Threading the needle: constraining the stellar content and dynamics of the Galactic Centre with hypervelocity stars
Verberne, S.

Threading the needle

Constraining the stellar content and dynamics of the Galactic
Centre with hypervelocity stars

Proefschrift

ter verkrijging van
de graad van doctor aan the Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op vrijdag 9 januari 2026
klokke 11:30 uur

door

Sill Verberne
geboren te Amsterdam, Nederland
in 1998

Promotores:

        Prof. dr. E.M. Rossi

        Prof. dr. K.H. Kuijken

Co-promotor:

        Prof. dr. Sergey E. Koposov   University of Edinburgh

Promotiecommissie:

        Prof. dr. I.A.G. Snellen

        Prof. dr. H. Hoekstra

        Prof. dr. R. Schödel   Instituto de Astrofísica de Andalucía

        Prof. dr. S. Naoz   University of California, Los Angeles

        Prof. dr. Y. Levin   Columbia University

        dr. W.R. Brown   Harvard University

CONTENTS

**3   Improved constraints on Galactic Centre ejection of hyper-
velocity stars based on novel search method   59**

1 | INTRODUCTION

The omnipresent and almost unchanging night sky has been has been a source of fascination for much of human history. Some of the earliest records of astronomical observations originate from Babylonia in ancient Mesopotamia—the very birthplace of human civilization—dating back some 4000 years (Steele 2021). Astoundingly, modern astronomy can be traced back through a continuous tradition to these early observations (Evans & Friedlander 2025). The practical use of astronomy for, in particular, time-keeping and navigation have allowed it to be remain relevant during much of this time. Of course, our understanding of the night sky has transformed almost beyond recognition over the ages.

Modern astronomy perhaps started with the first astronomical observations using a telescope published in Galilei (1610). Although revolutionary at its time, observational techniques have come a long way since this first study. Through incremental improvement in our telescopes and understanding of physics we are closer now than ever to understanding our place in the Universe. This is not to say that all the mysteries of the Universe will be solved within a few years. We have discovered many types of objects and phenomena that were simply not known even one hundred years ago, and continue to discover new ones. Sometimes a mysterious phenomenon is observed that eludes a rigorous theoretical explanation for decades, while other times a theoretical prediction is put forward, only to be observed much later. This thesis deals with one such class of object, known as hypervelocity stars (HVSs), which can travel at millions of kilometres per hour through our Galaxy, which we show schematically in Fig. 1.1.
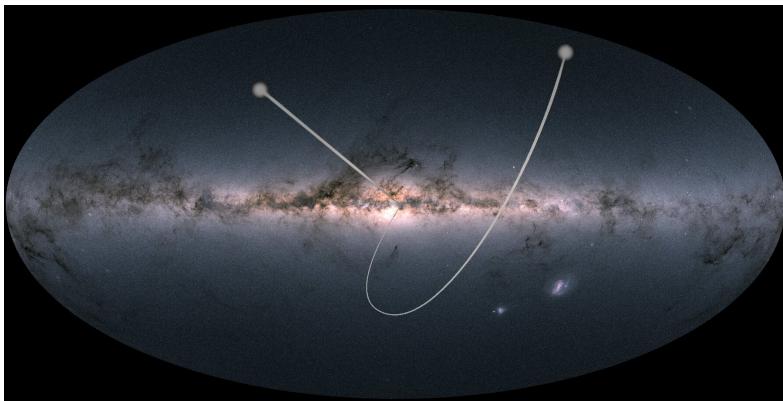
Figure 1.1: Schematic illustration of HVSs which travel through the Galaxy with extreme velocities. *Image credit: ESA + Gaia DPAC (modified).*

Already in 1988 the theoretical prediction was made that some stars should exist that are ejected from our Galaxy at extreme velocities through the interaction with a massive black hole, but it took until 2005 for the first such star to be discovered. To this day we know precious few HVSs despite significant efforts to discover more of them. This is in large part due to their extreme rarity; only about one in one hundred million stars in our Galaxy is expected to be an HVS (Verberne et al. 2024a).

HVSs are the thread throughout this thesis. We will discuss and apply various methods, combining observations and simulations, to search for these stars. Furthermore, we will discuss in depth how the limited observational knowledge of HVSs can nonetheless be used to place meaningful constraints not only on the properties of HVSs, but on the Galaxy itself.

In this introduction we will first discuss the major stellar components of the Galaxy in Section 1.1. In Section 1.2, we discuss HVSs, including how the field progressed from the first theoretical prediction to the present, the theoretical model for how these stars are ejected, a number of the science cases for the study of HVSs, and alternative mechanisms through which such stars might be produced. Section 1.3 describes the two most important astronomical surveys that are used in this thesis and covers the main instrumentation and data products. In Section 1.4 we cover future prospects of the field of HVSs. Finally, Section 1.5 provides an overview of the main scientific chapters of this thesis.

1.1   Structure and dynamics of the Milky Way

Advances over the past century have allowed us to gain a detailed understanding of our host galaxy: the Milky Way. This has allowed us to subdivide the Galaxy in a number of main components: the stellar halo, disc(s), bulge, bar, and Galactic Centre (GC). Fig. 1.2 shows a schematic overview of these components and their geometry. In this section, we will go over the main components of the Galaxy from the outside in.

1.1.1   Stellar halo

The classical view of the stellar halo is that it is a roughly spherical and isotropically distributed population of old, metal-poor stars that surrounds the Galaxy. We now know that is is not an accurate view of the stellar halo.

The current paradigm for galaxy formation is that galaxies grow through hierarchical mergers with satellites (Press & Schechter 1974; White & Rees 1978). These past mergers have build up a cloud of stars around the Galaxy at distances up to 100 kpc, forming the stellar halo (for recent reviews, see Helmi 2020; Deason & Belokurov 2024). The stars in the stellar halo are predominantly old and metal-poor and the atmospheres of these stars
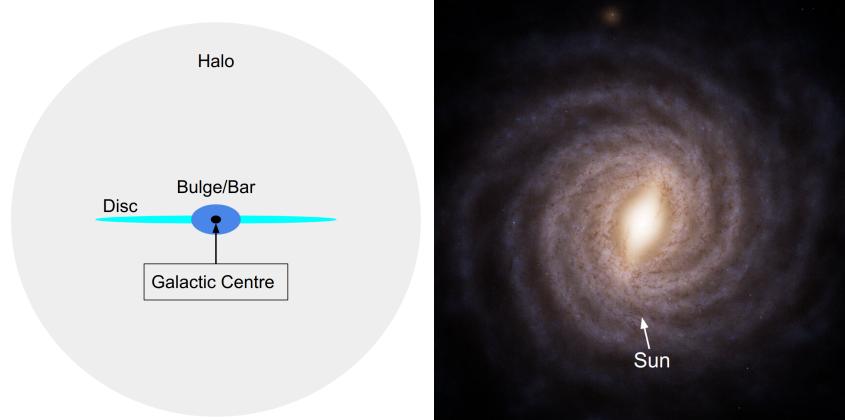
Figure 1.2: Left: side-on schematic overview of the Milky Way Galaxy (not to scale). Some of the major components of the Galaxy are labelled. Right: artist impression of the Milky Way from a top-down view (*adapted from image by: ESA/Gaia/DPAC, Stefan Payne-Wardenaar*).

contain crucial information on their satellite galaxy birth environments. These properties allow us to study the high-redshift progenitor satellites and make comparisons to observations of the distant Universe (e.g. Helmi 2020).

The stellar halo near the Sun is dominated by a single merger, which happened about 10 billion years ago, the remnant of which is known as the *Gaia* Sausage/Enceladus (GSE; Belokurov et al. 2018; Helmi et al. 2018; Haywood et al. 2018). The stars that are part of GSE are characterised by their highly elliptical orbits through the Galaxy and their relatively high metallicity for the stellar halo (mean [Fe/H] ~ -1.2; Feuillet et al. 2021). GSE is considered the last major merger of the Galaxy and has helped shape many of the current properties of the Galaxy (Naidu et al. 2021). The peculiar name of this structure is traced back to the two groups who discovered it (Belokurov et al. 2018; Helmi et al. 2018) and the shape it takes in Galactic radial against circular motion, as shown in Fig. 1.3.

The stellar halo is also marked by stars that orbit the Galaxy in coherent thin strings, known as stellar streams (for a recent review, see Bonaca & Price-Whelan 2025). These streams are formed by low mass dwarf galaxies and star clusters that merge with the Galaxy. The tidal force from the Galaxy causes some stars in the satellite to have slightly shorter or longer orbital periods. The effect is that stars escape from the satellite through the leading and trailing tidal tails (Binney & Tremaine 2008).

Furthermore, the stellar halo contains two notable dwarf galaxies called the Large and Small Magellanic clouds (LMC; SMC), which are the most
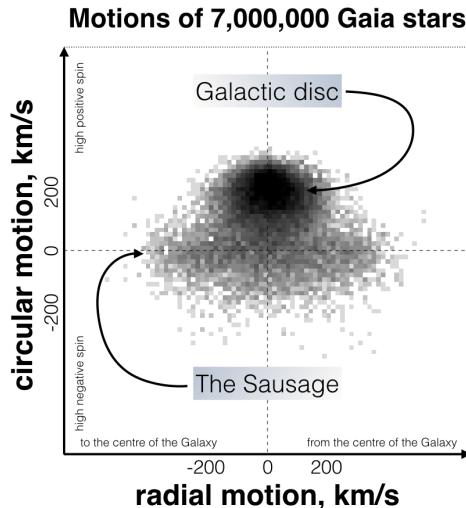
Figure 1.3: The GSE shown in Galactic radial against circular velocity. The GSE has an elongated 'sausage' shape in this space. *Image credit: V. Belokurov (Cambridge, UK and CCA, New York, US) and Gaia/ESA.*

massive Milky Way satellites. The LMC and SMC are visible by eye from the southern hemisphere and are located at about 50 and 62 kpc from the Sun respectively (e.g. Pietrzyński et al. 2019; Graczyk et al. 2020). Fig. 1.4 shows a picture of the LMC and SMC taken from Paranal, Chile. Especially the LMC is important for dynamics in the stellar halo due to its high mass. The enclosed mass of the LMC within 32.8 kpc is $\sim 7 \times 10^{10}$ M$_\odot$, while the enclosed mass of the Milky Way within 32.4 kpc is $(2.85 \pm 0.1) \times 10^{11}$ M$_\odot$ (Erkal et al. 2019; Vasiliev et al. 2020; Reino et al. 2021; Shipp et al. 2021; Koposov et al. 2023). Trajectories of stars in the stellar halo can therefore be significantly affected by the LMC and the reflex motion of the Milky Way in response to the LMC has already been measured along with the LMC induced wake (Garavito-Camargo et al. 2019; Erkal et al. 2021; Petersen & Peñarrubia 2021; Conroy et al. 2021; Vasiliev et al. 2021; Cavieres et al. 2025).

