{GM_{\rm d}}{\sqrt{R^2+\left(a+\sqrt{z^2+b^2}\right)^2}}, \tag{5.10}$$

with the following parameters: $M_{\rm b} = 3.4\times10^{10}M_{\odot}$, $r_{\rm b} = 0.7$ kpc, $M_{\rm d} = 1.0\times10^{11}M_{\odot}$, $a = 6.5$ kpc and $b = 0.26$ kpc. This Galactic model have been used in modelling both HVSs and stellar streams (e.g. Johnston et al. 1995; Price-Whelan et al. 2014; Hawkins et al. 2015, and with slightly different parameters by Kenyon et al. 2008). Observationally, our choice for the bulge's mass profile is supported by the fact that its density profile is very similar to that obtained by Kafle et al. (2014), fitting kinematic data of halo stars in SEGUE[10]. In addition Kafle et al. (2014) use our same model for the disc mass distribution and their best fitting parameters are very similar to our parameters (see their table 1 and 2). However, different choices may also be consistent with current data, and we will discuss the impact of different baryonic potentials on our results in Section 5.4.2.

In a potential constituted by the sum of all Galactic components,

$$\phi_{\rm T}(r, M_{\rm h}, r_{\rm s}) = \phi(r(R,z))_{\rm d} + \phi(r)_{\rm b} + \phi(r)_{\rm NFW}\,, \tag{5.11}$$

[10]The Kafle et al. (2014) model for the bulge is not spherical (see their table 1), therefore we compare to our model both their spherically averaged density profile and their density profile at 45° latitude (see Section 5.4 for a justification of this latter).

we integrate each star's trajectory from an inner radius $r_{\rm start} = 3$ pc, equal to Sgr A*'s sphere of influence but any starting radius $r_{\rm start} < 20$ pc gives very similar results. In fact, we find that the disc's sky-averaged deceleration is overall negligible with respect to that due to the bulge. To save computational time, we therefore set $R = z = r/\sqrt{2}$ in equation 5.10 (i.e. we only consider trajectories with a Galactic latitude of 45°), simplifying our calculations to one-dimensional (the Galactocentric distance r) solutions.

The star's initial velocity is drawn from the ejection velocity distribution, constructed as detailed in Section 5.2. Assumptions on HVS properties are those of our fiducial model. Informed by observations (Brown et al. 2014), we assigned a flight-time from a flat distribution between $[0, t_{\rm MS}]$. Each integration of 10^7 star orbits gives a sky realisation of the velocity PDF, but we actually find that the number of stars we are tracking is sufficiently high that differences between PDFs associated to different realisations are negligible.

An example of a halo velocity distribution is shown in Figure 5.3 with a black solid line. This accurate calculation of the star deceleration is well approximated by using eq.5.4 for $v \gtrsim 250$ km s^{-1}, when the escape velocity at 50 kpc is calculated as

$$V_{\rm G}^2 = 2(\phi_{\rm T}\,(50\ {\rm kpc}, M_{\rm h}, r_{\rm s}) - \phi_{\rm T}(r_{\rm start}, M_{\rm h}, r_{\rm s}))\,, \tag{5.12}$$

(red dashed line in Figure 5.3). Despite the discrepancy in the behaviour of the low velocity tail, the two approaches give very similar K-S test results when compared to current observations ($D = 0.26$ for the NFW model versus $D = 0.25$ for the "$V_{\rm G}$" model). With a random sampling, we tested that K-S results differ at most at percentage level in the whole extent of the parameter space of interest to us, validating our first approach, as an efficient and reliable exploratory method.

5.4.1 The low-velocity tail

We here pause to discuss and explain the difference in the velocity distribution around and below the peak calculated with our two approaches (see Figure 5.3). Without loss of indispensable information, the impatient reader may skip this section and proceed to the next one, where we discuss our results.

The low velocity tail discrepancy is due to our two main assumptions of our first method: i) neglecting the residual deceleration beyond 50 kpc; and ii) all stars reach 50 kpc before they evolve out of the main sequence.

The residual deceleration gives an excess of low velocity stars in the correct distribution (black solid line) that cannot be reproduced by our approximated calculation (red dashed line). On the other hand, a fraction of stars that should have ended up with velocities $\lesssim 150\,\mathrm{km\,s^{-1}}$beyond $50\,\mathrm{kpc}$ have in fact flight-times longer than their life-time and the low velocity excess is slightly suppressed in that range.

Let us be more quantitative. In the framework of our first approach, one can show that the PDF at low velocities increases linearly with v (Rossi et al. 2014). The calculation is as follows. The rate of HVSs crossing $r = r_{\rm in}$ with $v = \sqrt{v_{\rm ej}^2 - V_{\rm G}^2} \ll V_{\rm G}$ is given by

$$\frac{d\dot{N}}{dv} \sim \mathcal{R}\; P(v_{\rm ej})\big|_{v_{\rm ej}=V_{\rm G}}\; \frac{v}{V_{\rm G}}.$$

Moreover, for[11]

$$v < \Delta r/\langle t_{\rm MS}\rangle \approx 230\,\mathrm{km\,s^{-1}}(\Delta r/70\mathrm{kpc})(300/\,\mathrm{Myr}/\langle \mathrm{t_{MS}}\rangle),$$

the residence time within Δr is equal to (half of) the stars' life-time, therefore from eq.5.5 we conclude that

$$\frac{dn(v,\Delta r)}{dv} \propto\; P(v_{\rm ej})\big|_{v_{\rm ej}=V_{\rm G}}\; v \times \langle t_{\rm MS}\rangle\,,$$

recovering the linear dependence on v. In fact, $\langle t_{\rm MS}\rangle$ is not completely independent of v as it varies by a factor of ≈ 1.5 as $v \to 0$. Therefore dn/dv is slightly sub-linear in v. The dependence of $\langle t_{\rm MS}\rangle$ on v comes about because $v_{\rm ej}$ is proportional to $m_{\rm c}$. This causes low-velocity HVSs to be increasingly of lower masses ($\to 2.5M_\odot$), being ejected from binaries where their companions were all lighter $m_{\rm c} \lesssim 2.5M_\odot$ than the companions of more massive HVSs.

When considering instead the full deceleration of stars in a gravitational potential $a = -d\phi_{\rm T}(r)/dr$ as they travel towards $r_{\rm out}$, their velocity depends both on $v_{\rm ej}$ and r,

$$v(v_{\rm ej}, r) = \sqrt{v_{\rm ej}^2 - (V_{\rm esc}(0)^2 - V_{\rm esc}(r)^2)}, \tag{5.13}$$

where $V_{\rm esc}(r)$ is the escape velocity from a position r to infinity (i.e. $V_{\rm esc}(0)$ is the escape velocity from the GC to infinity). Note that $V_{\rm G} = \sqrt{V_{\rm esc}(0)^2 - V_{\rm esc}(r_{\rm in})^2}$.

[11] We remind the reader that $\Delta r = r_{\rm out} - r_{\rm in}$.

In the example shown in Figure 5.3, $V_{\rm esc}(0) \approx 826\,{\rm km\,s^{-1}}$, $V_{\rm esc}(r_{\rm in} = 50\,{\rm kpc}) \approx 323$ km s^{-1}, $V_{\rm esc}(r_{\rm out} = 120$ kpc$) \approx 257$ km s^{-1}and $V_{\rm G} \approx 760$ km s^{-1}. On the other hand, the distance r is a function of both $v_{\rm ej}$ and the flight-time $\tau(r) = \int dv(r)/\left|a(r)\right|$, and this latter is a preferable independent variable because uniformly distributed. Therefore we express $v = v(v_{\rm ej}, \tau)$ and

$$\frac{dn}{dv} \propto \int_0^{\langle t_{\rm MS}\rangle} \int_{v_{\rm ej,\,min}}^{v_{\rm ej,\,max}} \delta(v - v(v_{\rm ej}, \tau)) P(v_{\rm ej}) dv_{\rm ej} d\tau, \tag{5.14}$$

where the relevant ejection velocity range is that that gives low-velocity stars between $r_{\rm in}$ and $r_{\rm out}$: $v_{\rm ej,min} = \sqrt{v^2 + (V_{\rm esc}(0)^2 - V_{\rm esc}(r_{\rm in})^2)}$ and $v_{\rm ej,max} = \sqrt{v^2 + (V_{\rm esc}(0)^2 - V_{\rm esc}(r_{\rm out})^2)}$. Note that, for Galactic mass distribution where $V_{\rm esc}(0) > V_{\rm esc}(r_{\rm in}), V_{\rm esc}(r_{\rm out})$, the range $[v_{\rm ej,min} - v_{\rm ej,max}]$ is rather narrow and for $v \ll V_{\rm G}$ these limits may be taken as independent of v. This is the case in the example of Fig. 5.3, where $v_{\rm ej,min} \approx V_{\rm G} \approx 760 < v_{\rm ej}[{\rm km\,s^{-1}}] < {\rm v}_{\rm ej,max} \approx 785$.

It follows that the low-velocity tail is populated by stars that where ejected with velocities slightly higher than $V_{\rm G}$. If we further assume that the flight-time τ to reach any radius within $r_{\rm out}$ is always smaller than $\langle t_{\rm MS}\rangle$ (formally this means putting the upper integration limit in τ equal to infinity), then all HVSs ejected with that velocity reach 50 kpc. It may be therefore intuitive that, applying the above considerations, eq.5.14 reduces to

$$\frac{dn}{dv}(v, \Delta r) \propto P(v_{\rm ej})\big|_{v_{\rm ej}=V_{\rm G}} \int_{r_{\rm in}}^{r_{\rm out}} \frac{dr}{v_{\rm ej}(r)} \approx P(v_{\rm ej})\big|_{v_{\rm ej}=V_{\rm G}} \frac{\Delta r}{V_{\rm G}}, \tag{5.15}$$

where we substitute $d\tau = dv/|a|$ in eq.5.14 and we use eq.5.13. We therefore recover the flat behaviour for $v \lesssim 300$ km s^{-1}of the black solid line in Figure 5.3. We, however, also notice that below ~ 150 km s^{-1}there is a deviation from a flat distribution: this is because our assumption of $\tau(r_{\rm in}) \ll \langle t_{\rm MS}\rangle$ breaks down, as not all stars reach 50 kpc, causing a dearth of HVSs in that range.

As a concluding remark, we stress that, although we do not apply it here, the result stated in eq.5.15 can be used to further improve our first method, a necessity when low-velocity data will be available.

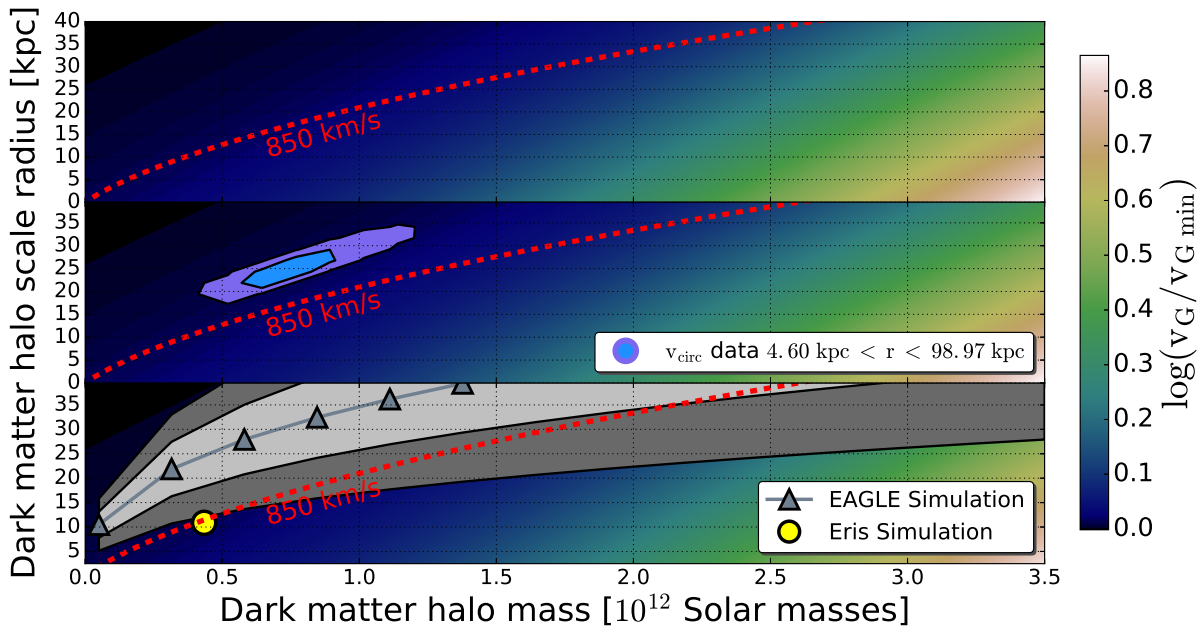


Figure 5.4: *Upper panel:* the "escape" velocity from the GC to 50 kpc, $V_{\rm G}$, over the minimum allowed by the presence of a baryonic disc and bulge ($V_{\rm G,min} = 725$ km s^{-1}) is mapped onto the $M_{\rm h} - r_{\rm s}$ parameter space for NFW dark halo profiles using eq. 5.12. The iso-contour line equal to $V_{\rm G} = 850$ km s^{-1}is explicitly marked as red dashed line. *Middle panel:* same as the upper panel but over-plotted are the results of our MCMC analysis of the Galactic circular velocity data from Huang et al. (2016) (see Appendix .1). *Lower panel:* the same as the upper panel but over-plotted are results from the Eris (Guedes et al. 2011) and EAGLE (Schaye et al. 2015) simulations. These are dark matter plus baryons simulations: the first one is a single realisation of a Milky Way-type galaxy, the latter are cosmological simulations that span a wider range of masses $(10^{10} - 10^{14} M_{\odot})$. Following Schaller et al. (2015), figure 11 middle panel, we plot the mass concentration relation found in EAGLE in our mass range, with a scatter in the concentration parameter of 25% at one sigma level.

5.4.2 Results

The relation given by eq. 5.12 allows us to map a given $V_{\rm G}$ value onto the $M_{\rm h} - r_{\rm s}$ parameter space. This is shown in Figure 5.4, upper panel. Note that for a given choice of the baryonic mass components of the potential, there is an absolute minimum for $V_{\rm G}$ (thereafter $V_{\rm G,min}$) , that corresponds to the absence of dark matter within 50 kpc. For our assumptions (eqs. 5.9 and 5.10), $V_{\rm G,min} \approx 725$ km s^{-1}. In other words, this is the escape velocity from the GC only due to the deceleration imparted by the mass in the disc and bulge components.

In Figure 5.4, the red dashed curve marks the iso-contour equal to $V_{\rm G} = 850$ km s^{-1}: above this curve $V_{\rm G,min} \lesssim V_{\rm G} < 850$ km s^{-1}. For a scale radius of $r_{\rm s} < 30$ kpc, this region corresponds to $M_{\rm h} < 1.5 \times 10^{12} M_{\odot}$, but, if larger $r_{\rm s}$ can be considered, the Milky Way mass can be larger. This parameter degeneracy is the result of fitting a measurement that — as far as deceleration is concerned — solely depends on the shape of the potential within 50 kpc: lighter, more concentrated haloes give the same net deceleration as more massive but less concentrated haloes. The $V_{\rm G} = 850$ km s^{-1}line stands as an indicative limit above which, for a given halo mass, HVS data can be fitted at $> 5\%$ significance level assuming a B-type binary population in the GC close to that inferred in the LMC. In fact, since in our case $V_{\rm G,min} > 630$ km s^{-1}, the observed Galactic binary statistics never gives a high significance level fit to current data (see Section 5.3.3).

To gain further insight into the likelihood of various regions of the parameter space, we compare our results to additional Milky Way observations and theoretical predictions. We compute the circular velocity $V_{\rm c} = \sqrt{GM(<r)/r}$ along the Galactic disc plane, where $M(<r)$ is the total enclosed mass (obtained integrating eq. 5.11). We compare it to a recent compilation of data from Huang et al. (2016), which traces the rotation curve of the Milky Way out to ~ 100 kpc. Specifically, using a Markov Chain Monte Carlo (MCMC) technique (see Appendix .1), we find that a relatively narrow region of the parameter space leads to a fair description of the circular velocity data. As shown in the middle panel of Fig. 5.4, the preferred combinations of $r_{\rm s}$ and $M_{\rm h}$ lie above our $V_{\rm G} \sim 850$ km s^{-1}iso-velocity line and the best fitting parameters are $M_{\rm h} \approx 8 \times 10^{11} M_{\odot}$ and $r_{\rm s} \approx 25$ kpc. More generally, $r_{\rm s}$ greater than ~ 30 (~ 35) kpc for our Galaxy can be excluded at, at least, one-sigma (two-sigma) level (see also Figure 7 right panel). This may be intuitively understood as follows. At distances where dark matter dominates, $r_{\rm s}$ sets the scale beyond which $V_{\rm c} \propto \sqrt{(M(<r)/r)} \sim \sqrt{\log r / r}$, while for $r < r_{\rm s}$ $V_{\rm c} \propto \sqrt{r}$. Therefore, a scale radius larger than ~ 30 kpc

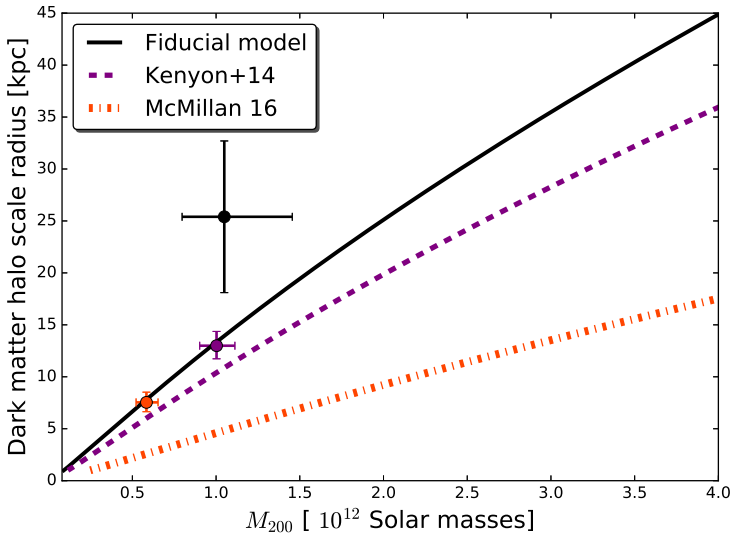


Figure 5.5: Dark Halo mass (M_{200}) versus dark matter scale radius (r_s) for 3 different models for the Galactic potential: the model presented in Section 5.4 ("Fiducial model"), the one adopted by Kenyon et al. (2014) and one which combines our disc model and a symmetric average of the bulge matter density profile, as reported by McMillan (2017). The plotted lines are combinations of mass and radius that give an escape velocity from the GC of 850 km s^{-1}. Over-plotted in matching colours for each Galactic potential model are the best fitting parameters for the Galactic circular velocity (see Appendix .1). Note that a mixed model with the **?**'s bulge and the Kenyon et al.'s parameters for the disc gives intermediate results.

cannot account for the observed rather flat/slowly decreasing behaviour of the circular velocity at distances of $\gtrsim 20$ kpc (see Figure 7 left panel). In addition, for a fixed M_h, large scale radii produce values of V_c lower than the measured $V_c \sim 200$ km s^{-1}in the halo region.

The lowest panel of Fig. 5.4 shows the values of M_h and r_s found in the EAGLE hydro-cosmological simulation (Schaye et al. 2015) and reported by Schaller et al. (2015). The region of parameter space within $V_G < 850$ km $s^{-1}$and $r_s \lesssim 35$ kpc fully overlaps with the one-sigma and two-sigma regions determined using the haloes in the EAGLE simulation. We also plot the M_h and r_s values that describe the halo in the Eris simulation (Guedes et al. 2011) and note that they lie at the edge of the lowest two-sigma confidence region.

5.4.3 Impact of different disc and bulge models

The mapping $V_{\rm G} \rightarrow (M_{\rm h} - r_{\rm s})$ depends on the assumed baryonic matter density distribution, upon which there is no full general agreement (see Bland-Hawthorn & Gerhard 2016, for a recent observational review on the Galactic content and structure). In particular, both the total baryonic mass and its concentration can have an impact. The most recent works point towards a stellar mass in the bulge around $1 - 2 \times 10^{10} M_{\odot}$ (e.g. Portail et al. 2015), but one should be aware of uncertainties given by the fact that different observational studies of the bulge constrain the mass in different regions and the size of the bulge is not universally defined. Moreover, the bulge's mass is distributed in a complex box/peanut structure, coexisting with an addition spherical component (see Gonzalez & Gadotti 2016, for an observational review on the bulge). The corresponding 3-dimensional density profile down to the sphere of influence of Sgr A*, is therefore uncertain. Likewise for the disc component, there are ongoing efforts to try and construct a fully consistent picture, that is currently missing (see Rix & Bovy 2013, for a recent review on the stellar disc). Recent estimates place the total disc mass around $5 \times 10^{10} M_{\odot}$, a factor of two lighter than the disc mass we adopt in Fig.5.4.

Given these uncertainties, we here explore the impact of adopting different baryonic components than the ones we assumed in Section 5.4, where a justification for that choices is stated. In particular, we explore lighter components, differently distributed. To do this, we compare in Figure 5.5 the loci of $V_{\rm G} = 850$ km s^{-1}in the plane $(M_{200} - r_{\rm s})$, given by other two Galactic potential models that together with ours should frame a plausible uncertainty range. We chose to plot here M_{200}[12] instead of $M_{\rm h}$ as it is commonly used to indicate the Milky Way dark matter mass and it can facilitate comparisons with results from other probes.

The potential adopted by Kenyon et al. (2014) and widely used in the HVS community is shown with a dashed line: the bulge and disc components are described by our eqs. 5.9 and 5.10 but with different parameters ($M_{\rm b} = 3.76 \times 10^9 M_{\odot}$, $r_{\rm b} = 0.1$ kpc, $M_{\rm d} = 6 \times 10^{10} M_{\odot}$, $a = 2.75$ kpc, $b = 0.3$ kpc). Comparing the solid and dashed lines one concludes that, for a given $r_{\rm s}$, the Kenyon et al.'s model gives $\sim 30\%$ more massive haloes. We then calculate the $V_{\rm G} = 850$ km s^{-1}iso-courve for a bulge potential advocated by McMillan (2017) plus our fiducial model for the disc (dash-dotted line).

[12]This is the mass enclosed within a sphere of mean density equal to 200 times the critical density of the Universe at $z = 0$

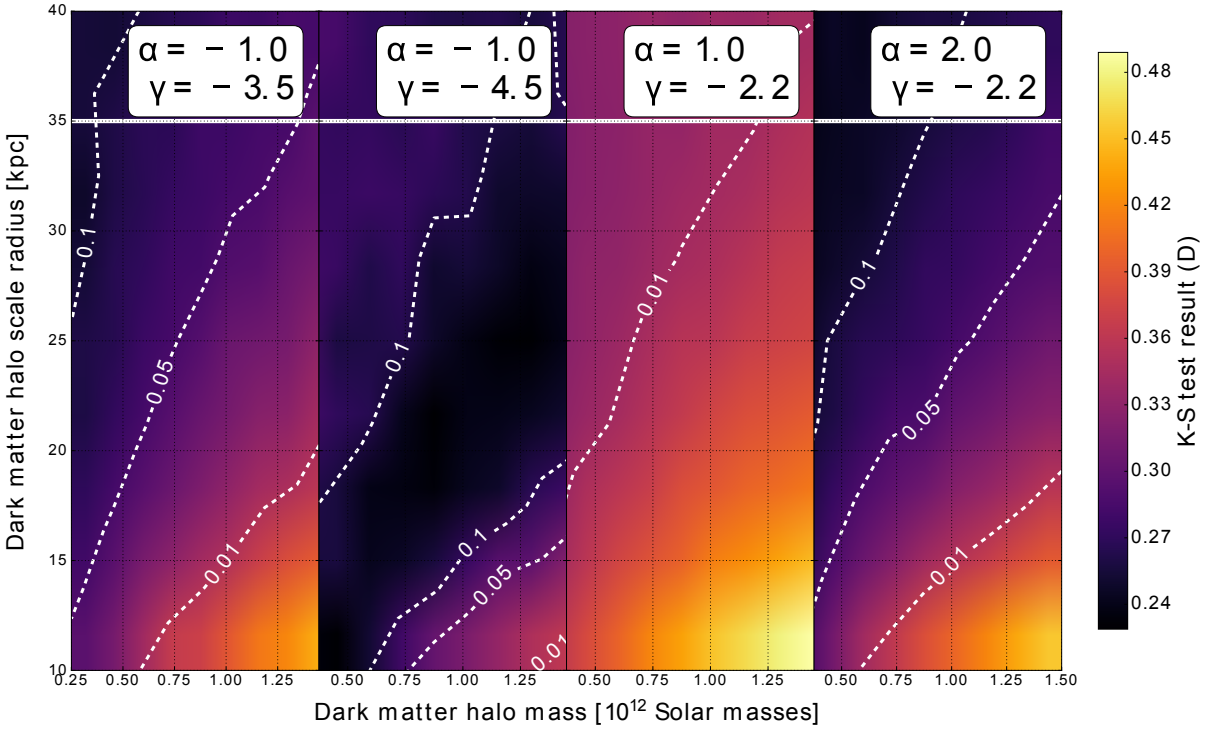


Figure 5.6: Contour plots for K-S test results in the parameter space $M_{\rm h} - r_{\rm s}$, for fixed α, γ pairs (see panels' label and star marks in Figure 5.4). Velocity distributions are computed radially decelerating each star in a given potential (see Section 5.4). The white dashed lines are iso-contour lines for a given significance level $\bar{\alpha}$. Regions at the left of of each line have a value of $\bar{\alpha}$ larger than that stated in the corresponding label.