1.1.2 Stellar discs

Embedded in the stellar halo are two aligned discs of stars, which also contain our solar system: the thin and thick discs. The thin disc contains most of the stars in our Galaxy and contains younger stars than the thick disc. The naming of these two discs stems from their scale heights; the thin disc having a significantly smaller scale height compared to the thick disc.

Figure 1.4: The Magellanic clouds seen from Paranal, Chile. The LMC and SMC are the bright clouds in the middle and top of the frame respectively. *Image credit: ESO/J. Colosimo.*

The existence of the thick disc was already suggested in Gilmore & Reid (1983), but has been debated till recently (e.g. Rix & Bovy 2013). The scale heights of the thin and thick disc are about 300 and 1000 pc respectively (e.g. Jurić et al. 2008). The Sun is part of the thin disc and is located about 20.8 pc above the Galactic mid-plane (Bennett & Bovy 2019).

In addition to the scale height, the thin and thick discs are also distinct in their abundance patterns. In particular, the thick disc is enhanced in $[\alpha/\text{Fe}]$ compared to the thin disc for a given $[\text{Fe}/\text{H}]$. $[\alpha/\text{Fe}]$ indicates the abundance of α elements with respect to iron. This ratio is highly informative on the origin of stars, because α-element enrichment mainly occurs through type II supernovae (e.g. Gilmore & Wyse 1991), which are the endpoints for massive stars, while iron is formed by all supernova.

An explanation of the thick disc that has been suggested is that it contains stars from a protodisc, which was dynamically heated by the last major merger of the Galaxy that resulted in the previously discussed GSE (Belokurov et al. 2018; Xiang & Rix 2022; Chandra et al. 2024).

The most dramatic substructures of the Galactic disc are the spiral arms (see Fig. 1.2), the formation of which are still a matter of debate (Lin & Shu 1964; Toomre 1964; Wada et al. 2011; Shu 2016; Shen & Zheng 2020; Castro-Ginard et al. 2021). Furthermore, determining the spiral structure of the Milky Way is challenging, because we observe the Galaxy from within

Figure 1.5: Picture including the central region of the Milky Way as seen from La Silla, Chile. *Picture by Sill Verberne.*

(e.g. Hou et al. 2009; Hou & Han 2014). It is believed that the Galaxy has four major spiral arms and that the Sun is located in the Local Arm, which is a subsidiary that extends between two of the major spiral arms (e.g. Shen & Zheng 2020; Poggio et al. 2021).

1.1.3   Bar and bulge

Closer to the centre of our Galaxy, the spiral arms make way for the bar and bulge (see Fig. 1.2). In the central few kiloparsecs of the Galaxy, we find a boxy/peanut bulge with an X-shape that forms an out-of-plane structure (Weiland et al. 1994; Dwek et al. 1995; Binney et al. 1997; McWilliam & Zoccali 2010; Wegg & Gerhard 2013; Wegg et al. 2015; Ness & Lang 2016; Barbuy et al. 2018). This bulge component is likely related to the bar. In fact, most of the stars in the bulge might belong to the central bar, which we view nearly end-on (Shen & Zheng 2020). The bar consists of stars on elliptical orbits and rotates with a pattern speed of $\Omega_{\mathrm{bar}} = 38.1^{+2.6}_{-2.0}$ km s^{-1} (Gaia Collaboration et al. 2023b). The bar and the spiral arms are the two most notable non-axisymmetric stellar components of the Milky Way.

1.1.4   Galactic Centre

At the very centre of our Galaxy we find a complex region where many stellar populations overlap. Fig. 1.5 shows an image of the GC region and surrounding area taken from Chile with a phone camera. At optical wavelengths such as shown in the image, the Galactic Centre (GC) itself is hidden behind a strong extinction layer ($A_{\mathrm{V}} \sim 50$; Fritz et al. 2011). For this reason,

Nuclear star cluster
~ 5 pc

Sgr A*
sphere of influence
~ 3 pc

Clockwise disk
~ 0.5 pc

S-star cluster
~ 0.01 pc

Binary tidal
disruption
~ 100 AU

Stellar tidal
disruption
~ 1 AU
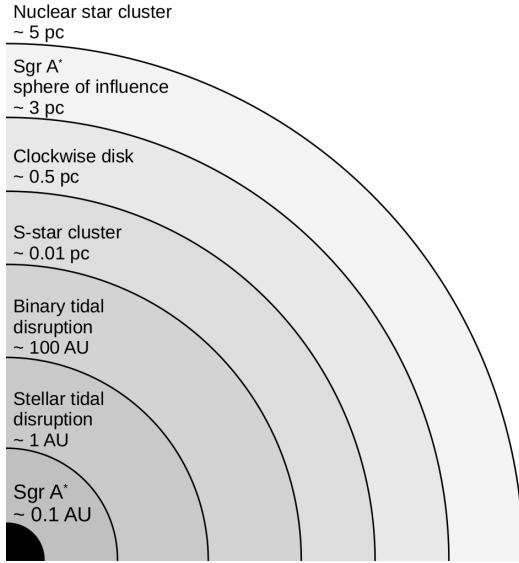
Sgr A*
~ 0.1 AU

Figure 1.6: Schematic view of the different structures in the central parsecs and their distance from Sgr A*.

the GC has often been studied at lower frequencies (e.g. Ghez et al. 1998; Genzel et al. 2003; Event Horizon Telescope Collaboration et al. 2022a) or X-rays (e.g. Baganoff et al. 2001; Degenaar & Wijnands 2009; Hailey et al. 2018) where the extinction is lower. To provide an understanding of the physical scales at the GC, we include Fig. 1.6. At the centre of the Galaxy sits the super massive black hole Sagittarius A* (Sgr A*), first discovered at radio wavelengths in Balick & Brown (1974). Sgr A* is located about 8.3 kpc from the Sun and has a mass of about 4.3×10^6 M$_\odot$ (Eisenhauer et al. 2005; Ghez et al. 2008; Schödel et al. 2009; Gillessen et al. 2009; Boehle et al. 2016; Gillessen et al. 2017; Do et al. 2019; GRAVITY Collaboration et al. 2019, 2022b), which means that the Schwarzschild radius $R_s = 2GM/c^2$ is close to 0.1 AU. Recent advances in very-long baseline interferometry have made it possible to obtain observations on scales similar to the Schwarzschild radius, resulting in the first publication of an image of Sgr A* (Event Horizon Telescope Collaboration et al. 2022a,b,c,d,e).

At about 10 times the Schwarzschild radius is the tidal disruption radius, where a star is torn apart by the tidal force from the black hole in a transient event known as a tidal disruption event (TDE; Hills 1975; Rees 1988; Rossi et al. 2021; Gezari 2021). Further out at roughly 100 AU is the radius at which binary stars are separated by the tidal force of Sgr A*, discussed in detail in Section 1.2.2.

The closest population of stars to Sgr A* is known as the S-star clus-

ter, which is a cluster of stars on isotropically distributed orbits with high eccentricities and semi-major axes up to ~ 0.04 pc (Genzel et al. 2010). The cluster is mainly composed of main-sequence B-type stars, but some evolved stars have also been found in the S-star cluster (e.g. Gillessen et al. 2017). It is unclear how these stars formed, since the strong tidal force so close to Sgr A* should inhibit star formation (Ghez et al. 2003). This lack of understanding on the origin of this cluster persists despite the fact that the S-star cluster has been intensely studied over the past decades (Eckart & Genzel 1996; Ghez et al. 1998). The main reason that the S-star cluster has been studied so much is its proximity to Sgr A*. Because these stars are so close to Sgr A*, their orbital periods are on a timescale that can now be resolved. The star S2 has been of particular importance for our understanding of the GC and Sgr A*, since it is a bright star ($m_K = 14$) on an orbit of ~ 16 years (Habibi et al. 2017a). Only a single star has been detected on a shorter orbit (S55; ~ 13 years), but since that star is much fainter at $m_k = 17.5$ it is also less constraining due to the difficulty of observing it. The monitoring of orbits of individual stars in the S-star cluster is how researchers have been able to determine the mass and distance to Sgr A* with an unprecedented precision. In recent years, tests of general relativity have even become possible based on observations of the S-star cluster, since Schwarzschild precession has now been measured for multiple stars (GRAVITY Collaboration et al. 2020; Gravity Collaboration et al. 2024). As an excellent example of how important the S-star cluster has been for astronomy, half of the Nobel Prize in Physics 2020 was awarded to Reinhard Genzel and Andrea Ghez 'for the discovery of a supermassive compact object at the centre of our galaxy'. Observations of the S-star cluster were what made this discovery possible.

At ~ 0.04 pc a transition occurs between the S-star cluster and at least one (warped) disc of massive stars (e.g. Levin & Beloborodov 2003; Paumard et al. 2006; Bartko et al. 2009; von Fellenberg et al. 2022). This so called clockwise disc (CWD) contains many O/WR stars out to ~ 0.5 pc, while no O/WR stars are detected within the S-star cluster. In addition, stars in the CWD travel on relatively low eccentricity orbits in comparison with the S-star cluster. The CWD is one of the only regions in the Galaxy where a top-heavy initial mass function (IMF) has been suggested (Paumard et al. 2006; Bartko et al. 2009; Lu et al. 2013; Gallego-Cano et al. 2024). The most likely origin of the CWD is through a burst of star formation in a gaseous disc (Levin & Beloborodov 2003; Paumard et al. 2006; Levin 2007). It is still uncertain if and how the CWD and S-star cluster are related.

Encompassing Sgr A*, the S-star cluster, and the CWD is the nuclear star cluster (NSC), with a half-light radius of about 5 pc (Schödel et al. 2014). The NSC is a much older population of stars compared to the S-star cluster and CWD, with most of the stellar content having formed billions of years ago (e.g. Pfuhl et al. 2011; Chen et al. 2023). The NSC is the densest

stellar cluster in the Galaxy with on the order of 10^6 stars per cubic parsec (e.g. Schödel et al. 2007). The Milky Way is not the only galaxy with a NSC, in fact they are observed at the centres of most galaxies (for a recent review, see Neumayer et al. 2020).

1.2   Hypervelocity stars

Now that we have discussed the general structure of the Milky Way, we will discuss the main subject of this thesis: HVSs (for review, see Brown 2015). We will start by going over some of the original discoveries, before going over the basic concepts of how these objects are formed.

1.2.1   From the first prediction to the present

Hills (1988) first predicted that if a stellar binary would approach a massive black hole (MBH) to within the tidal radius, where the tidal force from the MBH overcomes the self-gravity of the binary, one star would get captured into a tight elliptical orbit while the companion would get ejected at velocities potentially in excess of the escape velocity from the Galaxy. Hills dubbed the ejected stars HVSs, while the ejection mechanism has since been referred to as the Hills mechanism. When this prediction was first published, there was still active debate about the mere existence of MBHs. In fact, Hills (1988) suggested that the discovery of an HVS would provide near definitive evidence for a MBH. In Hills (1991), the original work was expanded upon with additional simulations exploring the parameter space.