The McMillan's bulge model adopts a total mass of $\simeq 8.9 \times 10^9 M_\odot$ and it is not spherically symmetric. We therefore radially average the axisymmetric density profile before computing the corresponding potential[13]. Note that the McMillan's bulge model is more massive than the Kenyon et al.'s one but equally concentrated, resulting in a very different density profile. Consequently, this model gives significantly more massive haloes (by a factor $\gtrsim 2$) than we obtain with either Kenyon et al.'s or our fiducial model.

We conclude that the impact of these uncertainties on the determination of the halo mass with HVS data is large and cannot be ignored. In order to put robust constraints on the dark matter halo of our Galaxy through our method a multi-parameter fit of data is therefore required where both the disc and bulge parameters need to be left free to vary. We defer this more sophisticated analyses, however, when more and better HVS data will be available.

On the positive side, the main features of the two regions in the $M_\mathrm{h} - r_\mathrm{s}$ parameter space defined by our $V_\mathrm{G} = 850$ km s^{-1}remain the same, regardless of the specific baryonic potentials: the best fitting models for the circular velocity data always lie within the $V_\mathrm{G} < 850$ km s^{-1}region (see crosses in Figure 5.5 and Appendix .1), as do the EAGLE's predictions for ΛCDM compatible haloes.

5.5 Discussion and conclusions

The analysis presented in the paper yields the following main results:

1 For a $> 5\%$ ($> 1\%$) significance level fit, HVS velocity data *alone* require a Galactic potential with an escape velocity from the GC to 50 kpc $\lesssim 850$ km s^{-1}($\lesssim 930$ km s^{-1}), when assuming that binary stars within the innermost few parsecs of our Galaxy are not dissimilar from binaries in other, more observationally accessible *star forming* regions. For $V_\mathrm{G} \sim 630$ km s^{-1}, the binary statistics for late B-type stars observed in the Solar neighbourhood also provide a fit at the same significance level.

2 When specialising to a NFW dark matter halo, we find that the region $V_\mathrm{G} \lesssim 850$ km s^{-1}contains models that are compatible with both HVS

[13]Indeed, we are comparing our models with a radially averaged observed distribution of HVS velocities beyond 50 kpc, we can therefore assume a spherically symmetric bulge, since its spatial extension is no more than a few kpc.

and circular velocity data. These models also correspond to ΛCDM-compatible Milky Way haloes. In principle, we cannot exclude the parameter space $V_G \gtrsim 850$ km s^{-1}. However, it would require us to face both an increasingly different statistical description of the binary population in the GC with respect to current observations *and* dark matter haloes that are inconsistent with predictions in the ΛCDM model at one-sigma level or more (see lower panel of Figure 5.4).

3 The result stated in point 2 is *independent* of the assumed baryonic components of the Galactic potential, across a wide range for plausible masses and scale radii.

4 However, the *specific* mapping of V_G values onto the $M_h - r_s$ parameter space *is* highly dependent on the assumed bulge and disc models (see Section 5.4.3). Both the baryonic total mass and its distribution affect the results. In general, works that try to infer the dark matter halo mass from HVS data should fold in the uncertainties linked to our imperfect knowledge of the baryonic mass distribution.

These results rely on certain assumptions for the binary population in the GC whose impact we now discuss. Following the same computational procedure previously presented for our fiducial model, we have found that a different mass function for the primary stars (either a Salpeter or a top-heavy mass function) or a change in metallicity (from super-solar to solar) do not substantially alter our results. However, the choice of the minimum companion mass (i.e. m_{min} in eq. 5.3) does lead to different conclusions. In particular, the higher m_{min}, the steeper the binary distributions should be to fit the data, even for low (< 850 km s^{-1}) V_G. For example, for $m_{min} = 0.3 M_\odot$ (instead of 0.1 $M_\odot$) and $V_G = 760$ km s^{-1} the stripe of minima for the K-S test runs along the $\gamma \approx -6.5$ and $\alpha \approx 4.5$ directions, very far from the observed values. Currently, there is no observational or theoretical reason why we should adopt a higher minimum mass than the one usually assumed ("the brown dwarf" limit), but this exercise shows that better quality and quantity HVS data has the potential to statistically constrain the minimum mass for a secondary, which may shed light on star and/or binary forming mechanisms at work in the GC.

A second set of uncertainties that may affect our conclusions pertain to the observed binary parameter distributions in the 30 Doradus region, that we use as guidance. The 30 Doradus B-type sample of Dunstall et al. (2015) is based on 6 epochs of spectra, that do not allow for a full orbital solution

for each system. These authors' results are mainly based on the distribution of the maximum variation in radial velocities per system, from where they statistically derive constraints for the full sample. Another point worth stressing is that the 30 Doradus B-type sample is of *early* type stars (mass roughly around $10 M_\odot$) and distributions for *late* B-type star binaries in star forming regions may be different. However, these latter are not currently available, and therefore the Dustall et al. sample remains the most relevant to guide our analysis in those regions. Our statement is therefore that the statistical distributions derived from this sample (including the statistical errors on the power-law indexes) can reproduce HVS data at a several percentage confidence level. Far more reliable is the statistical description of observed late B-type binaries in the Solar neighbourhood, that can be easily reconciled with HVS data *only* for quite low V_G potentials.

A possibility that we have not so far discussed is that dynamical processes that inject binaries within Sgr A*'s tidal sphere modify the natal mass ratio and separation distributions. Unfortunately, as far as we know, dedicated studies are missing and we will then only discuss the consequence of the classical loss-cone[14] theory" dealing with two-body encounters (e.g. Frank & Rees 1976; Lightman & Shapiro 1977) as derived in Rossi et al. (2014, section 3). Their considerations show that even allowing for extreme regimes, one would expect no modification in the mass ratio distribution and a modification in the separation distribution by no more than a factor of "a" (i.e. a natal Öpik's law would evolve into $f_a \sim$ const.). This would increase the V_G range ($V_G \lesssim 750$ km s^{-1}) compatible with Solar neighbourhood observations (see Fig. 5.2). Beside that, all our results remain unchanged.

We would also like to remark here that, although observed binary parameters give acceptable fits for $V_G < 930$ km s^{-1}, the K-S test results currently prefer even steeper mass ratio and binary separation distributions ($\gamma \sim -4.5$ instead of $\gamma \sim -3.5$ and/or $\alpha \sim 2$ instead of -1, see Fig. 5.6). This larger $|\gamma|$ value gives a steeper high velocity tail, which better match the lack of observed > 700 km s^{-1}HVSs. From the above considerations, modification of the natal distribution by standard two-body scattering into the binary loss cone may not be held responsible. Assuming that the halo actually has $V_G < 930$ km s^{-1}, one possible inference is indeed that $\gamma \sim -4.5$ is

[14]The loss cone theory deals with processes by which stars are "lost" because they enter the tidal sphere, in which they will suffer tidal disruption on a dynamical time. The name comes from the fact that the tidal sphere is defined in velocity space at a fixed position as a "cone" with an angle proportional to the angular momentum needed for the (binary) star to be put on an orbit grazing the tidal radius (see for e.g. Alexander 2005, section 6.1.1).

a better description of the B-type binary natal distribution in the GC, close but not identical to that in the Tarantula Nebula.

It is of course possible that some other dynamical interactions (e.g. binary softening/hardening, collisions) or disruption of binaries in triples could be indeed responsible for a change in γ and a larger one in α. However, for massive binaries dynamical evolution of their properties may be neglected in the GC, because it would happen on timescales longer than their lifetime (Pfuhl et al. 2014). On the contrary, it may be relevant for low mass binaries, but only within the inner 0.1 pc (Hopman 2009). Nevertheless, these possibilities would be very intriguing to explore in depth, if more and better data on HVSs together with a more solid knowledge of binary properties in different regions will still indicate the need for such processes.

Finally, given the paucity of data, we did not use any spatial distribution information but we rather fitted the velocity distribution integrated over the observed radial range. This precluded the possibility to meaningfully investigate anisotropic dark matter distributions and we preferred to confine ourselves to spherically symmetric potentials.

All the above uncertainties and possibilities can and should be tested and explored when a HVS data sample that extends below and above the velocity peak is available. Such a data set would allow us to break the degeneracy between halo and binary parameters, as the rise to the peak and the peak itself are mostly sensitive to the halo properties, whereas the high velocity tail is primarily shaped by the binary distributions. This will be achieved in the coming few years thanks to the ESA mission *Gaia*, whose catalogue should contain at least a few hundred HVSs with precise astrometric measurements. Moreover *Gaia* will greatly improve our knowledge of binary statistics in the Galaxy (but not directly in the GC, where infrared observations are required) and in the LMC allowing us to draw more robust inferences.

In conclusion, this paper shows for the first time the potential of HVS data combined with our modelling method to extract joint information on the GC and (dark) matter distribution. It is clear, however, that the full realisation of this potential requires a larger and less biased set of data. The ESA *Gaia* mission is likely to provide such a sample within the coming five years.

Acknowledgements

We thank S. de Mink, R. Schödel for a careful reading of the manuscript and O. Gnedin, M. Viola, H. Perets, E. Starkenburg, J. Navarro, A. Helmi, S. Kobayashi and A. Brown for useful discussions and comments. We also thank the anonymous referee for the careful reading of the manuscript and his/her useful suggestions. E.M.R. and T.M. acknowledge the supported from NWO TOP grant Module 2, project number 614.001.401.

.1 Markov Chain Monte Carlo to fit the observed circular velocity

To assess which ranges of the halo mass and scale radius are compatible with current constraints of the Milky Way halo, we employ circular velocity measurements presented in Huang et al. (2016) where the rotation curve of the Milky Way out to ~ 100 kpc has been constructed using ~ 16,000 primary red clump giants in the outer disc selected from the LAMOST Spectroscopic Survey of the Galactic Anti-centre (LSS-GAC) and the SDSS-III/APOGEE survey, combined with ~ 5700 halo K giants selected from the SDSS/SEGUE survey. These measurements are reported in Figure 7 left panel as green points with error bars.

We remind the reader that our model for the matter density (and thus the circular velocity) of the Milky Way consists of three components: a bulge, a disc, and an extended (dark matter) halo. While bulge and disc dominate the circular velocity at relatively small scales (below about 30 kpc), larger scales are dominated by the dark matter halo. Each of these components for all models we consider is described in detail in the main body of the paper (see Sections 5.4 and 5.4.3). To fit the data described above we fix the parameters that refers to the bulge and the disc, whereas we consider as free parameters those related to the dark matter halo. We remind that dark matter halo is assumed to have a NFW matter density profile, completely characterised by two parameters: the total halo mass, $M_{\rm h}$, and the scale radius, $r_{\rm s}$.

The two-dimensional parameter space $(M_{\rm h}, r_{\rm s})$ is sampled with an affine invariant ensemble Markov Chain Monte Carlo (MCMC) sampler (Goodman & Weare 2010). Specifically, we use the publicly available code Emcee (Foreman-Mackey et al. 2013). We run Emcee with three separate chains with 200 walkers and 4 500 steps per walker. Using the resulting 2 700 000 model evaluations, we estimate the parameter uncertainties. We assess the

convergence of the chains by computing the auto-correlation time (see e.g. Akeret et al. 2013) and finding that our chains are about a factor of 20 times longer than it is needed to reach 1% precision on the mean of each fit parameter.

The left panel of Figure 7 shows the circular velocity as a function of distance from the GC. Green points with error bars are taken from table 3 of Huang et al. (2016), whereas orange and yellow shaded regions correspond to the 68th and 95th credibility intervals obtained from the MCMC procedure described above for our fiducial model (Section 5.4). Different line styles and colours refer to the different contributions as detailed in the legend. The MCMC leads to a best-fit χ^2 of 39.07 with $N_{\rm data} = 43$ data points and $N_{\rm par} = 2$ model parameters, thus resulting in a satisfactory reduced $\chi^2_{\rm red} = \chi^2/(N_{\rm data} - N_{\rm par}) = 0.95$. Comparable level of agreement between models[15] and data is obtained when adopting i) a model that combines our fiducial disc parameters with a lighter bulge from McMillan (2017) ($\chi^2_{\rm red} = 1.34$) or ii) Kenyon et al. (2014)'s much lighter disc and bulge models ($\chi^2_{\rm red} = 0.88$).

The right panels of Figure 7 show the posterior distribution of the halo parameters for the three baryonic models mentioned above. As expected, the two halo parameters are strongly degenerate but the sampling strategy has nevertheless finely sampled the region of high likelihood. For our fiducial baryonic model, we find that $\log[M_{\rm h}/M_\odot] = 11.89 \pm 0.18$, and $r_{\rm s} = 25.4 \pm 7.3$ kpc, where we quote the median and errors are derived from the 16th and 84th percentiles. For i) instead the best fitting parameters are $\log[M_{\rm h}/M_\odot] = 11.42 \pm 0.06$, and $r_{\rm s} = 7.5^{+1.0}_{-0.9}$ kpc, while ii) gives intermediate results: $\log[M_{\rm h}/M_\odot] = 11.72 \pm 0.06$, and $r_{\rm s} = 12.99^{+1.4}_{-1.3}$ kpc.

[15] A mixed model that combines Kenyon at al.'s disc and McMillan's bulge gives results very similar to that obtained with Kenyon et al. (2014) disc and bulge models, so we will not discuss it further.

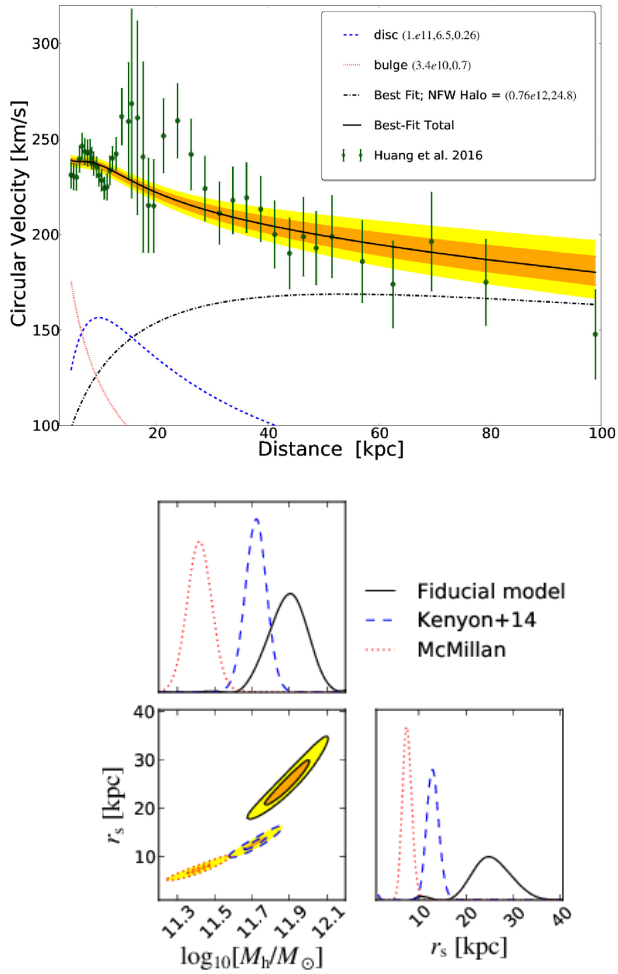


Figure 7: *Top panel:* Galactic circular velocity. Data points with error bars are taken from Huang et al. (2016). The orange and yellow regions correspond to the 68th and 95th credibility interval obtained with the MCMC described in the text for our fiducial Galactic Potential model. Red dotted and blue dashed lines represent the contribution from the bulge and the disc, respectively, whereas the dash-dotted black line indicates the contribution from the best-fitting NFW halo. The solid black line corresponds to the total circular velocity for the best-fitting model ($\chi^2_{\rm red} = 0.95$). *Bottom panel:* Posterior distributions of the two halo parameters, $\log_{10}[\mathrm{M_h/M_\odot}]$ and $r_{\rm s}$, as obtained from the MCMC used to fit the Galaxy circular velocity measurements with the three models discussed in the text (see also legend). The diagonal panels show the the posterior distributions for each parameter. The lower left panel shows the two-dimensional marginalised posterior distributions. As expected, the two parameters are strongly degenerate. Orange (yellow) region indicates the extent of the 68% (95%) credibility interval.

Bibliography

Aarseth S. J., 1974, A&A, 35, 237
Abadi M. G., Navarro J. F., Steinmetz M., 2009, ApJ, 691, L63
Akeret J., Seehars S., Amara A., Refregier A., Csillaghy A., 2013, Astronomy and Computing, 2, 27
Alexander T., 2005, Phys. Rep., 419, 65
Andrae R., et al., 2018, A&A, 616, A8
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
Bailer-Jones C. A. L., Rybizki J., Fouesneau M., Mantelet G., Andrae R., 2018, AJ, 156, 58
Balick B., Brown R. L., 1974, ApJ, 194, 265
Barbier-Brossat M., Petit M., Figon P., 1994, A&AS, 108
Bartko H., et al., 2010, ApJ, 708, 834
Bessell M. S., 1990, PASP, 102, 1181
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
Böker T., 2010, in de Grijs R., Lépine J. R. D., eds, IAU Symposium Vol. 266, Star Clusters: Basic Galactic Building Blocks Throughout Time and Space. pp 58–63 (arXiv:0910.4863), doi:10.1017/S1743921309990871
Boubert D., Evans N. W., 2016, ApJ, 825, L6
Boubert D., Erkal D., Evans N. W., Izzard R. G., 2017a, MNRAS, 469, 2151
Boubert D., Fraser M., Evans N. W., Green D. A., Izzard R. G., 2017b, A&A, 606, A14
Boubert D., Guillochon J., Hawkins K., Ginsburg I., Evans N. W., Strader J., 2018, MNRAS, 479, 2789
Boubert D., et al., 2019, MNRAS, 486, 2618
Bovy J., 2015a, ApJS, 216, 29
Bovy J., 2015b, ApJS, 216, 29

Bovy J., Rix H.-W., Green G. M., Schlafly E. F., Finkbeiner D. P., 2016, ApJ, 818, 130
Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2011, MNRAS, 415, L40
Bromley B. C., Kenyon S. J., Geller M. J., Barcikowski E., Brown W. R., Kurtz M. J., 2006, ApJ, 653, 1194
Bromley B. C., Kenyon S. J., Brown W. R., Geller M. J., 2009, ApJ, 706, 925
Bromley B. C., Kenyon S. J., Geller M. J., Brown W. R., 2012, ApJ, 749, L42
Bromley B. C., Kenyon S. J., Brown W. R., Geller M. J., 2018, ApJ, 868, 25
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
Brown W. R., Lattanzi M. G., Kenyon S. J., Geller M. J., 2018, ApJ, 866, 39
Bullock J. S., 2002, in Natarajan P., ed., The Shapes of Galaxies and their Dark Halos. pp 109–113 (arXiv:astro-ph/0106380), doi:10.1142/9789812778017_0018
Bullock J. S., Stewart K. R., Kaplinghat M., Tollerud E. J., Wolf J., 2010, ApJ, 717, 1043
Capuzzo-Dolcetta R., Fragione G., 2015, MNRAS, 454, 2677
Cardelli J. A., Clayton G. C., Mathis J. S., 1989, ApJ, 345, 245
Carollo C. M., Stiavelli M., de Zeeuw P. T., Mack J., 1997, AJ, 114, 2366
Carollo D., et al., 2010, ApJ, 712, 692
Carrasco-Davis R., et al., 2018, arXiv e-prints, p. arXiv:1807.03869
Carson D. J., Barth A. J., Seth A. C., den Brok M., Cappellari M., Greene J. E., Ho L. C., Neumayer N., 2015, AJ, 149, 170
Chiba M., Beers T. C., 2000, AJ, 119, 2843
Cignoni M., Ripepi V., Marconi M., Alcalá J. M., Capaccioli M., Pannella M., Silvotti R., 2007, A&A, 463, 975
Clements E. D., Swifte R. H. D., Alexander J. B., 1980, The Observatory, 100, 5
Contigiani O., Rossi E. M., Marchetti T., 2019, MNRAS, 487, 4025
Cropper M., et al., 2018, arXiv e-prints, p. arXiv:1804.09369
Cui X.-Q., et al., 2012, Research in Astronomy and Astrophysics, 12, 1197
Dai J.-M., Tong J., 2018, arXiv e-prints, p. arXiv:1807.10406
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
Do T., Kerzendorf W., Winsor N., Støstad M., Morris M. R., Lu J. R., Ghez A. M., 2015, ApJ, 809, 143
Dormand J., Prince P., 1980, Journal of Computational and Applied Mathematics, 6, 19–26
Drimmel R., Cabrera-Lavers A., López-Corredoira M., 2003, A&A, 409, 205

Duarte de Vasconcelos Silva M., 2012, PhD thesis, PhD Theses Collection, 2299, 8115
Duchêne G., Kraus A., 2013, ARA&A, 51, 269
Duchi J., Hazan E., Singer Y., 2011, J. Mach. Learn. Res., 12, 2121
Dunstall P. R., et al., 2015, A&A, 580, A93
Edelmann H., Napiwotzki R., Heber U., Christlieb N., Reimers D., 2005, ApJ, 634, L181
Eldridge J. J., Langer N., Tout C. A., 2011, MNRAS, 414, 3501
Erkal D., Boubert D., Gualandris A., Evans N. W., Antonini F., 2019, MNRAS, 483, 2007
Evans N. W., Sanders J. L., Williams A. A., An J., Lynden-Bell D., Dehnen W., 2016, MNRAS, 456, 4506
Event Horizon Telescope Collaboration et al., 2019, ApJ, 875, L1
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
Frank J., Rees M. J., 1976, MNRAS, 176, 633
Fritz T. K., et al., 2016, ApJ, 821, 44
Gaia Collaboration et al., 2016a, A&A, 595, A1
Gaia Collaboration et al., 2016b, A&A, 595, A2
Gaia Collaboration et al., 2018a, A&A, 616, A1
Gaia Collaboration et al., 2018b, A&A, 616, A11
García Cole A., Schuster W. J., Parrao L., Moreno E., 1999, Rev. Mexicana Astron. Astrofis., 35, 111
Geier S., et al., 2015, Science, 347, 1126
Genzel R., Eisenhauer F., Gillessen S., 2010, Reviews of Modern Physics, 82, 3121
Georgiev I. Y., Böker T., 2014, MNRAS, 441, 3570
Ghez A. M., et al., 2003, ApJ, 586, L127
Ghez A. M., Salim S., Hornstein S. D., Tanner A., Lu J. R., Morris M., Becklin E. E., Duchêne G., 2005, ApJ, 620, 744
Ghez A. M., et al., 2008, ApJ, 689, 1044
Gibbons S. L. J., Belokurov V., Evans N. W., 2014, MNRAS, 445, 3788
Gillessen S., Eisenhauer F., Trippe S., Alexander T., Genzel R., Martins F., Ott T., 2009, ApJ, 692, 1075
Gillessen S., et al., 2017, ApJ, 837, 30
Gilmore G., et al., 2012, The Messenger, 147, 25
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
Gonzalez O. A., Gadotti D., 2016, in Laurikainen E., Peletier R., Gadotti D., eds, Astrophysics and Space Science Library Vol. 418, Galactic Bulges. p. 199 (arXiv:1503.07252), doi:10.1007/978-3-319-19378-6_9
Goodman J., Weare J., 2010, Comm. App. Math. Comp. Sci., 5, 65