It was not until the early 2000s that the subject started to gain more attention, for instance in Gould & Quillen (2003), who suggested that the star S2 might have been put onto its close orbit around Sgr A* by a tidal binary separation. The ejection of stars from the GC was revisited in Yu & Tremaine (2003), where the authors discussed three mechanisms through which stars might be ejected from the GC: close encounters of single stars, binary disruptions by a MBH (Hills mechanism), and the ejection of a single star by a binary black hole (BBH). The real breakthrough for the subject, however, came with the discovery of the first HVS in Brown et al. (2005) called HVS1. This serendipitous discovery was soon followed up by more HVS discoveries (Hirsch et al. 2005; Edelmann et al. 2005; Brown et al. 2006a,b) and marked the start of a wave of renewed interest in the subject.

About a decade later, there were a few tens of HVSs found in the literature, most notably by the MMT survey (Brown et al. 2009, 2012, 2014), which targeted blue stars in the stellar halo. These HVS discoveries tended to be distant ($\gtrsim 50$ kpc) main-sequence stars with Galactocentric velocities up to ~ 700 km s^{-1}. A side effect of the large distances to these stars is

the difficulty of constraining their past trajectories. Nonetheless, in further studies, it was found that at least some of the HVSs in fact cannot originate in the GC (e.g. Gualandris & Portegies Zwart 2007; Brown et al. 2018; Boubert et al. 2018) and alternative ejection origins proved difficult to rule out in general (see Section 1.2.3). This highlighted a potential issue for the science exploitation of these objects, because many of the science cases specifically rely on a GC ejection origin (see Section 1.2.4). The first, and currently only, unambiguous discovery of an HVS originating in the GC was presented in Koposov et al. (2020): a ~ 1700 km s^{-1} star that is relatively nearby (~ 9 kpc) and whose trajectory traces back to the GC called S5-HVS1. The combination of extreme velocity—much higher than previously discovered HVSs—and the relative proximity were what truly set this star apart from the previous discoveries. Interestingly, HVS1 and S5-HVS1, arguably the two most important stars in the field currently known, were both serendipitously discovered.

It should be noted at this point that the term HVS has not been consistently used in the literature. Brown (2015) define an HVS to be characterised by an MBH ejection origin and an unbound velocity. A complicating factor here is that in reality it is often uncertain if a star is unbound because of uncertainties in the Galactic potential, the stellar velocity, and its distance from the Sun. Additionally, alternative mechanisms exist that can accelerate stars to near, or in excess of, the Galactic escape velocity (see Section 1.2.3). Throughout this thesis, we define HVSs to be stars ejected from the vicinity of Sgr A*.

1.2.2 Hills mechanism

Having discussed some of the most important advances in the field of HVSs, we will now go over some of the theoretical aspects of the Hills mechanism. First we discuss the general principle through which a gravitational interaction near an MBH can cause a star to gain significant energy, where we follow the argumentation from Yu & Tremaine (2003). Suppose we have a star with specific energy $E = \frac{1}{2}v^2 + \Phi(r)$, where v is the velocity and $\Phi(r)$ is the specific potential energy. Close to the MBH $|\Phi(r)| \gg |E|$, which means that the star will travel approximately at the free-fall velocity. If, at this time, the star experiences a velocity perturbation $\delta v \ll v$, the specific energy of the star will change by $\delta E = \frac{1}{2}(v + \delta v)^2 - \frac{1}{2}v^2 \simeq v\delta v$. For a relatively modest increase in velocity δv, δE can become much larger than $|E|$, which means that the star is ejected. In the Hills mechanism, the velocity perturbation is caused by the tidal separation of a stellar binary. As mentioned before, a stellar binary is separated by an MBH when the tidal force of the MBH exceeds the self-gravity of the binary. The radius at which
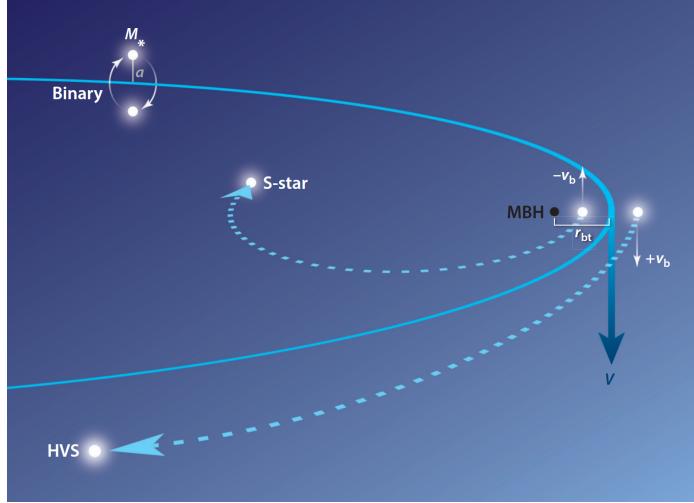
Figure 1.7: Schematic illustration of the Hills mechanism in which a stellar binary is separated by the tidal force from an MBH. *Image from Brown (2015).*

this occurs is the tidal separation radius

$$r_{\rm t} = a_{\rm b} \left(\frac{M_{\rm MBH}}{m_{\rm b}} \right)^{1/3}, \tag{1.1}$$

where $a_{\rm b}$ is the semi-major axis of the progenitor binary, $M_{\rm MBH}$ the mass of the MBH, and $m_{\rm b}$ the mass of the stellar binary. Note that in reality a binary can be disrupted at larger radii and survive at smaller radii (Sersante et al. 2025). A schematic illustration of this process is shown in Fig. 1.7.

   If the progenitor binary is on a parabolic orbit around the MBH, the ejection probability is equal between the two stars in the binary, independent on mass ratio (Sari et al. 2009a; Kobayashi et al. 2012). If we consider the isolated system of the MBH and the stellar binary, the velocity of the ejected star at infinity is

$$v = \alpha \sqrt{\frac{2Gm_{\rm cap}}{a_{\rm b}}} \left(\frac{M_{\rm MBH}}{m_{\rm b}} \right)^{1/6}, \tag{1.2}$$

where α is a prefactor of order unity, G the gravitational constant, and $m_{\rm cap}$ the mass of the captured companion (e.g. Rossi et al. 2014). The prefactor α depends on the exact geometry of the encounter including the phase of the binary, orbital inclination, eccentricity, and diving factor, where the diving factor is defined as the tidal separation radius over the pericentre of the centre-of-mass orbit (Sersante et al. 2025). Because the MBH is embedded in the Galaxy, the velocity defined above can be seen as the velocity at

which a star is ejected from the GC. For an equal mass binary, we can rewrite equation 1.2 to

$$v \simeq 1500 \text{ km s}^{-1} \left(\frac{m_{\text{b}}}{2M_{\odot}}\right)^{1/3} \left(\frac{a_{\text{b}}}{0.1 \text{ AU}}\right)^{-1/2} \left(\frac{M_{\text{MBH}}}{4.3 \times 10^6 \ M_{\odot}}\right)^{1/6} \alpha, \quad (1.3)$$

which clearly demonstrates the extreme velocities that the Hills mechanism can eject stars at. The captured star on the other hand is captured into an orbit around the MBH with a semi-major axis

$$a_{\text{cap}} = \frac{a_{\text{b}} M_{\text{MBH}}}{2\alpha^2 m_{\text{ej}}} \left(\frac{m_{\text{b}}}{M_{\text{MBH}}}\right)^{1/3}, \quad (1.4)$$

with m_{ej} the mass of the ejected star. Similarly to above, we can rewrite this as

$$a_{\text{cap}} \simeq 0.005 \text{ pc } \left(\frac{a_{\text{b}}}{0.1 \text{ AU}}\right) \left(\frac{m_{\text{b}}}{2 \ M_{\odot}}\right)^{-2/3} \left(\frac{M_{\text{MBH}}}{4.3 \times 10^6 \ M_{\odot}}\right)^{2/3} \frac{1}{\alpha^2}, \quad (1.5)$$

which shows that the captured stars can end up on orbits similar to the S-stars in the GC (see Section 1.1.4).

1.2.3 Alternative ejection mechanisms

We have discussed the Hills mechanism as an origin of HVSs, however there are alternative methods to accelerate stars to the escape velocity of the Galaxy. Here we will briefly discuss a number of these mechanisms. Before we do so, we should recognise that extragalactic HVSs should exist according to the same mechanisms as Galactic HVSs and that some might have already been discovered (Edelmann et al. 2005; Gualandris & Portegies Zwart 2007; Przybilla et al. 2008a; Evans & Massey 2015; Han et al. 2025). Similarly, the Hills mechanism can also operate with intermediate mass black holes in star clusters (Häberle et al. 2024).

As mentioned before, Yu & Tremaine (2003) identify two additional ejection mechanisms from the GC: gravitational encounters of single stars and ejections by a BBH. An ejection by a gravitational encounter between single stars is limited by the escape velocity from the surface of the star. This means that ejection velocities tend to be significantly lower compared to the Hills mechanism. Only for rare, close encounters can this mechanism eject a star with sufficient energy to escape from the Galactic bulge. The rate of these ejections is estimated to be orders of magnitude lower compared to the Hills mechanism with only $\sim 10^{-9}$ yr^{-1} able to reach out to the solar circle. For ejections involving a massive BBH, the average ejection speed is

$$v_{\text{ej}} \simeq \left(2.7 \times 10^3 \text{ km s}^{-1}\right) \left(\frac{\nu}{0.1}\right)^{1/2} \left(\frac{M_1}{4.3 \times 10^6 \ M_{\odot}}\right)^{1/2} \left(\frac{1 \text{ mpc}}{a_{\text{BH}}}\right)^{1/2},$$
$$(1.6)$$

with ν the mass ratio between the less massive black hole and the BBH, M_1 the mass of the more massive black hole, and a_{BH} the semi-major axis of the BBH (Yu & Tremaine 2003). This ejection velocity is similar to the Hills mechanism, but requires an, as of yet, undetected companion to Sgr A* (e.g. Gualandris & Merritt 2009; Naoz et al. 2020; GRAVITY Collaboration et al. 2020; Evans et al. 2023). As opposed to the Hills mechanism, the BBH slingshot mechanism thus requires additional assumptions and it is not clear if this mechanism currently operates at the GC.