Gravity Collaboration et al., 2018, A&A, 615, L15
Gualandris A., Portegies Zwart S., Sipior M. S., 2005, MNRAS, 363, 223
Guedes J., Callegari S., Madau P., Mayer L., 2011, ApJ, 742, 76
Gvaramadze V. V., Gualandris A., 2011, MNRAS, 410, 304
Gvaramadze V. V., Gualandris A., Portegies Zwart S., 2009, MNRAS, 396, 570
Habibi M., et al., 2017, ApJ, 847, 120
Hattori K., Valluri M., Bell E. F., Roederer I. U., 2018a, ApJ, 866, 121
Hattori K., Valluri M., Castro N., 2018b, ApJ, 869, 33
Hawkins K., Wyse R. F. G., 2018, MNRAS, 481, 1028
Hawkins K., et al., 2015, MNRAS, 447, 2046
Haykin S., 2009, Neural Networks and Learning Machines. No. v. 10 in Neural networks and learning machines, Prentice Hall, https://books.google.nl/books?id=K7P36lKzI_QC
Heber U., Edelmann H., Napiwotzki R., Altmann M., Scholz R.-D., 2008, A&A, 483, L21
Helmi A., 2004, MNRAS, 351, 643
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
Høg E., et al., 2000, A&A, 355, L27
Holmberg J., Nordström B., Andersen J., 2007, A&A, 475, 519
Hoogerwerf R., de Bruijne J. H. J., de Zeeuw P. T., 2001, A&A, 365, 49
Hopman C., 2009, ApJ, 700, 1933
Houk N., 1978, Michigan catalogue of two-dimensional spectral types for the HD stars
Huang Y., et al., 2016, MNRAS, 463, 2623
Hunter J. D., 2007, Computing In Science & Engineering, 9, 90
Hurley J. R., Pols O. R., Tout C. A., 2000, MNRAS, 315, 543
Irrgang A., Kreuzer S., Heber U., 2018, A&A, 620, A48
Johnson D. R. H., Soderblom D. R., 1987, AJ, 93, 864
Johnston K. V., Spergel D. N., Hernquist L., 1995, ApJ, 451, 598
Jordi K., Grebel E. K., Ammon K., 2005, Astronomische Nachrichten, 326, 657
Jordi C., et al., 2010, A&A, 523, A48
Kafle P. R., Sharma S., Lewis G. F., Bland-Hawthorn J., 2014, ApJ, 794, 59
Katz D., Brown A. G. A., 2017, in Reylé C., Di Matteo P., Herpin F., Lagadec E., Lançon A., Meliani Z., Royer F., eds, SF2A-2017: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics. pp 259–263 (arXiv:1710.10816)
Katz D., et al., 2019, A&A, 622, A205
Kennedy J., Eberhart R., 1995, in Neural Networks, 1995. Proceedings., IEEE International Conference on. pp 1942–1948 vol.4, doi:10.1109/ICNN.1995.488968

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
Kenyon S. J., Bromley B. C., Brown W. R., Geller M. J., 2018, ApJ, 864, 130
Khan S., et al., 2019, in The Gaia Universe. p. 13 (arXiv:1904.05676), doi:10.5281/zenodo.2635051
Kharchenko N. V., Scholz R.-D., Piskunov A. E., Röser S., Schilbach E., 2007, Astronomische Nachrichten, 328, 889
Klypin A., Kravtsov A. V., Valenzuela O., Prada F., 1999, ApJ, 522, 82
Kobayashi S., Hainick Y., Sari R., Rossi E. M., 2012, ApJ, 748, 105
Kobulnicky H. A., et al., 2014, ApJS, 213, 34
Kollmeier J. A., Gould A., Knapp G., Beers T. C., 2009, ApJ, 697, 1543
Kollmeier J. A., et al., 2010, ApJ, 723, 812
Kordopatis G., Recio-Blanco A., de Laverny P., Bijaoui A., Hill V., Gilmore G., Wyse R. F. G., Ordenovic C., 2011a, A&A, 535, A106
Kordopatis G., et al., 2011b, A&A, 535, A107
Kordopatis G., et al., 2013a, AJ, 146, 134
Kordopatis G., et al., 2013b, MNRAS, 436, 3231
Kordopatis G., et al., 2013c, A&A, 555, A12
Kordopatis G., et al., 2015, A&A, 582, A122
Kordopatis G., Amorisco N. C., Evans N. W., Gilmore G., Koposov S. E., 2016, MNRAS, 457, 1299
Kouwenhoven M. B. N., Brown A. G. A., Portegies Zwart S. F., Kaper L., 2007, A&A, 474, 77
Kroupa P., 2001, MNRAS, 322, 231
Kroupa P., 2002, Science, 295, 82
Kunder A., et al., 2017, AJ, 153, 75
Laevens B. P. M., et al., 2015, ApJ, 813, 44
Latham D. W., Stefanik R. P., Torres G., Davis R. J., Mazeh T., Carney B. W., Laird J. B., Morse J. A., 2002, AJ, 124, 1144
Law D. R., Majewski S. R., 2010, ApJ, 714, 229
LeCun Y., 1993, in Tutorial presented at Neural Information Processing Systems. p. 49
LeCun Y. A., Bottou L., Orr G. B., Müller K.-R., 2012, Efficient BackProp. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 9–48, doi:10.1007/978-3-642-35289-8_3, http://dx.doi.org/10.1007/978-3-642-35289-8_3
Leavitt H. S., 1908, Annals of Harvard College Observatory, 60, 87
Leavitt H. S., Pickering E. C., 1912, Harvard College Observatory Circular, 173, 1
Leonard P. J. T., 1991, AJ, 101, 562
Leonard P. J. T., Duncan M. J., 1990, AJ, 99, 608
Li Y., Luo A., Zhao G., Lu Y., Ren J., Zuo F., 2012, ApJ, 744, L24
Li Y.-B., et al., 2015, Research in Astronomy and Astrophysics, 15, 1364
Lightman A. P., Shapiro S. L., 1977, ApJ, 211, 244

Lindegren L., Lammers U., Hobbs D., O'Mullane W., Bastian U., Hernández J., 2012, A&A, 538, A78
Lindegren L., et al., 2016, A&A, 595, A4
Lindegren L., et al., 2018a, https://www.cosmos.esa.int/documents/29201/1770596/Lindegren_GaiaDR2_Astrometry_extended.pdf/1ebddb25-f010-6437-cb14-0e360e2d9f09
Lindegren L., et al., 2018b, A&A, 616, A2
Loebman S. R., et al., 2014, ApJ, 794, 151
Lu J. R., Do T., Ghez A. M., Morris M. R., Yelda S., Matthews K., 2013, ApJ, 764, 155
Luri X., et al., 2014, A&A, 566, A119
Luri X., et al., 2018, A&A, 616, A9
Madigan A.-M., Pfuhl O., Levin Y., Gillessen S., Genzel R., Perets H. B., 2014, ApJ, 784, 23
Magrini L., et al., 2017, A&A, 603, A2
Maiolino R., et al., 2017, Nature, 544, 202
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., Rossi E. M., Brown A. G. A., 2018a, MNRAS, p. 2466
Marchetti T., Contigiani O., Rossi E. M., Albert J. G., Brown A. G. A., Sesana A., 2018b, MNRAS, 476, 4697
Martell S. L., et al., 2017, MNRAS, 465, 3203
Matthews B., 1975, Biochimica et Biophysica Acta (BBA) - Protein Structure, 405, 442
McMillan P. J., 2017, MNRAS, 465, 76
McWilliam A., Zoccali M., 2010, ApJ, 724, 1491
Meyer L., et al., 2012, Science, 338, 84
Michalik D., Lindegren L., Hobbs D., 2015, A&A, 574, A115
Miyamoto M., Nagai R., 1975, PASJ, 27, 533
Moe M., Di Stefano R., 2017, ApJS, 230, 15
Monari G., et al., 2018, A&A, 616, L9
Monson A. J., et al., 2017, AJ, 153, 96
Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, ApJ, 524, L19
Morris M., 1993, ApJ, 408, 496
Muno M. P., Pfahl E., Baganoff F. K., Brandt W. N., Ghez A., Lu J., Morris M. R., 2005, ApJ, 622, L113
Natali F., Natali G., Pompei E., Pedichini F., 1994, A&A, 289, 756
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
O'Leary R. M., Loeb A., 2008, MNRAS, 383, 86
Oort J. H., 1927, Bull. Astron. Inst. Netherlands, 3, 275
Öpik E., 1924, Publications of the Tartu Astrofizica Observatory, 25

Ott T., Eckart A., Genzel R., 1999, ApJ, 523, 248
Palladino L. E., Schlesinger K. J., Holley-Bockelmann K., Allende Prieto C., Beers T. C., Lee Y. S., Schneider D. P., 2014, ApJ, 780, 7
Paumard T., et al., 2006, ApJ, 643, 1011
Pereira C. B., Jilinski E., Drake N. A., de Castro D. B., Ortega V. G., Chavero C., Roig F., 2012, A&A, 543, A58
Perets H. B., Šubr L., 2012, ApJ, 751, 133
Perets H. B., Hopman C., Alexander T., 2007, ApJ, 656, 709
Perets H. B., Wu X., Zhao H. S., Famaey B., Gentile G., Alexander T., 2009, ApJ, 697, 2096
Peters P. C., 1964, Physical Review, 136, 1224
Pfuhl O., et al., 2011, ApJ, 741, 108
Pfuhl O., Alexander T., Gillessen S., Martins F., Genzel R., Eisenhauer F., Fritz T. K., Ott T., 2014, ApJ, 782, 101
Piffl T., et al., 2014, A&A, 562, A91
Planck Collaboration et al., 2016, A&A, 594, A13
Portail M., Wegg C., Gerhard O., Martinez-Valpuesta I., 2015, MNRAS, 448, 713
Portegies Zwart S. F., 2000, ApJ, 544, 437
Portegies Zwart S. F., Verbunt F., 1996, A&A, 309, 179
Portegies Zwart S., et al., 2009, New A, 14, 369
Posti L., Helmi A., 2019, A&A, 621, A56
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
Price-Whelan A. M., Hogg D. W., Johnston K. V., Hendel D., 2014, ApJ, 794, 4
Przybilla N., Fernanda Nieva M., Heber U., Butler K., 2008, ApJ, 684, L103
Przybylski A., 1967, MNRAS, 136, 185
Przybylski A., 1978, PASP, 90, 451
Quinlan G. D., 1996, New A, 1, 35
Randich S., Gilmore G., Gaia-ESO Consortium 2013, The Messenger, 154, 47
Rasskazov A., Fragione G., Leigh N. W. C., Tagawa H., Sesana A., Price-Whelan A., Rossi E. M., 2019, ApJ, 878, 17
Reid M. J., Menten K. M., Zheng X. W., Brunthaler A., Xu Y., 2009, ApJ, 705, 1548
Renzo M., et al., 2019, A&A, 624, A66
Rimoldi A., Portegies Zwart S., Rossi E. M., 2016, Computational Astrophysics and Cosmology, 3, 2
Rix H.-W., Bovy J., 2013, A&A Rev., 21, 61
Robbins H., Monro S., 1951, Ann. Math. Statist., 22, 400
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
Saerens M., Latinne P., Decaestecker C., 2002, IEEE Trans. Neural Networks, 13, 1204
Sana H., et al., 2012, Science, 337, 444
Sana H., et al., 2013, A&A, 550, A107
Sari R., Kobayashi S., Rossi E. M., 2010, ApJ, 708, 605
Schaller M., et al., 2015, MNRAS, 451, 1247
Schaye J., et al., 2015, MNRAS, 446, 521
Schlafly E. F., Finkbeiner D. P., 2011, ApJ, 737, 103
Schödel R., Feldmeier A., Neumayer N., Meyer L., Yelda S., 2014a, Classical and Quantum Gravity, 31, 244007
Schödel R., Feldmeier A., Kunneriath D., Stolovy S., Neumayer N., Amaro-Seoane P., Nishiyama S., 2014b, A&A, 566, A47
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
Sesar B., Fouesneau M., Price-Whelan A. M., Bailer-Jones C. A. L., Gould A., Rix H.-W., 2017, ApJ, 838, 107
Silk J., Antonuccio-Delogu V., Dubois Y., Gaibler V., Haas M. R., Khochfar S., Krause M., 2012, A&A, 545, L11
Silva M. D. V., Napiwotzki R., 2011, MNRAS, 411, 2596
Singh B., De S., Zhang Y., Goldstein T., Taylor G., 2015, CoRR, abs/1510.04609
Smith M. C., et al., 2007, MNRAS, 379, 755
Smith M. C., et al., 2009, MNRAS, 399, 1223
Soubiran C., Jasniewicz G., Chemin L., Crifo F., Udry S., Hestroffer D., Katz D., 2013, A&A, 552, A64
Stivaktakis R., Tsagkatakis G., Moraes B., Abdalla F., Starck J.-L., Tsakalides P., 2018, arXiv e-prints, p. arXiv:1809.09622
Tauris T. M., 2015, MNRAS, 448, L6
Tauris T. M., Takens R. J., 1998, A&A, 330, 1047
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
Tody D., 1986, in Crawford D. L., ed., Proc. SPIEVol. 627, Instrumentation in astronomy VI. p. 733, doi:10.1117/12.968154
Venn K. A., Irwin M., Shetrone M. D., Tout C. A., Hill V., Tolstoy E., 2004, AJ, 128, 1177
Vera-Ciro C., Helmi A., 2013, ApJ, 773, L4
Vera-Ciro C. A., Helmi A., Starkenburg E., Breddels M. A., 2013, MNRAS, 428, 1696

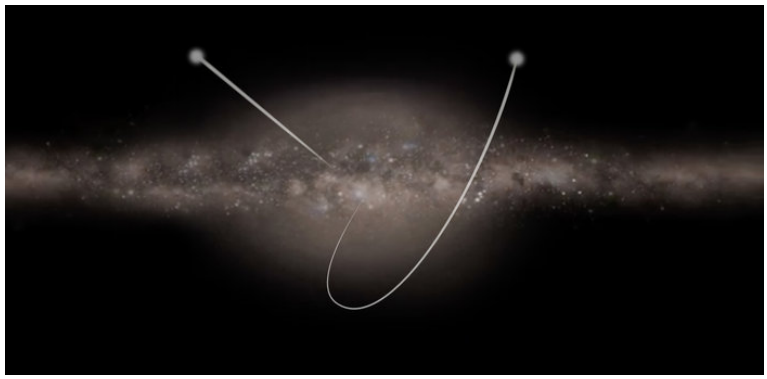
Vickers J. J., Smith M. C., Grebel E. K., 2015, AJ, 150, 77
Wang W., Han J., Cooper A. P., Cole S., Frenk C., Lowing B., 2015, MNRAS, 453, 377
Wegg C., Gerhard O., 2013, MNRAS, 435, 1874
Westera P., Buser R., 2003, in Piotto G., Meylan G., Djorgovski S. G., Riello M., eds, Astronomical Society of the Pacific Conference Series Vol. 296, New Horizons in Globular Cluster Astronomy. p. 238
Williams A. A., Evans N. W., 2015, MNRAS, 454, 698
Williams A. A., Belokurov V., Casey A. R., Evans N. W., 2017, MNRAS, 468, 2359
Xu Y., Newberg H. J., Carlin J. L., Liu C., Deng L., Li J., Schönrich R., Yanny B., 2015, ApJ, 801, 105
Yu Q., Madau P., 2007, MNRAS, 379, 1293
Yu Q., Tremaine S., 2003, ApJ, 599, 1129
Zasowski G., et al., 2013, AJ, 146, 81
Zhang F., Lu Y., Yu Q., 2013, ApJ, 768, 153
Zhang Y., Smith M. C., Carlin J. L., 2016, ApJ, 832, 10
Zheng Z., et al., 2014, ApJ, 785, L23
Ziegerer E., Volkert M., Heber U., Irrgang A., Gänsicke B. T., Geier S., 2015, A&A, 576, L14
Ziegerer E., Heber U., Geier S., Irrgang A., Kupfer T., Fürst F., Schaffenroth J., 2017, A&A, 601, A58
Zinn J. C., Pinsonneault M. H., Huber D., Stello D., 2019, ApJ, 878, 136
Zubovas K., Nayakshin S., Sazonov S., Sunyaev R., 2013, MNRAS, 431, 793
Zucker S., 2003, MNRAS, 342, 1291
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

Nederlandse Samenvatting

De zon is slechts één van de honderden miljarden sterren in het sterrenstelsel waarin wij wonen; de Melkweg. Hoewel het overgrote deel van deze sterren rondom het galactisch centrum beweegt, bestaat er ook een klein aantal wat deze draaiing niet volgt, en in plaats daarvan met een buitengewoon hoge snelheid dwars door de Melkweg reist. De snelste leden van deze groep sterren worden zogeheten *hypervelocity* sterren genoemd, en reizen met snelheden van duizenden kilometers per seconde door ons sterrenstelsel. Dit is een aantal miljoen kilometer per uur, meer dan zesduizendmaal sneller dan de snelste trein op Aarde!

We denken dat deze sterren een dusdanig hoge snelheid hebben omdat ze uit het galactisch centrum afkomstig zijn. Oorspronkelijk waren ze onderdeel van een dubbelster-systeem, wat bestaat uit twee sterren die om elkaar heen draaien. Het galactisch centrum herbergt het zwaarste object in de gehele Melkweg, Sagittarius A*, een zwart gat wat meer dan vier miljoen keer zo zwaar is als de zon. De zwaartekracht-wisselwerking met een dusdanig zwaar object kan een dubbelster volledig uit elkaar trekken, en de sterren hierbij voorgoed scheiden. Één van de twee sterren zal hierna in een baan rond het zwarte gat terechtkomen, terwijl de andere – de hypervelocity ster – wordt weggeschoten met een enorme snelheid. Deze snelheid is zodanig dat hypervelocity sterren de zwaartekracht van de gehele Melkweg volledig overwinnen, en er voor eeuwig vanaf zullen bewegen. Afbeelding S.1 laat een artist impression van uit het Melkwegcentrum weggeschoten hypervelocity sterren zien.

Hypervelocity sterren zijn interessant omdat ze kunnen worden gebruikt om meer te weten te komen over de verschillende onderdelen van de Melkweg. Het galactisch centrum is moeilijk om direct waar te nemen omdat het wordt afgeschermd door interstellair stof, zodat het ontdekken van hypervelocity sterren ver van het zwarte gat ons kan leren over hun geboorteplaats. Op een compleet andere schaal kunnen hypervelocity sterren ons



Figuur S.1: Een artist impression van hyperveloctity sterren weggeschoten uit het centrum van de Melkweg. Bron: ESA.

ook inzicht geven in de halo van donkere materie waarin de Melkweg zich bevindt. Donkere materie is een bepaalde soort materie die geen wisselwerking met elektromagnetische straling vertoont, wat inhoudt dat het voor onze ogen en telescopen volledig onzichtbaar is. De halo van donkere materie is dusdanig zwaar dat het de baan van de hypervelocity sterren doet buigen, zodat deze kunnen worden gebruikt om fundamentele parameters van de halo te bepalen, zoals de vorm en massa. Daarnaast kunnen hypervelocity sterren ons in het algemeen meer inzicht geven in de aard van deze raadselachtige donkere materie. Tot op heden is slechts een klein aantal hypervelocity sterren geïdentificeerd, maar door de komst van *Gaia*, de nieuwe satelliet van de Europese Ruimtevaartsorganisatie (*European Space Agency*, ESA) zit onze kennis van de snelste sterren in de Melkweg momenteel in de lift. Het doel van *Gaia* is om de gehele geschiedenis van de Melkweg in kaart te brengen door middel van het produceren van de grootste en meest precieze catalogus van sterren tot nu toe, met daarin posities, afstanden en geprojecteerde snelheden van meer van een miljard sterren.

Dit Werk

Het doel van deze proefschrift is het vinden en karakteriseren van de verzameling snelste sterren in de Melkweg, en om te laten zien dat deze objecten gebruikt kunnen worden gebruikt om verschillende onderdelen van ons sterrenstelsel te onderzoeken. Hiervoor maken we gebruik van verscheidene datamining technieken; astrometrische, photometrische en spectrosco-

pische datasets; astronomische waarnemingen, en theoretische modellen. In het bijzonder streeft deze proefschrift naar het beantwoorden van de volgende vier vragen:

- **Hoofdstuk 2:** Hoeveel hypervelocity sterren verwachten we te vinden in de *Gaia* catalogus?
- **Hoofdstuk 3:** Kunnen we deze hypervelocity sterren in de eerste *data release* van *Gaia* vinden?
- **Hoofdstuk 4:** Wat is de oorsprong van de snelste sterren in de tweede *Gaia* data release?
- **Hoofdstuk 5:** Wat kunnen we leren over het Melkwegcentrum en de halo van donkere materie met de verzameling ontdekte hypervelocity sterren?

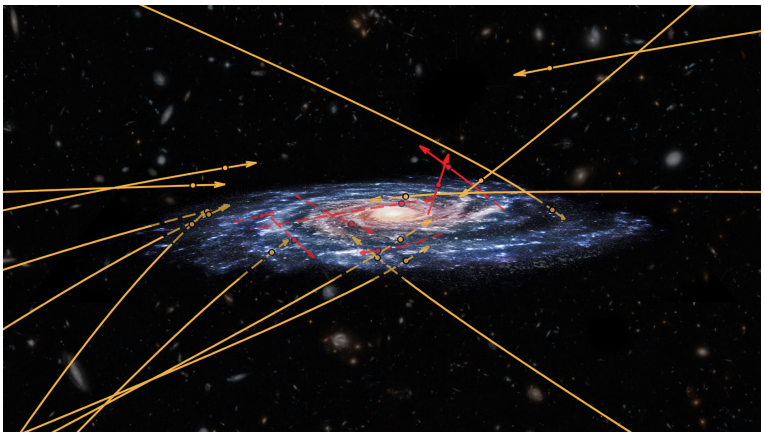
In wat volgt, presenteren we een korte samenvatting van elk van de bovengenoemde hoofdstukken, waarin we deze vragen behandelen.

De eerste vraag die we willen beantwoorden, betreft hoeveel hypervelocity sterren we verwachten aan te treffen in de door *Gaia* geproduceerde catalogus van sterren. Hiervoor hebben we in **Hoofdstuk 2** een catalogus met gesimuleerde hypervelocity sterren ontwikkeld, om de verwachte verzameling te kwantificeren en te karakteriseren. We veronderstellen drie mogelijke mechanismen die hypervelocity sterren hun extreme snelheden mee kunnen geven: i) we bevolken de hele Melkweg met ongebonden sterren op radiële banen vanuit het Melkwegcentrum, ii) we nemen aan dat hypervelocity sterren het resultaat zijn van de zwaartekracht-wisselwerking tussen een dubbelster-systeem en het massieve zwarte gat in het centrum van ons sterrenstelsel (het zogeheten *Hills mechanisme*) en iii) we modelleren het verval van de omwentelingsbaan in een systeem van twee zwarte gaten die, bij het kruisen van de baan van individuele sterren, deze hierbij met grote snelheid afstoten. We maken gebruik van simpele modellen van ster-evolutie om de schijnbare magnitude van elke hypervelocity ster te bepalen, zodat we een schatting verkrijgen van de nauwkeurigheid waarmee *Gaia* de astrometrische parameters van de ster kan bepalen. In elk van de scenario's vinden we zeer bemoedigende resultaten: we verwachten honderden tot duizenden nauwkeurig gemeten hypervelocity sterren in de laatste data release van *Gaia* aan te treffen, hoewel het grootste deel hiervan niet helder genoeg zal zijn om ook een radiële snelheidsmeting van

Gaia te hebben.