Neither of the alternatives to the Hills mechanism suggested in Yu & Tremaine (2003) thus provide an appealing explanation for currently observed HVSs travelling with velocities in excess of the escape velocity of the Galaxy. For completeness we also point out two more mechanisms that have been suggested in the literature: a variety of the first mechanism of Yu & Tremaine (2003), where stars are scattered of stellar mass black holes in the GC rather than other stars (O'Leary & Loeb 2008) and the disruption of a dwarf galaxy near the GC (Abadi et al. 2009). These two mechanisms both predict distinct properties of the ejected stars in comparison to the Hills mechanism and neither can convincingly explain the existence of S5-HVS1. The Hills mechanism is therefore the leading theory for the ejection of the currently observed HVSs. However, stars have been found in the stellar halo that are unbound to the Galaxy, but do not originate in the GC (e.g. Irrgang et al. 2018; Kreuzer et al. 2020; Irrgang et al. 2021). Note that according to our definition of HVSs—having been ejected from the vicinity of Sgr A*—these stars do not classify as HVSs. In the literature, often no distinction is made on the origin of the stars, with 'high velocity' sometimes being the only defining feature. Since such stars can easily be confused with HVSs, we discuss the two leading ejection mechanisms. Because these stars did not originate in the deep potential well of the GC, their ejection velocities can be lower while still being energetic enough to escape from the Galaxy.

The first mechanism to accelerate stars to velocities in excess of the escape velocity of the Galaxy is through core-collapse supernovae in binary systems (Blaauw 1961; Tauris & Takens 1998; Portegies Zwart 2000; Przybilla et al. 2008b; Evans et al. 2020). The rapid mass loss of the companion during the supernova in combination with a kick induced by an asymmetric explosion can accelerate the companion to velocities in excess of the escape velocity (e.g. Tauris 2015). Typical ejection velocities through this mechanism are, however, thought to be much lower at up to a few tens of $\mathrm{km\,s^{-1}}$ (e.g. Renzo et al. 2019). An alternative way to eject stars is through dynamical interactions in dense stellar systems (Poveda et al. 1967; Leonard & Duncan 1988, 1990; Leonard 1991; Perets & Šubr 2012; Oh & Kroupa 2016; Evans et al. 2025). The ejection velocity in this scenario is maximised for an encounter in which a physical collision is narrowly avoided, which limits the ejection velocity to below the escape velocity from the stellar surface (on

the order ~ 600 km s^{-1} for a Sun-like star). The rates at which these two mechanisms produce ejections in excess of the escape velocity are thought to be significantly lower compared to the ejection rate from the GC, but remain highly uncertain (e.g. Silva & Napiwotzki 2011; Brown 2015; Evans et al. 2020; Irrgang et al. 2021; Evans et al. 2025).

1.2.4 HVS science cases

Now that we have established what HVSs are and how they can be formed, we go into the potential science cases. The science case that we attempt to exploit most throughout this work is the ability of HVSs to inform us about properties of the GC. The GC is a highly challenging region to study due to the relatively large distance (~ 8 kpc), high extinction ($A_{\mathrm{V}} \sim 50$; Fritz et al. 2011), and difficulty of resolving individual stars due to source crowding. However, this is also a very interesting region, since it is host to the MBH Sgr A*, the densest stellar environment in the Galaxy, and a plethora of stellar and dynamical phenomena that are not yet understood (for a review, see Genzel et al. 2010). HVSs offer a unique probe into this region of the Galaxy, because they are ejected out of the GC and are currently observable in the stellar halo, with little extinction and without the issue of source crowding. If these stars are ejected through the Hills mechanism, we know they were formed in a binary, likely near the GC (e.g. Penoyre et al. 2025). Their properties, such as metallicity, mass, and age are therefore informative for the same properties of stars near the GC, which is what some of the work in this thesis focusses on (Verberne et al. 2024a, 2025). These properties are extremely difficult to currently measure for stars near Sgr A* (e.g. Habibi et al. 2017b), while they are informative on the assembly history of the GC, star formation in extreme environments, and the evolution of Sgr A*. In addition, their velocity distribution is a measure of the progenitor binary semi-major axis and mass ratio distributions (Evans et al. 2022a), which are uncertain even for field stars (Moe & Di Stefano 2017). Furthermore, their distribution on sky can be used to infer if progenitor binaries approach the MBH isotropically and even if Sgr A* has (or had) a massive companion black hole (e.g. Evans et al. 2023, see also Section 1.2.3).

Another potential of HVSs is to measure the Galactic potential. Because we know the origin of these stars (the GC), their current position and velocity describe the Galactic potential (Gnedin et al. 2005; Yu & Madau 2007; Kenyon et al. 2008; Perets et al. 2009a; Rossi et al. 2017; Fragione & Loeb 2017; Kenyon et al. 2018; Contigiani et al. 2019; Gallo et al. 2022; Armstrong et al. 2025). Using HVSs to measure the Galactic potential has, however, proved challenging. The HVSs that are most sensitive to the Galactic potential are the ones that are ejected with a velocity similar to the escape velocity of the Galaxy (Rossi et al. 2014). The problem here arises because

these HVSs are typically also the most difficult to unambiguously associate with a GC ejection due their overlap in parameter space with other stellar populations (e.g. Przybilla et al. 2008b,a; Kreuzer et al. 2020; Irrgang et al. 2021). Meanwhile, the single unambiguous HVS (largely unambiguous due to its extreme velocity) is too fast to have been significantly affected by the Galactic potential (Koposov et al. 2020). Another challenge for measuring the Galactic potential using HVSs is that the number of observed HVSs required for these methods to be competitive is typically much larger than is currently observed.

A science case which has already proved successful is measuring the Solar motion and distance to the GC using HVSs, in particular S5-HVS1 (Hattori et al. 2018; Koposov et al. 2020). With a second HVS added to the analysis, the precision of this measurement will likely be comparable to that presented in GRAVITY Collaboration et al. (2019) (Koposov et al. 2020).

1.3   Large observational surveys

Critical for our understanding of the Galaxy and HVSs are the large observational surveys that have allowed us to study the stars around us in unprecedented detail. In order to determine the orbit of a star, we need to measure the radial velocity, the angular velocity and position in the plane of the sky, and the distance to the star. Especially over the past decades, various surveys have made significant strides in each of these areas (e.g. Perryman et al. 1997; York et al. 2000; Zhao et al. 2012; De Silva et al. 2015; Gaia Collaboration et al. 2016; Majewski et al. 2017a). Here we will discuss the two most important surveys for this thesis.

1.3.1   *Gaia*

Most prominent among the recent large surveys is the *Gaia* mission (Gaia Collaboration et al. 2016), which was a space-based observatory that had the ambitious goal of measuring astrometry for more than one billion stars. Astrometry deals with measuring the position and movement of stars and can be used to determine distances to stars and their velocities perpendicular to our line of sight. A full astrometric solution includes sky position, proper motion, and parallax. The proper motion of a star is its angular velocity in the plane of the sky, measured in two directions, with respect to a stationary reference frame of quasars. The parallax is the apparent movement of the position of a star by the change in viewing angle caused by observing a star at multiple points throughout the orbit of the telescope around the Sun. In Fig. 1.8, we schematically show the position of a star as a function of time under the effect of proper motion and parallax. The parallax of a star can
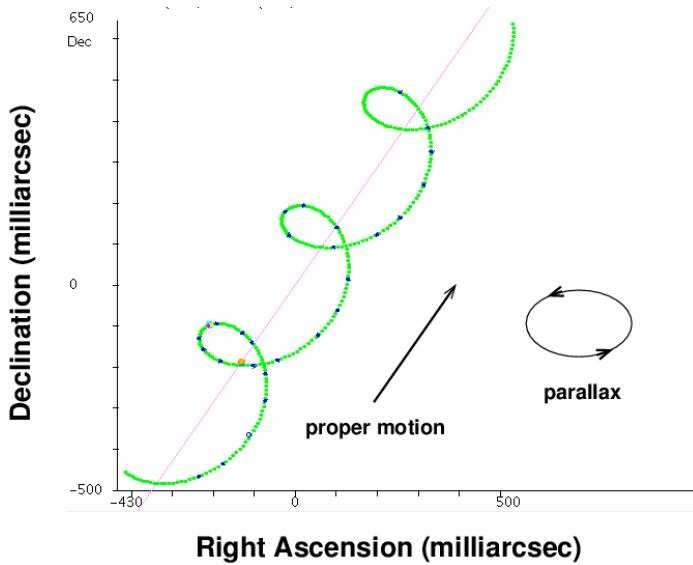
Figure 1.8: The on-sky movement of a star, which is the super position of the effects of proper motion and parallax. *Image credit: Michael Richmond.*

be used to determine the distance following

$$d = \frac{1}{\varpi},\tag{1.7}$$

where d is the distance in kpc and ϖ the parallax in mas. *Gaia* works in an ingenious way to obtain these measurements. Unlike most telescopes, *Gaia* looks in two directions at the same time using two mirrors with fields-of-view separated by 106.5 degrees. The entire spacecraft spins at a fixed rate of 60 arcseconds per second around its spin axis. This rate has to be maintained precisely, since it is matched to the speed at which electrons are transferred inside the CCD. This also means that *Gaia* does not take pictures of the sky in the way we are familiar with for most other optical telescopes. The spin axis is such that one of the telescopes 'leads' while the other 'follows', making the two fields of view pass the same parts of the sky. The spin axis itself is inclined by 45 degrees with respect to the Sun to allow optimal parallax sensitivity. Finally, *Gaia* also precesses with a period of 63 days to allow it to observe the entire sky. The way that *Gaia* effectively observes the sky is called the scanning law, since the way *Gaia* observes is akin to scanning. Light from both fields of view ends up at the focal plane of *Gaia* shown in Fig. 1.9. The blocks correspond to individual CCDs, while their colours show the different functions. On the left we see the Sky Mapper CCDs, which handle object detection. In the middle are the Astrometric
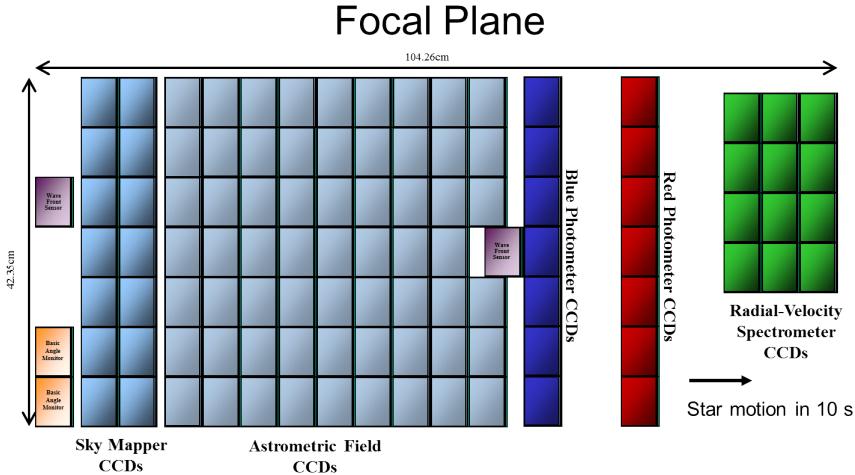
Focal Plane

Figure 1.9: Focal plane of *Gaia* showing the different types of CCDs and their layout. *Image credit: ESA. Acknowledgement: Alex Short.*

Field CCDs, which perform the astrometric measurements. Next to that, we see in blue and red the Photometer CCDs, which provide low-resolution spectro-photometric measurements. Finally there are the Radial-velocity Spectrometer (RVS; Cropper et al. 2018) CCDs, which take high resolution spectra of bright sources. We will now discuss the main data products that these instruments provide.