In **Hoofdstuk 3** ontwikkelen we een nieuwe datamining techniek, waarbij we gebruik maken van kunstmatige neurale netwerken, om hypervelocity sterren in de eerste data release van *Gaia* te identificeren. We ontwikkelen hiervoor een *binary classification* algoritme dat, gebruikmakend van de positie, parallax en eigenbeweging van een ster (dus zonder een radiële snelheidsmeting), een enkele waarde produceert die we interpreteren als de kans dat de ster een hypervelocity ster is. We trainen dit algoritme op zowel de gesimuleerde verzameling hypervelocity sterren uit Hoofdstuk 1 als een nagebootste *Gaia* catalogus. Het toepassen van dit algoritme op de data leidt tot een verzameling van 80 sterren met een grote kans een hypervelocity ster te zijn. Om de betrouwbaarheid van deze kandidaat-hypervelocity sterren vast te stellen, nemen we een subset van de sterren waar met de Isaac Newton telescoop in La Palma, één van de Canarische Eilanden, zodat we deze radiële snelheden en afstanden kunnen toekennen, en de stereigenschappen (zoals massa, temperatuur, leeftijd en metalliciteit) kunnen bepalen. Deze spectroscopische waarnemingen bevestigen dat de datamining techniek werkt, en dat deze succesvol is in het ontdekken van snelle sterren: we vinden een zestal sterren die mogelijkerwijs vanuit het centrum van de Melkweg zijn weggeschoten.

In **Hoofdstuk 4** maken we gebruik van de nauwkeurige waarnemingen in de tweede data release van *Gaia* om de hoge-snelheids uitschieters van de meer dan 7 miljoen sterren in deze catalogus te karakteriseren. We bepalen de afstanden en totale snelheden van alle sterren in deze set, en vinden hierin 20 sterren met dusdanig hoge snelheden dat ze niet langer gebonden zijn aan de Melkweg. We zorgen er in het bijzonder voor dat onechte detecties en instrumentele fouten worden weggefilterd, aangezien deze potentieel hypervelocity sterren kunnen nabootsen. We gebruiken de nauwkeurige waarnemingen van *Gaia* om de baan van deze sterren door de Melkweg terug te rekenen, om op deze manier hun geboorteplaats te achterhalen. Een aantal van deze sterren lijkt afkomstig te zijn vanuit de galactische schijf van de Melkweg. Verrassend genoeg lijkt de meerderheid van de sterren echter niet afkomstig uit één van de bekende stervormingsgebieden in ons sterrenstelsel, wat een extragalactische oorsprong suggereert. Zulke sterren zijn voorheen voorspeld in numerieke simulaties die de zwaartekracht-



Figuur S.2: De banen van de ontdekte ongebonden sterren, weggeschoten vanuit de Melkweg (rood) of van extragalactische oorsprong (geel). De achtergrond stelt een artist impression van de Melkweg voor. Bron: ESA / NASA / Hubble / Marchetti et al. 2018.

interactie tussen de Melkweg en kleinere sterrenstelsels modelleren. Een artist impression van deze 'extragalactische indringers' is te zien in Afbeelding S.2.

Ten slotte heeft **Hoofdstuk 5** als doel om de kracht van hypervelocity sterren als meetinstrument voor de eigenschappen van dubbelsterren in het Melkwegcentrum, alsmede de karakteristieke parameters van de halo van donkere materie, te toetsen. Door middel van een statistische aanpak vergelijken we de snelheidsdistributie van de verzameling van ~ 20 ongebonden hypervelocity sterren met analytische voorspellingen van het Hills mechanisme. We concluderen hieruit dat het huidig aantal bekende hypervelocity sterren onvoldoende is om de fysische eigenschappen van de eerdergenoemde onderdelen van de Melkweg nauwkeurig te bepalen. Onder de aanname dat dubbelster-systemen in het galactisch centrum soortgelijke eigenschappen vertonen als deze in andere stervormingsgebieden, vinden we overeenstemming tussen onze data en modellen van donkere materie halo's, afkomstig van voorspellingen uit kosmologische simulaties. De duizenden hypervelocity sterren die aanwezig zijn in de waarnemingen van *Gaia* zijn cruciaal om deze nieuwe en enerverende methode voor het bestuderen van het sterrenstelsel waarin wij wonen te onthullen.

English Summary

The Sun is just one star among the hundreds of billions living in our Galaxy, the Milky Way. While most of these stars rotate around the Galactic centre on almost circular orbits, a few others do not follow this motion, but move through the Galaxy with a surprising high speed. The fastest of these stars are known as *hypervelocity stars*, and travel through the Galaxy with velocities of thousands of kilometers per second. This corresponds to a few millions of kilometers per hour, more than six thousand times faster than the fastest train on Earth!

The reason why we think these stars have such an incredibly high velocity is that they come from the centre of our Galaxy. These stars were originally part of a binary system: two stars orbiting around each other. The centre of our Galaxy is the residence of the most massive single object in the Milky Way, Sagittarius A*, a black hole with a total mass of more than four million times the one of our Sun. The interaction with such an incredible massive object can break the binary system, separating the two stars forever. One of the two will start orbiting around Sagittarius A*, while the other one, the hypervelocity star, will be ejected with an incredibly high velocity. This velocity is so high that these stars do not feel anymore the gravitational pull of the Galaxy, but fly away forever from it. Figure S.3 shows an artistic impression of hypervelocity stars flying away from the centre of the Milky Way.

The reason why hypervelocity stars are interesting is that they can be used to gain knowledge on different environments of our Galaxy. The Galactic centre is very difficult to observe because of interstellar dust, so detecting hypervelocity stars far from it can tell us something on how stars form and interact in the vicinity of the massive black hole. Moving to a completely different scale, the Galaxy is embedded into a vast halo composed of dark matter: a particular kind of matter that does not interact with the electromagnetic radiation, and is thus invisible to our eyes (and

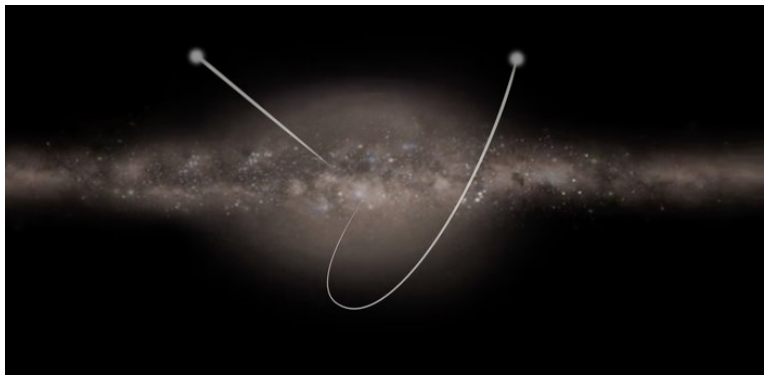


Figure S.3: Artistic impression of hypervelocity stars ejected from the centre of the Milky Way. Image credits: ESA.

telescopes). The dark matter halo is so massive that it bends the trajectories of hypervelocity stars, so that these stars can be used to determine some of its fundamental parameters (such as mass and shape) and investigate on the puzzling nature of this invisible component. Until now, only a few hypervelocity stars have been identified, but the advent of the European Space Agency (ESA) satellite *Gaia* is currently revolutionizing our knowledge on the fastest stars in the Galaxy. The aim of *Gaia* is to reconstruct the evolutionary history of the Milky Way by providing the largest and most precise stellar catalogue ever produced: positions, distances, and projected velocities for more than one billion stars.

This work

The goal of this thesis is to search for and characterize the population of the fastest stars in our Galaxy, and to show how these incredible objects can be used to probe different Galactic environments. To do that, we make use of data mining techniques, astrometric, photometric and spectroscopic datasets, observations, and theoretical modelling. In particular, this thesis aims at answering these four questions:

- **Chapter 2:** How many hypervelocity stars are we expecting to find in the *Gaia* catalogue?
- **Chapter 3:** Can we find any hypervelocity star candidates in the first *Gaia* data release?

- **Chapter 4:** What is the origin of the fastest stars in the second *Gaia* data release?

- **Chapter 5:** What can we learn about the centre of the Milky Way and its dark matter halo using the known sample of hypervelocity stars?

In the following we will present a short summary of the content of each scientific chapter, illustrating how we answer each of the questions above.

The first question we want to answer concerns how many hypervelocity stars we expect to find in the stellar catalogue provided by the *Gaia* satellite. To do so, in **Chapter 2** we create simulated catalogues of hypervelocity stars, to quantify and characterize the predicted population. We make three different assumptions on the ejection mechanism responsible for their extreme velocities: i) we populate the whole Galaxy with unbound stars on radial orbits from the Galactic Centre, ii) we assume hypervelocity stars to be the result of the interaction between a binary star and the massive black hole in the Galactic Centre (the *Hills mechanism*), and iii) we model the orbital decay of a massive black hole binary, ejecting single stars interacting with it. We use simple stellar evolution prescriptions to derive the apparent magnitude of each hypervelocity star. This allows us to estimate the error with which *Gaia* will measure its astrometric parameters. In all cases, our predictions are extremely encouraging: we find hundreds to thousands of precisely measured hypervelocity stars to be contained in the final *Gaia* catalogue, but the majority of these stars will not be bright enough to have a radial velocity determination from *Gaia*.

In **Chapter 3** we introduce and develop a novel data mining technique, based on artificial neural networks, to identify hypervelocity stars in the first data release of the *Gaia* satellite. We create a binary classifier algorithm which, taking in input the position, parallax, and proper motions of a star (no radial velocity), outputs a single real number which we can interpret as the probability of the star being a hypervelocity star. The algorithm is trained on mock populations built in Chapter 1, and on a simulated *Gaia* catalogue. The application to the data results in 80 stars with a high probability to be hypervelocity stars. To confirm the goodness of our candidates, we observe a subset of stars at the Isaac Newton Telescope in La Palma, Canary Islands, deriving radial velocities, distances and stellar parameters (mass, temperature, age and metallicity). The spectroscopic observations confirm the working of the data mining routine, which succeeded in find-

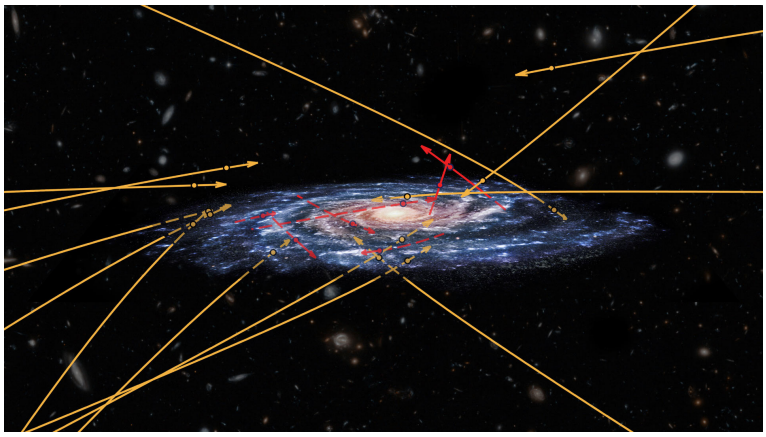


Figure S.4: Past orbits of runaway (red) and extragalactic (yellow) unbound candidates, on top of an artistic impression of the Milky Way. Image credits: ESA / NASA / Hubble / Marchetti et al. 2018.

ing high velocity stars: we report the discovery of 6 stars might be ejected from the centre of our Galaxy.

In **Chapter 4** we use the precise data provided by the second data release of *Gaia* to characterize the high velocity tail of the velocity distribution of more than 7 million stars in the Milky Way. We derive distances and total velocities for all of the stars in the sample, and we are able to discover a sample of 20 stars with unbound velocities. Particular care is taken to filter out spurious measurements and instrumental artifacts, which might mimic high velocity stars. We use the precise *Gaia* data to reconstruct the past trajectories of these stars in the Galaxy, to identify their birth place. Some of these are consistent with coming from the stellar disk of the Milky Way. Surprisingly, the remaining majority of stars is not consistent with coming from any known Galactic star forming region, suggesting an extragalactic origin. These stars were previously predicted in numerical simulation following the gravitational interaction of other small galaxies with our Milky Way. An artistic impression of these intergalactic interlopers is shown in Fig. S.4.

Finally, the aim of **Chapter 5** is to quantify the power of known hypervelocity stars to constrain the binary properties in the Galactic Centre and the characteristic parameters of the dark matter halo. Using a statistical ap-

proach, we compare the velocity distribution of the sample of ~ 20 unbound hypervelocity stars with analytic predictions assuming the Hills mechanism. We find that the number of known hypervelocity stars is not sufficient to give tight constraints on both these environments. Assuming that binaries in the Galactic Centre have similar properties to binaries in other star forming regions, we find a good match between the data and the model for haloes consistent with predictions from cosmological structure formation simulations. The sample of thousands of hypervelocity stars in the *Gaia* catalogue will be crucial to unveil the power of this new, exciting tool to study the Galaxy we are living in.

List of Publications

First author

Marchetti, T.; Rossi, E. M.; Brown, A. G. A.; *Gaia DR2 in 6D: Searching for the fastest stars in the Galaxy*, 2018, MNRAS.

Marchetti, T.; Contigiani, O.; Rossi, E. M.; Albert, J. G.; Brown, A. G. A.; Sesana, A.; *Predicting the hypervelocity star population in Gaia*, 2018, MNRAS, 476, 4697-4712.

Marchetti, T.; Rossi, E. M.; Kordopatis, G.; Brown, A. G. A.; Rimoldi, A.; Starkenburg, E.; Youakim, K.; Ashley, R.; *An artificial neural network to discover hypervelocity stars: candidates in Gaia DR1/TGAS*, 2017, MNRAS, 470, 1388-1403.

Contributing author

Contigiani, O.; Rossi, E. M.; **Marchetti, T.**; *On measuring the Galactic dark matter halo with hypervelocity stars*, 2019, MNRAS.

Rossi, E. M.; **Marchetti, T.**; Cacciato, M.; Kuiack, M.; Sari, R.; *Joint constraints on the Galactic dark matter halo and Galactic Centre from hypervelocity stars*, 2017, MNRAS, 467, 1844-1856.

Coppi, G.; **Marchetti, T.**; de Bernardis, P.; Masi, S.; *Measurements of the polarization properties of foam materials useful for mm-wave polarimeters windows*, 2016, JIMTW, 37, 8, 815-824.

Conference proceedings

Marchetti, T.; Rossi, E. M.; Kordopatis, G.; Brown, A. G. A.; Rimoldi, A.; Starkenburg, E.; Youakim, K.; Ashley, R.; *Hypervelocity star candidates in Gaia DR1/TGAS*, 2018, Astrometry and Astrophysics in the Gaia sky, Proceedings of the International Astronomical Union, IAU Symposium, Volume 330, pp. 181-184.

Legg, S.; Lamagna, L.; Coppi, G.; de Bernardis, P.; Giuliani, G. M.; Gualtieri, R.; **Marchetti, T.**; Masi, S.; Pisano, G.; Maffei, B.; *Development of the multi-mode horn-lens configuration for the LSPE-SWIPE B-mode experiment*, 2016, proceedings of the SPIE.

Lamagna, L.; Coppi, G.; de Bernardis, P.; Giuliani, G. M.; Gualtieri, R.; Legg, S.; Maffei, B.; **Marchetti, T.**; Masi, S.; Pisano, G.; *Development of the multi-moded pixels for the LSPE/SWIPE focal plane*, 2015, proceedings for the 36th Antenna Workshop on Antennas and RF Systems for Space Science.

Curriculum Vitae

I was born on the 26th of December 1991 in Rome, Italy. I would be lying if I said that I have always been fascinated by astronomy. As a kid, my true passion was dinosaurs, and my dream was to become a paleontologist. Moving from elementary to middle school I started being more and more attracted to scientific subjects, with a particular interest in maths. This led me to take the decision to enroll in a scientific high school, the Liceo Scientifico "Isaac Newton" in the centre of Rome. This school was part of a project called PNI (literally national plan for computer technology, *Piano Nazionale Informatica* in Italian), a special program with mandatory classes of physics for all the 5 years. Nevertheless, for a few years, I seriously considered the idea of enrolling in history at university, to become an archaeologist. It was only in the fourth year of high school that I decided to focus my interest and future career on another type of archaeology: astrophysics. The idea came to my mind unexpectedly in 2009 while visiting an exhibition called "Astri e Particelle" (Stars and Particles) at "Palazzo delle Esposizioni" in Rome. Being fascinated by the possibility to study stars and distant astrophysical processes, I attended a two weeks school in astrophysics at the IAPS (Institute for astrophysics and planetology in Rome). This helped me make the final decision to enroll for the Bachelor in Physics and Astrophysics at the University "La Sapienza" in Rome. Fascinated by cosmology and the recent preliminary results from the Planck satellite, my Bachelor thesis focused on relativistic Doppler effects on the cosmic microwave background, under the supervision of Prof. Paolo de Bernardis. A natural step was then to continue my study in this field, enrolling for the Master in Astronomy and Astrophysics in the same university, where I chose a curriculum heavily based on theoretical and observational cosmology. For my thesis, I worked again in the research group led by Prof. Paolo de Bernardis on the optimization of optical components for the polarimeter SWIPE on board of the balloon-borne experiment LSPE. The work done in the G31 laboratory

allowed me to learn more about instrumentation and experimental cosmology. During the thesis, I had the chance to work and be familiar with cryostats, microwave antennas, and low-temperature detectors. My work on the polarization properties of foam materials for polarimeters windows resulted in my first scientific publication. As much as I enjoyed the technical and experimental work, I decided that I wanted to focus more on the astrophysical side, and thus to change completely topic for my scientific career.

In October 2015 I started a position as a PhD candidate at Leiden Observatory, working under the supervision of Dr. Elena Maria Rossi on the search for hypervelocity stars in the *Gaia* catalogue. This allowed me to familiarize with big data, machine learning, spectroscopic observations, and astrometric data handling. In the four years spent in Leiden, I had the chance to travel around the world, to present my results at international conferences and visit collaborators in several institutes. I was lucky enough to travel to China, France, Germany, Greece, Hungary, UK and USA, an experience that I really enjoyed and greatly enriched my ability to efficiently communicate my scientific results to a broad audience. My works with *Gaia* data (chapters 3 and 4 of this thesis) resulted in two press releases from the European Space Agency, and had a large impact on magazines and online scientific blogs. In Leiden I have been a teaching assistant for the bachelor courses "Introduction to General Relativity and Astrophysical Applications" and "Radiative Processes". During the PhD I have also been official co-supervisor for 8 Master students at Leiden Observatory and 2 high school students in Amsterdam, an experience that has greatly improved my ability of mentoring. In April 2019 I have been guest lecturer on Machine Learning in the master course "Numerical Recipes for Astrophysics". In the first semester of 2017 I organized, together with Eleonora Zari, PhD talks, specially meant for fellow candidates to share their scientific results with their colleagues in a friendly environment. In summer 2018 I had the privilege to volunteer to the "Discover Club" programme (UNAWE): an astronomy outreach programme for children in the emergency refugee center in Katwijk (The Netherlands).

On December 1st 2019 I will start a fellowship at the European Southern Observatory Headquarters in Garching, Germany. There I will have the chance to continue my scientific career focused on high velocity stars and to expand my expertise working with some of the most important state-of-the-art telescopes.

Acknowledgements

These four years spent at Leiden Observatory flew by, and they would have not been the same without the presence of so many nice people working there. So my first acknowledgment goes to the Leiden Sterrewacht, for being such a friendly, international, positive and inclusive environment. That was the first thing that struck me back in February 2015 when I was invited for the PhD interviews, and more than four years later I'm glad that this impression was confirmed and exceeded.

I want to deeply thank Elena, for offering me this PhD position and believing in me, at a time when my background was completely different from what would have then become my own research field. I knew nothing about the Milky Way and high velocity stars, and it is thanks to you if I started my scientific career in this fascinating field of research. Thank you for your continuous availability, your constant enthusiasm in the work, for always pushing me to give my best in all the occasions, and simply for starting such an awesome research group. Thank you Alex, Clément, Fraser, Stella and Valeria for the exciting lectures, lively discussions and interesting meetings. Thanks to Anthony Brown for being my unofficial second supervisor, your expertise has been invaluable for my project. Thanks to all the master students I had the pleasure to supervise over these years: Bart, Brendon, Esmee, Luther, Marco, Omar, Tom, Yuejia. As often in science, it has not always been a linear and simple path, but I hope that you enjoyed working together as much as I did, and that some of the skills you acquired during the project have been (or will be) useful in your future careers.

A huge thanks to all the amazing people I met at the Observatory. Christos and Jit, you two are the best friends I could meet. I am sure that even if our future paths will diverge from Leiden, things among us will never change. I will miss our nights at Lemmy's, the random last-minute holiday

to Crete, the trip to India, jamming together, playing PS4, watching movies and simply chatting about life and everything. I'm so happy I have you as my best friend(s). To Ele: from supervising students and going to conferences together to hanging out, travelling, cooking and watching Italian TV shows at home, spending time with you has always been great fun. Thank you, life in Leiden would have not been the same without you. To the (extended) AstroTelegram and AstroLunch people: Aayush, Alberto, Andrej, Clément, Daniel, Dilovan, Emanuele, Francesco, Fraser, Gabriella, Gaby & Marco, Igone & David, Kim, Kirsty, Lammim, Lorenzo, Luke, Omar, Pedro, Salvatore, Santiago, Valeria, Vincenzo, thanks for all the great dinners, lunches, brunches, nights, parties, barbecues, and all kinds of social activities. Thanks Pier, daje forte sempre. To my fellow PhD friends Alex, Hiddo (thanks for helping with the Nederlandse Samenvatting), Josh, Luis, Mieke, Stijn, Stella, for all the great memories in Kaiser Lounge and at the borrels. Thanks to the Observatory band: Christos, James, Olivier and Peter, for the fun times playing together. Finally, special thanks go to Maria Cristina, for being the curious, enthusiastic and always stimulating person that you are. Thank you for the great time spent together, from the amazing trip to Morocco to everyday life. What we did is just a taste of what's yet to come.

A special thanks to all my colleagues (and friends) I met at the University "La Sapienza" in Rome during my bachelor and master: Adriano, Gabriele, Fabrizio, Francesca M., Francesca S., Luca, Michele, Pietro, Silvia, Simone, Stefano, Valerio. Without you, I would have never been able to make it through the exams, the bureaucracy, and the infinite amount of hours spent trying to give meaning to physics.

Un abbraccio speciale ai miei amici di sempre: Bea, Betta, Cella, Fede, Francesco, Gabriele, Genny, Giulio, Johnny, Miriam, Natan, Quattro, Riccardino, Riccardone, Sarella, Silvia, Tanna, Timmy, Vitto. Non è sempre semplice rimanere in contatto vista la distanza, ma ogni volta che torno a Roma è come se non fossi mai partito. Grazie per esserci sempre. Grazie ad Alice, Anna & Bruno, Livio & Marco, Luca & Irene. Ci vediamo sempre meno di quanto vorremmo, ma certi rapporti non cambiano mai. Un grazie particolare a Daniele: se non fosse stato per quel concerto allo Strike non sarei la persona che sono oggi. Grazie di cuore. Un abbraccio enorme agli Otsunami: Alex, Lorenzo, Luca e Marco. Condividere il palco con voi è stata una delle esperienze più importanti della mia vita. Quello che abbiamo cre-

ato insieme in sala è stato (e sarà) unico ed irripetibile. In particolare, grazie Marco, per le cene, i concerti, le lunghe passeggiate e le chiacchierate al parco (a Roma e a Rotterdam), un abbraccione frate.

Infine, voglio ringraziare specialmente la mia famiglia. Grazie ai miei nonni: Dina, Elide, Lorenzo e Sergio, per tutti i meravigliosi momenti passasti insieme e per tutti quelli in cui purtroppo non sono potuto essere lì con voi. Sono sicuro che sareste stati orgogliosi. Alle mie cugine Cecilia, Elisa e Flavia. A Marco, per il suo sostegno, dagli aiuti con i compiti di fisica al liceo fino alla lettura di questa tesi. Grazie per essere sempre un punto di riferimento. A mamma, per il continuo e incondizionato supporto, per i bei viaggi che continuiamo (e continueremo) a fare insieme, per le piccole litigate che ogni tanto ci piacciono tanto, e per tutte quelle altre cose che mi ci vorrebbe una tesi intera per elencarle. A papà, vorrei tanto che tu fossi qui.