The latest *Gaia* data release at the time of writing is data release 3 (DR3; Gaia Collaboration et al. 2023a), which has provided full astrometric solutions to about 1.46×10^9 sources with a limiting magnitude of $G = 20.7$, including magnitudes in three broad-band filters: G, $G_{\rm BP}$, and $G_{\rm RP}$. For the brightest of these stars, some 34 million, radial velocity measurements are available from the RVS (Katz et al. 2023). Although this is the largest precision radial velocity catalogue available, it only covers a fraction of the sources with full astrometric solutions in *Gaia*. New in *Gaia* DR3 compared to DR2 have been the low-resolution BP/RP spectra from the red and blue photometers (Gaia Collaboration et al. 2016; Montegriffo et al. 2023; De Angeli et al. 2023). Currently, these spectra are published for some 220 million sources, while with the next data release they will be published for all sources in *Gaia*. Although much lower resolution compared to the RVS spectra, the wavelength range covers the entire optical spectrum for sources with a limiting magnitude of $G = 17.65$. These low-resolution spectra are the subject of Chapter 2 of this thesis (Verberne et al. 2024b).

Now that we have gone over the main data products that the *Gaia* cata-

logue provides, we will discuss the limitation in using parallax measurements from *Gaia* to measure distance. For stars with $G < 15$, the median parallax uncertainty from *Gaia* data release 3 (DR3) is $0.02 - 0.03$ mas (Gaia Collaboration et al. 2023a). It should be clear that *Gaia* can therefore not effectively constrain distances beyond a few tens of kpc, given the inverse relation between distance and parallax. However, it turns out that obtaining a distance from simply inverting the parallax becomes problematic much sooner (Luri et al. 2018). While parallax can be negative, distance cannot. This, in combination with the non-linear relation between distance and parallax, causes the normally distributed uncertainties on the parallax to not translate into normally distributed uncertainties in distance space. It should also be noted that there are systematic biases present, such as the zero-point offset in the parallax (Lindegren et al. 2018, 2021). Bailer-Jones (2015) argue that once the relative uncertainty on the parallax exceeds 20%, it becomes non-trivial to obtain an unbiased distance estimate. Simple distance estimates from inverting the parallax can therefore typically only be used up to at most a few kpc with the data from *Gaia*. Beyond these distances, additional information should be incorporated when considering the distance, such as colour, magnitude, and sky position (Bailer-Jones et al. 2018, 2021).

Also, although the limiting magnitude of *Gaia* is $G = 20.7$ as mentioned before, that does not mean that every source brighter than this limit is represented in the *Gaia* catalogue. If we imagine a fictitious star with properties $\vec{p}$ located at different positions in the sky, the chance of this star appearing in the *Gaia* catalogue is not fixed. For various reasons including a limit on the number of sources that *Gaia* can detect simultaneously, source crowding in some parts of the sky, and uneven sky coverage due to the scanning law, *Gaia* is more 'complete' in some directions than others. We show the chance of a star with $G = 21$ appearing in the *Gaia* DR3 catalogue in Fig. 1.10. In the figure we can recognise some of the dependencies mentioned above. Most obvious is the pattern of stripes, which is caused by the scanning law: the red areas tend to be observed more often. In addition, we see that the GC, at the centre of the figure, has a low completeness at $G = 21$, which is caused by the large number of sources saturating *Gaia*. The chance for a star with properties $\vec{p}$ to be observed at various positions in the sky is referred to as the selection function. Note that $\vec{p}$ are the properties as seen from the observer, which means that interstellar extinction does not directly influence the selection function. In the case of *Gaia*, the selection function is fairly well defined thanks to the efforts of the GaiaUnlimited team (Cantat-Gaudin et al. 2023; Castro-Ginard et al. 2023) and can be mostly characterised by sky position and G magnitude.
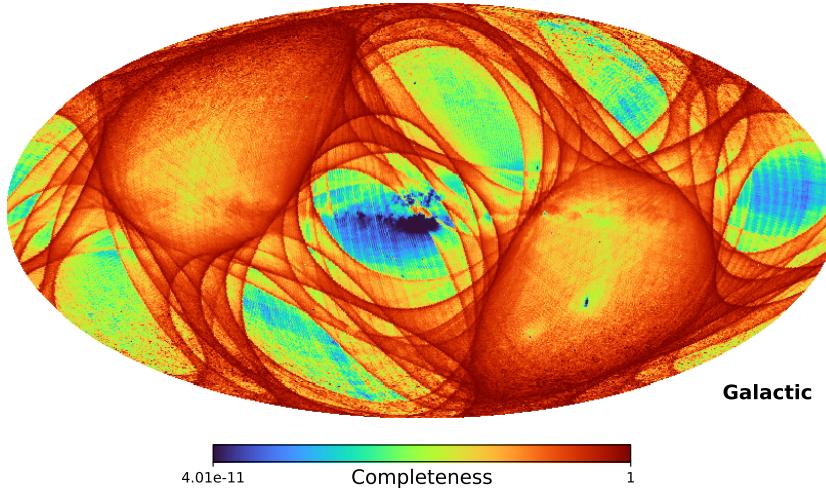
Figure 1.10: The completeness of the *Gaia* catalogue at $G = 21$ as a function of sky position in Galactic coordinates.

1.3.2   DESI

The Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration et al. 2016a) is a multi-object spectrograph located at Kitt Peak National Observatory in Arizona, capable of taking 5000 spectra simultaneously (DESI Collaboration et al. 2016b). The instrument is designed to study cosmology, but also observes stars as secondary targets as part of its Milky Way Survey (MWS; Cooper et al. 2023; Koposov et al. 2025). Despite stars being a secondary target, the MWS survey is still expected to observe more than 12 million stars in total (Koposov et al. 2025). The wavelength coverage of DESI is $360 - 980$ nm spread over three channels, which covers the entire visible wavelength range and part of the near-infrared. The spectral resolution $R = \lambda/\Delta\lambda$ over this range varies between 2000 and 5000. *Gaia* RVS in comparison has a higher resolution at $R = 11\,500$, but only a very narrow wavelength range $\lambda \in [845, 872]$ nm. Another key difference between DESI and *Gaia* RVS is the magnitude range covered, which we show in Fig. 1.11. We can see that *Gaia* RVS is overall the largest catalogue, but only covers relatively bright stars. DESI on the other hand is the largest catalogue for faint sources with $17 \lesssim G \lesssim 21$. In addition to above mentioned characteristics, a key difference of DESI compared to *Gaia* is that DESI only covers part of the sky, which is a consequence of both survey strategy and the limitation of being a ground based observatory.
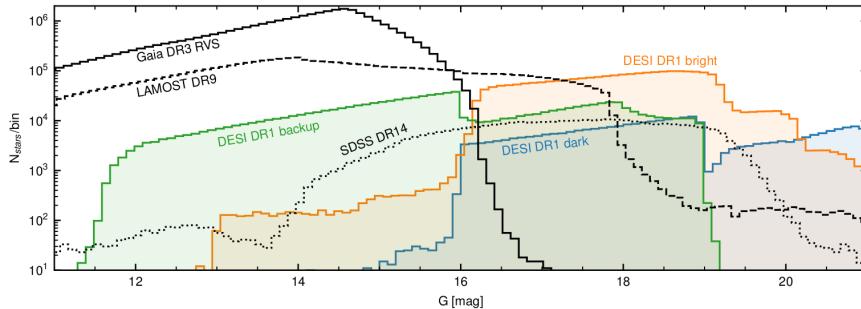
Figure 1.11: The number of stars with spectroscopic (radial velocity) measurements as a function of G magnitude in various surveys. Note that DESI DR1 is currently the largest catalogue for stars with $17 \lesssim G \lesssim 21$. *Figure from Koposov et al. (2025).*

In the final chapter of this thesis, we will use DESI DR1 (DESI Collaboration et al. 2025a), which was released in March 2025. This data release includes spectra and data products for ~ 2.5 million stars as part of the bright program, which is the main program for the MWS (Koposov et al. 2025) and the program we use in Chapter 5. Apart from radial velocities, the value-added catalogues of DESI DR1 also contain measurements of distance, metallicity, α enhancement, and a number of individual elements.

1.4 Future observational prospects

Now that we have gone over the background for this thesis, we will discuss the future prospects for HVSs, where we focus on the observational aspect.

Science exploitation of HVSs has so far been limited by the number of known HVSs. Identifying additional HVSs is therefore of prime importance in the field. The most exciting prospects for identifying more HVSs are the ongoing and upcoming large spectroscopic surveys by *Gaia* (including astrometry), *WEAVE* (Dalton et al. 2014; Jin et al. 2024), DESI, and *4MOST* (de Jong et al. 2019).

Although *Gaia* has recently been decommissioned, a wealth of data remains unpublished and will be released in DR4 and DR5. The latest DR3 only publishes data from the first 34 months of *Gaia* operations between 2014 and 2017. Meanwhile *Gaia* continued to observe till it was decommissioned in early 2025 (Gaia Collaboration et al. 2021). In DR4, *Gaia* will publish radial velocities for more than 100 million stars, in addition to individual epoch measurements (Katz et al. 2023). The large increase in radial velocities in combination with improved accuracy and precision for the astrometric measurements is expected to allow for much more sensitive

HVS searches (e.g. Evans et al. 2022a,b).

Despite *Gaia* increasing the number of stars with measured radial velocities by a factor of at least a few in the next data release, these measurements will still be limited to the relatively bright stars with $G_{\mathrm{RVS}} \lesssim 16$. The aforementioned large ongoing and upcoming spectroscopic surveys performed by DESI, *WEAVE*, and *4MOST* will provide millions of radial velocity measurements for stars up to *Gaia*'s limiting magnitude of $G = 21$. In addition, these instruments, unlike *Gaia*, also allow for targeted follow-up observations of promising HVS candidates. These HVS candidates can be identified using available data, such as distance, colour, brightness, and astrometry. Spectroscopic follow-up observations can then be used to measure the radial velocities to these stars, which might allow confirmation or rejection of the HVS candidates. This method is much more efficient compared to magnitude limited radial velocity surveys such as *Gaia* in identifying HVSs. The multi-object spectrographs on the aforementioned instruments allow the simultaneous measurement of a thousand or more radial velocities. This means that follow-up observations of large numbers of HVS candidates become feasible. For instance about 20 000 HVS candidates will be observed as part of the high-latitude low-resolution survey of *WEAVE*. This is in addition to the possibility of serendipitous HVS discoveries in these surveys of course.

Apart from larger surveys, more effective identification of HVS candidates is an area with a lot of potential for improvement. For instance, the first HVS survey only relied on colour-magnitude information to identify HVS candidates. More recently, many 'blind' searches have been conducted for HVSs, in which all stars with 6D phase-space information are analysed to search for stars with either high Galactocentric velocities and/or a trajectory that is consistent with originating in the GC. However, the astrometric data provided by *Gaia* also facilitates more sophisticated selection procedures, such as presented in Chapter 3. The challenge in identifying HVS candidates often comes down to a trade-off between completeness and purity. Completeness refers to the fraction of true HVSs that a method will identify, while purity refers to the ratio of true HVSs over the total number of identified candidates. Applying more stringent requirements to consider a particular star to be an HVS candidate will typically decrease the completeness, while improving the purity.

1.5   Thesis content

In this thesis we use state-of-the-art observations and simulations to provide the best constraints to date on the HVS ejection rate and constraints on the stellar content and dynamics at the GC, including the progenitor binary populations and their star formation history. The challenge in identifying

HVSs is often the observational volume of the sample that is used, which is a common thread through most of the chapters in this thesis. We first attempt to identify new HVSs by increasing the number of stars with radial velocity solutions using the 220 million low resolution spectra provided by *Gaia* DR3, before switching to a method that looks for the fastest (> 800 km s^{-1}) HVSs in all ~ 1.5 billion sources in *Gaia*. In Chapter 4 we combine these measurements with observations of the GC to perform a robust analysis linking the S-stars with the stars captured by the Hills mechanism and in the final chapter we use the newly released DESI DR1 MWS to further increase our observational footprint and search instead for HVSs ejected over billions of years that have remained bound to the Galaxy. We briefly summarise each chapter below.

In Chapter 2 we start out by attempting to measure radial velocities from the aforementioned *Gaia* BP/RP spectra in *Gaia* DR3. Our goal is to increase the number of sources with 3D velocities, since that makes it more straightforward to identify HVS candidates. Using model grids of stellar spectra, we fit every available BP/RP spectrum and marginalise over all stellar parameters other than radial velocity. We find that BP/RP spectra in *Gaia* can be used to measure radial velocities, albeit at much lower precision compared to, for instance, *Gaia* RVS. We perform calibration of the radial velocities in colour-magnitude-extinction space and publish a catalogue of BP/RP measured radial velocities. Our extended catalogue, containing some 125M radial velocity measurements is to our knowledge the largest radial velocity catalogue to date. Despite the size of our catalogue, we find that identifying HVSs from our catalogue is non-trivial, due to the large nominal uncertainties and a fraction of sources with unreliable measurements.

In Chapter 3 we follow a different approach to identify HVS candidates. Instead of only looking at stars with 3D velocity measurements, we use a novel method to select stars that appear to be travel on radial trajectories from the GC, which can be applied to all ~ 1.5 billion stars with astrometric solutions in *Gaia*. We create a catalogue of 600 HVS candidates, based on their proper motions, sky position, and colour and perform follow-up observations for about 200 of those using ground-based spectroscopic instruments. Because we use both northern and southern hemisphere telescopes, we can cover our HVS candidates across the sky. We use the non-detection of new HVSs among our 200 observed HVS candidates in combination with sophisticated simulations to significantly improve constraints on the ejection rate and mass function of HVSs and provide predictions for the undiscovered population of HVSs in *Gaia*.

Chapter 4 in a way is the odd one out in this thesis, since we do not attempt to discover new HVSs. Instead we combine observations of the GC with our own observational catalogue described in Chapter 3 and compare to simulations to investigate if the Hills mechanism alone can explain the

observed properties of the S-star cluster. We use realistic star formation histories for stars near the GC and simulate binary disruptions. We find that no single progenitor binary population can explain all the observed properties of both the S-star cluster and HVSs simultaneously. Instead, we find that at least two progenitor populations are required: an old population and a young population. Assuming that the young population is the CWD near Sgr A* and the old population undergoes Hills mechanism disruptions at a constant rate, we find that the recent star formation episode at the GC which formed the CWD has boosted the Hills mechanism disruption rate by about an order of magnitude compared to the background rate over the past ~ 10 Myr.

Finally in Chapter 5 we again aim to identify new HVSs by going to an even larger potential discovery space: stars ejected from the GC that have remained bound to the Galaxy. Since HVS ejections will have been happening for billions of years and a significant fraction of these is expected to remain bound to the Galaxy, a population of stars ejected from the GC should have accumulated in the stellar halo. Using the recently released DESI DR1 MWS survey data, we aim to statistically identify an overdensity of high metallicity stars on low angular momentum orbits, characteristic of stars ejected from the GC. Although we obtain a null-detection for a GC ejected population of stars, we use this null-detection to put upper limits on the ejection rate of stars from the GC over the past ~ 5 Gyr, which is ejection model independent.

2 | Radial velocities from *Gaia* BP/RP spectra

Abstract

The *Gaia* mission has provided us full astrometric solutions for over 1.5B sources. However, only the brightest 34M of them have radial velocity measurements. This paper aims to close that gap by obtaining radial velocity estimates from the low-resolution BP/RP spectra that *Gaia* now provides. These spectra are currently published for about 220M sources, with this number to increase to the full $\sim$ 2B *Gaia* sources with *Gaia* Data Release 4. To obtain the radial velocity measurements, we fitted *Gaia* BP/RP spectra with models based on a grid of synthetic spectra. From this we obtained the posterior probability on the radial velocity for each object. Our measured velocities show systematic biases that depend mainly on the colours and magnitudes of stars. We corrected for these effects by using external catalogues of radial velocity measurements. We present a catalogue of about 6.4M sources with our most reliable radial velocity measurements and uncertainties of < 300 km s^{-1} obtained from BP/RP spectra. About 23% of them have no radial velocity measurement from the *Gaia* RVS. Furthermore, we provide an extended catalogue that contains all 125M sources for which we were able to obtain radial velocity measurements. This catalogue, however, also contains a fraction of measurements for which the reported radial velocities and uncertainties are inaccurate. Although typical uncertainties in the catalogue are significantly higher compared to those obtained with precision spectroscopy instruments, the number of potential sources to which this method can be applied is orders of magnitude higher than any previous radial velocity catalogue. Further development of the analysis could therefore prove extremely valuable in our understanding of Galactic dynamics.

2.1 Introduction

The *Gaia* mission (Gaia Collaboration et al. 2016) has been collecting data since 2014, with its primary scientific data products being the positions, proper motions, and parallaxes of about 1.5B objects. In the recent *Gaia* Data Release 3 (DR3) (Gaia Collaboration et al. 2022), low-resolution spectra of $\sim 220M$ objects have additionally been published. These spectra were obtained from two low-resolution prism spectrographs: BP observes in the wavelength range 330–680 nm and RP in the 640–1050 nm range; together they are referred to as XP spectra. Their primary purpose is to provide source classification and astrophysical information for the astrometric sources observed, for example stellar metallicity and line-of-sight extinction (Bailer-Jones et al. 2013). In order to measure stellar parameters, such as radial velocities and elemental abundances, *Gaia* is equipped with the Radial Velocity Spectrometer (RVS; Katz et al. 2023). However, RVS spectra, being limited to $G_{RVS} \leq 16$ in *Gaia* DR4, will not be available for all *Gaia* sources (Katz et al. 2023). XP spectra, on the other hand, will be published in DR4 for all sources in the astrometric catalogue with a limiting magnitude of $G \approx 20.7$ (Gaia Collaboration et al. 2016). The current magnitude limit of XP spectra in *Gaia* DR3 is G = 17.65 (Gaia Collaboration et al. 2022).

XP spectra have been recognised as a rich source of astrophysical information, with efforts to measure [M/H], [α/M] [Fe/H], $\log g$, T_{eff}, and line-of-sight extinction, among other quantities (e.g. Rix et al. 2022; Zhang et al. 2023; Andrae et al. 2023b; Andrae et al. 2023a; Guiglion et al. 2023; Li et al. 2023). This work focuses on obtaining radial velocity measurements, for the first time, from the low-resolution XP spectra. Although the precision of any radial velocity measurement from XP spectra is expected to be lower than that of conventional spectroscopic surveys, the scientific content would still be very significant due to the number of objects (220M currently and $\sim$ 2B in DR4). This would constitute a factor of $\sim$ 6.5 increase in the total number of sources with radial velocity measurements compared to the currently largest radial velocity catalogue, *Gaia* DR3, which contains $\sim$ 34M measurements. Additionally, the recently launched *Euclid* space telescope (Laureijs et al. 2011) also includes a low-resolution slitless spectrograph. Although the *Euclid* instrument operates in the near infrared, the spectral resolution is significantly higher[1] compared to *Gaia* XP spectra, though still considered low-resolution. The method presented in this paper could in principle also be applicable to those data.

This work should be seen as a proof of concept. We focus on demonstrating our ability to obtain radial velocity information from *Gaia* XP spectra rather than obtaining the most accurate and precise measurements possible

---

[1] https://sci.esa.int/web/euclid/-/euclid-nisp-instrument

for all sub-types of objects appearing in *Gaia* DR3 XP spectra. We therefore advise readers to carefully consider if our measurements are appropriate for their specific use case.

This paper is structured as follows: Section 2.2 discusses the properties and format of *Gaia* XP spectra. Section 2.3 describes the analysis of the XP spectra we carried out to obtain radial velocity measurements. Section 2.4 provides the post-calibration we applied to our radial velocity measurements using reference radial velocities. Section 2.5 presents the random forest classifier (RFC) we trained for data quality assurance. Section 2.6 provides a validation of our calibrated results. Section 2.7 describes our main and extended catalogues, which we publish together with this paper. In Section 2.8 we discuss the science case of hypervelocity stars (HVSs) for our catalogue. In Section 2.9 we discuss our results and provide prospects for *Gaia* DR4. In Section 2.10 we give closing remarks.

2.2   BP/RP spectra

In the following section we discuss a number of important points on the calibration and representation of XP spectra in *Gaia* DR3. We only considered XP spectra from sources with $G < 17.65$, which is the main XP catalogue from *Gaia* consisting of $\sim$ 219M sources. The few hundred thousand sources fainter than this limit mainly consist of white dwarfs and quasi-stellar objects (QSOs) (Gaia Collaboration et al. 2022).

2.2.1   *Gaia* XP calibration

The *Gaia* mission relies on self-calibration where possible. In the case of spectra, this is only possible to a limited degree. The calibration of multiple measurements, taken possibly years apart using different charge-coupled devices (CCDs) and fields of view, into a single mean spectrum is described by De Angeli et al. (2023) and is done using self-calibration. The calibration onto a physical wavelength and flux scale is described by Montegriffo et al. (2023) and is performed using external measurements.

At wavelengths below 400 nm and above 900 nm, the wavelength calibration is less accurate, because there is an insufficient number of calibrator QSOs with emission lines in that part of the spectrum (see Fig. 19 of Montegriffo et al. 2023). In addition, there is a systematic offset in the RP spectra (Montegriffo et al. 2023), which would lead to a systematic offset in radial velocity if not corrected for. The exact origin of this offset is unknown, but Montegriffo et al. (2023) note that it might be caused by a systematic error in the line-spread-function model.

In terms of flux calibration, the uncertainties are typically underestimated (see Fig. 18 of De Angeli et al. 2023). This underestimation is more

pronounced for bright sources but affects the majority of spectra and is wavelength dependent. The underlying cause of this underestimation is unknown.

2.2.2 Basis function representation

Gaia observes the same sources multiple times over a time span of years, using two different fields-of-view and an array of CCDs (Gaia Collaboration et al. 2016). Small differences in the dispersion, wavelength coverage, instrument degradation, etc. between observations gives the opportunity to extract more spectral information (i.e. higher resolution spectra) from the sources than would be possible given a single observation. Representing this information in flux-wavelength space (henceforth referred to as sampled spectra), would be highly inefficient due to the small nature of the variations compared to the spectral resolution. For this reason, the *Gaia* consortium instead chose to represent the spectra as a series of coefficients for basis functions that describe the spectra. The BP and RP spectra are represented by 55 such spectral coefficients each, making for a total of 110 coefficients that describe every source. The first few coefficients contain most of the spectral information, since they are optimised for representing 'typical' *Gaia* sources (De Angeli et al. 2023). Alongside the spectral coefficients, *Gaia* has published their uncertainties and correlation coefficients, allowing us to construct the full covariance matrix for the coefficients of each source.

While providing more information than a sampled spectrum with the same number of samples, the representation in spectral coefficients also introduces challenges. Due to the individual basis functions being continuous functions over the entire wavelength range, the uncertainties on all spectral coefficients are correlated. This means that when converting the basis function representation into sampled flux-wavelength space, all data points are correlated. Random noise in the initial *Gaia* observations in particular causes random wiggles in the sampled XP spectra that could be mistaken for physical spectral features.

2.3 Spectral analysis

Now that we have discussed some of the important features of the XP spectra, we now describe the spectral analysis of these data we carried out to obtain radial velocity measurements.

In this study we chose to convert the spectral coefficients to sampled spectra using `GaiaXPy`[2]. The conversion is performed through the design

---

[2]`https://gaia-dpci.github.io/GaiaXPy-website/`
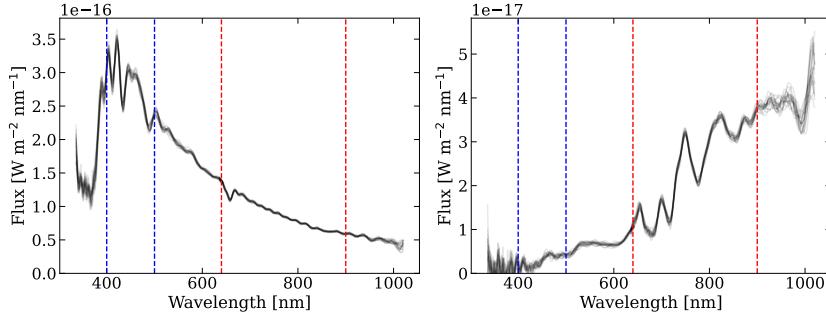
Figure 2.1: Example of two sampled *Gaia* XP spectra, in black. The uncertainties in the spectral coefficients are sampled over to indicate the uncertainties in the sampled spectra. The dashed blue and red lines indicate the BP and RP spectral ranges used in the fitting procedure, respectively. On the left we show a hot star, *Gaia* DR3 `source_id 191594196746880`, that displays prominent Balmer features. On the right we show a red source, *Gaia* DR3 `source_id 31958852451968,` that contains broad molecular absorption bands.

matrix as

$$\mathbf{s} = A \cdot \mathbf{b}, \tag{2.1}$$

with $\mathbf{s}$ the mean sampled spectrum, A the design matrix provided by `GaiaXPy`, and $\mathbf{b}$ the spectral coefficients. This gives us two spectra for each source, BP and RP, which we chose to sample on a grid of $\Delta\lambda = 2$ nm. However, we did not use the entire spectra for our analysis: we selected the wavelength range 400–500 nm for the BP spectra, while for the RP spectra we selected 640–900 nm. Two example sampled *Gaia* XP spectra are shown in Fig. 2.1. The BP range is chosen to include the prominent Balmer lines, but exclude the region below 400 nm, where the wavelength calibration might be problematic, and the region above 500 nm, where for many stars the continuum would dominate the fit. The RP-range is much wider and includes most of the RP spectral range, except the region above 900 nm, where again the wavelength calibration might be problematic (see Sect. 2.2.1).

Now that we have discussed how we handle the data, we describe in the following how we produced model spectra. We created the model $\mathbf{M}(T_{\text{eff}}, \log g, [\text{Fe/H}], v_{\text{r}}, \text{E}(\text{B} - \text{V}))$, where T_{eff}, $\log g$, and [Fe/H] correspond to the effective temperature, surface gravity, and metallicity from the PHOENIX spectral library, respectively (Husser et al. 2013), v_{r} is the radial velocity, and E(B−V) the extinction along the line of sight. For the PHOENIX models, we only considered atmospheres with $[\alpha/\text{H}] = 0$ for computational reasons. We shifted each of these models by a radial velocity

---

`https://dx.doi.org/10.5281/zenodo.7566303`

Table 2.1: Extrema of the parameter ranges of the grid.

Parameter	Minimum	Maximum
Radial velocity ($\mathrm{km\,s^{-1}}$)	-3000	3000
T_{eff} (K)	2300	15000
$\log g$ (dex)	-0.5	6.5
[Fe/H] (dex)	-3.0	1.0

with a step size of 30 $\mathrm{km\,s^{-1}}$. The parameter ranges for radial velocity, T_{eff}, $\log g$, and [Fe/H] are displayed in Table 2.1. The resulting models were convolved with the resolution of the externally calibrated XP spectra. This was done by interpolating the values from Table 1 in Montegriffo et al. (2023). These interpolated values were then used at each wavelength in the sampled data to spread the flux in wavelength space using a Gaussian with standard deviation $\sigma = \text{FWHM}/\left(2\sqrt{2\ln 2}\right)$. Lastly, we applied extinction on a source-to-source basis using the 2D extinction map from Schlegel et al. (1998) and the re-calibration from Schlafly & Finkbeiner (2011), assuming all sources to be behind the extinction layer. The extinction law we used is from Fitzpatrick (1999).

Having described how we prepared the XP spectra and created model spectra, we now describe how we fitted the spectral models to the XP spectra. Given an XP spectrum from *Gaia*, we could determine the likelihood of the data given a model, $P(\mathbf{D} \mid \mathbf{M})$, from

$$P(\mathbf{D} \mid \mathbf{M}) = \frac{1}{\sqrt{(2\pi)^k |\mathbf{C}|}} \exp\left(-\frac{1}{2}\left[\mathbf{D} - \mathbf{M}\right]^\top \mathbf{C}^{-1}\left[\mathbf{D} - \mathbf{M}\right]\right), \qquad (2.2)$$

where $\mathbf{D}$ is the data, $\mathbf{M}$ the model, k the number of dimensions, and $\mathbf{C}$ the covariance matrix of the data (e.g. Hogg et al. 2010). This holds in the case where the uncertainties are Gaussian with correctly estimated variances, which is not strictly true in our case. Because we over-sampled our sampled spectra with respect to the orthogonal bases, our covariance matrix in sampled space does not have full rank. To allow an inversion of the covariance matrix in sampled space, we only considered the diagonal elements and thus discarded correlation information. This caused the uncertainties on the radial velocity measurements to be further underestimated.

We calculated this likelihood for all models, after which we marginalised over the nuisance parameters T_{eff}, $\log g$, and [Fe/H]. We used flat priors on our parameters between the extrema of the parameter grid shown in Table 2.1. T_{eff} is the exception; for computational reasons, we instead only considered models differing by no more than 500K from an initial guess for T_{eff} we made based on the $BP - RP$ colour of a source and the extinction.

The initial guess for T_{eff} is described in Appendix 2.A. During analysis of the results, we noticed that the performance of this initial guess is poor for $E(B-V) \gtrsim 0.5$. For this reason we only report results for sources with $E(B-V) < 0.5$, for which the method works well.

In order to determine the radial velocity and corresponding uncertainty, we assumed a Gaussian posterior probability on the radial velocity. We fitted a parabola to the log-posterior probability by selecting all radial velocity points with a log-posterior probability of no less than 10 from the maximum log-posterior probability. If fewer than five points in radial velocity space met this requirement, we reduced the threshold by increments of 10 until we had more than five points. If the resulting fit peaked outside our radial velocity range of ± 3000 km s^{-1}, we considered the fit to have failed and report no radial velocity.

Now that we have laid the foundation of our method, we can describe the skewness and goodness-of-fit measurements we used to evaluate the reliability of the radial velocity (uncertainty) measurements (see Sect. 2.5).

By fitting a parabola to the log-posterior probability, we were assuming symmetric uncertainties. To evaluate if this is a reasonable assumption, we determined the skewness for the log-posterior probability distribution using

$$g_1 = \frac{\sum_{i=1}^{n} P_i \left(v_{\text{r}i} - \overline{v_{\text{r}}} \right)^3}{\left(\sum_{i=1}^{n} P_i \left(v_{\text{r}i} - \overline{v_{\text{r}}} \right)^2 \right)^{3/2}}, \tag{2.3}$$

where P_i is the posterior probability per radial velocity bin, $v_{\text{r}i}$ the corresponding radial velocity, and $\overline{v_{\text{r}}}$ the mean radial velocity given by

$$\overline{v_{\text{r}}} = \sum_{i=1}^{n} v_{\text{r}i} \cdot P_i. \tag{2.4}$$

This allowed us to identify cases in which the posterior probability distribution is asymmetric and for which the symmetric uncertainties might not be reliable. In addition, we calculated the reduced χ^2 of our best fit by approximating the number of degrees of freedom as the number of data points we had (i.e. number of flux vs wavelength points) minus the number of parameters we fitted (four).

2.3.1   Results of spectral analysis

Here we discuss the results from the spectral analysis presented above. As mentioned, the analysis was applied to all ~ 219M XP sources with $G < 17.65$. There are generally three outcomes possible for our spectral analysis. The first outcome is that we obtain a measurement for the radial velocity

Table 2.2: Summary of the raw results of this work: the number of sources analysed and the number of those for which we obtained a radial velocity estimate.

XP spectra analysed	218 969 408
Predicted T_{eff} outside grid range	15 041 788
E(B−V) $\geq$ 0.5	55 594 623
Failed fits	23 187 507
Radial velocities obtained	125 145 490

and corresponding uncertainty of a particular source. It is also possible that a fit failed, because the best-fit radial velocity was outside our parameter range of $\pm 3000 \, \mathrm{km \, s^{-1}}$, or a column in *Gaia*, such as the BP colour, required by our processing was not measured or was unavailable. The third outcome is that the initial guess for T_{eff} was outside our model range, in which case we did not perform a fit (see Table 2.1). We summarise the relevant numbers in Table 2.2.

2.4   Radial velocity calibration

Because of the calibration issues described in Sect. 2.2, we expected to see systematic offsets in our measurements of radial velocities that are a function of colour, magnitude, and extinction, in addition to an underestimation of uncertainties. To make matters more complicated, we expected the presence of an 'outlier' population, which is a population of objects for which the measured radial velocity spread is well beyond formal errors. In this section we describe how we corrected for these systematics in our radial velocities and their uncertainties obtained from the spectral analysis described in the previous section. We made use of reference radial velocity measurements from dedicated radial velocity surveys. We begin by describing our set of reference radial velocities in Sect. 2.4.1, followed by the statistical model to describe our XP radial velocities compared to the reference measurements in Sect. 2.4.2, and finally the fitting procedure of the model to the data in Sect. 2.4.3.

2.4.1   Reference dataset

The reference radial velocity measurements we used are *Gaia* RVS DR3 (Katz et al. 2023), Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) DR8 (low-resolution) (Zhao et al. 2012), and Apache Point Observatory Galactic Evolution Experiment (APOGEE) DR17 (Majewski et al. 2017a; Abdurro'uf et al. 2022). These catalogues were chosen because of their large size and sky coverage. Importantly, APOGEE and

Table 2.3: Summary of the reference radial velocity catalogue used to calibrate and validate our results. The combined size is smaller than the sum since there is overlap between the surveys.

Catalogue	Radial velocities matched to BP/RP	Magnitude limit
Gaia (RVS)	22 018 897	$G_{RVS} \leq 14^a$
LAMOST	3 666 919	$r \leq 17.8^b$
APOGEE	445 137	$H \lesssim 12.8^c$
Combined	23 900 765	

Notes. [a] Katz et al. (2023) ; [b] Yan et al. (2022) ; [c] Santana et al. (2021)

LAMOST contain sources fainter than the *Gaia* RVS magnitude cut, which is needed to calibrate and validate our results for faint sources. A summary of the relevant statistics from these catalogues is included in Table 2.3. The *Gaia* RVS radial velocities are measured with a different instrument and technique and can therefore be considered fully independent of the measurements we provide here. Cross-referencing LAMOST and APOGEE was done using the *Gaia* DR3 `source_id` provided for each measurement by both LAMOST and APOGEE. When more than a single measurement was available for either LAMOST or APOGEE, we took the median of all measurements. This provides us with a total number of 23 900 765 sources for which we have both an XP and reference radial velocity measurement. We considered reference radial velocity measurements to be 'ground truth' and did not consider uncertainties in them. The reason is that our measurements will have uncertainties much larger than typical uncertainties in any of the reference catalogues. In our calibration we only considered sources that have no neighbours in *Gaia* within 2 arcseconds. The reason for this is that these sources tend to have blended spectra, due to the size of the spectral extraction window for XP spectra. Our models are not set up to account for blending, which means that radial velocity uncertainty and offset will be different for many of these sources. This further reduces the total number of sources used in calibration to 22 397 143.

2.4.2   Calibration model

To characterise the systematics in our radial velocity measurements, we adopted a Gaussian mixture model with a likelihood given by

$$
\begin{aligned}
\mathcal{L} \ \propto\ \prod_{i=1}^{N} &\left[\frac{1-f}{\sqrt{2\pi\sigma_i^2}} \ \exp\left(-\frac{\left[v_{\mathrm{rref}} - v_{\mathrm{rxp}} - b\right]^2}{2\sigma_i^2} \right) \right. \\
&\left. + \frac{f}{\sqrt{2\pi\sigma_{\mathrm{out}}^2}} \ \exp\left(-\frac{\left[v_{\mathrm{rxp}} - y\right]^2}{2\sigma_{\mathrm{out}}^2} \right) \right],
\end{aligned}
\tag{2.5}
$$

with f the outlier fraction, σ_i the radial velocity uncertainty, v_{rref} the reference radial velocity measurement, v_{rxp} the XP radial velocity, b the systematic offset between the reference and XP radial velocities, σ_{out} the standard deviation of the outlier population, and y the offset of the outlier population. The uncertainties on radial velocity measurements (σ_i) are described by

$$
\sigma_i = \sqrt{(a\sigma_{\mathrm{m}})^2 + c^2},
\tag{2.6}
$$

with a the underestimation factor on the uncertainties, σ_{m} the uncertainty determined from the posterior probability, and c the noise floor parameter.

We used bins in $\mathrm{BP} - \mathrm{RP}$ colour, apparent G magnitude, and extinction to fit for the free parameters f, a, b, c, σ_{out}, and y. We used 20 equally spaced bins in the range $5 \leq G \leq 17.65$ and 40 bins in the range $-0.3 \leq \mathrm{BP} - \mathrm{RP} \leq 5$. For extinction, we used bins with a width of $\Delta\mathrm{E(B-V)} = 0.1$. Additionally, we used two bins in $\log g$ for sources with $\mathrm{BP} - \mathrm{RP} \geq 1.6875$, with a divide at $\log g = 3.5$. We used the $\log g$ measurements from Zhang et al. (2023) for this purpose. The split in $\log g$ was used because we observed a high degree of systematic offset in the radial velocities we measured between dwarfs and giants at these colours. A description and justification for this split in $\log g$ is given in Appendix 2.B. We required at least 64 sources in a particular bin for fitting, with a maximum of $100\,000$, above which we selected $100\,000$ sources from the sample at random for computational efficiency. We ran the same calibration procedure using ten equally spaced bins in the range $5 \leq G \leq 17.65$ and 20 bins in the range $-0.3 \leq \mathrm{BP} - \mathrm{RP} \leq 5$ (i.e. using bins twice the default size). This ensured that we had calibrations for most sources, even in sparsely populated areas of the colour-magnitude space. If there were still not enough sources in the colour-magnitude bin for a particular source, we did not apply calibration.

2.4.3   Fitting of the model

To estimate the parameters in our calibration model we used the Markov chain Monte Carlo (MCMC) implementation in `emcee` (Foreman-Mackey
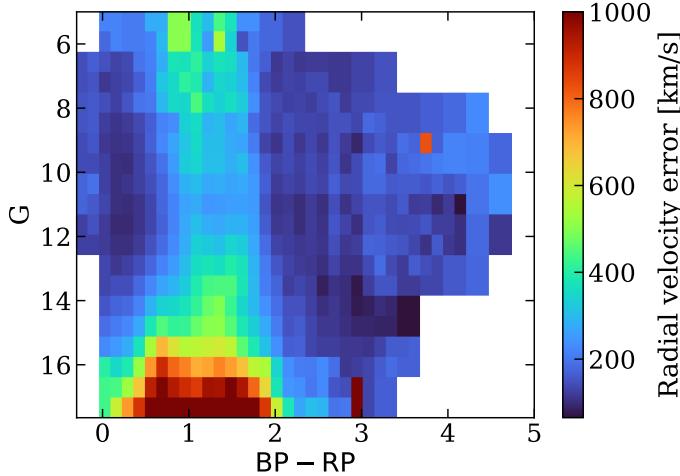
Figure 2.2: Median calibrated uncertainties as a function of colour and apparent magnitude for sources with E(B−V) < 0.1.

et al. 2013). We used flat priors throughout, except for σ_{out}, for which we used a log-uniform prior. Our MCMC approach is as follows: we initialised 64 walkers that we first propagated for 1000 steps to explore the parameter space. To avoid walkers getting stuck in local minima, we rejected walkers that finished with a log-likelihood outside 8.4 of the maximum log-likelihood over all walkers. The value 8.4 ensures that 99% of the walkers would remain if they traced a 6D Gaussian distribution. Another 1000 steps were performed with, again, 64 walkers that were drawn randomly from the last 100 steps of the walkers that remained from the previous run. The last 800 steps of this run were used to compute the medians of the free parameters.

2.4.4   Results of radial velocity calibration

The number of calibrated radial velocities is 123 835 034 out of a total of 125 145 490. This means that calibration was performed for ∼ 99% of our radial velocity measurements from Sect. 2.3.

Here we present the radial velocity calibration results for low extinction (E(B−V) < 0.1) sources. Results for higher extinction sources are similar, unless specified. We show the calibrated uncertainties (see Eq. 2.6) on the radial velocities measured from the XP spectra in Fig. 2.2. In the region BP − RP ≥ 1.6875, where we used two bins in log g, we took the source-number average for each bin between the giants and dwarfs. In general, we can see that the lowest uncertainties are obtained from blue and red sources.
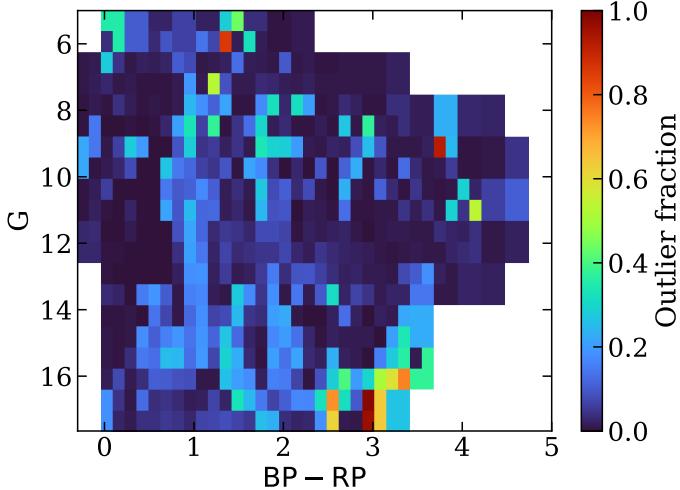
Figure 2.3: 2D histogram of the outlier fraction as a function of colour and apparent magnitude for sources with $E(B-V) < 0.1$.

We observe higher uncertainties for $1 \lesssim BP - RP \lesssim 2$ and the uncertainty generally increases for faint sources. In addition, the figure shows that uncertainties down to ~ 100 km s^{-1} are possible for red and blue sources. The reason why uncertainties are relatively high for $1 \lesssim BP-RP \lesssim 2$ is that there are few spectral features in the XP spectra for those sources. Without strong spectral features such as the Balmer lines and molecular absorption bands (see Fig. 2.1), fitting for a radial velocity becomes less precise.

In Fig. 2.3 we show the outlier fraction as a function of colour and magnitude. The outlier fraction tends to be low (smaller than 0.1), with a few regions containing notably more outliers. For higher extinction, the outlier fraction increases substantially, which we show in Fig. 2.16. In Appendix D we also show the underestimation factor, offset, and noise floor for sources with $E(B-V) < 0.1$.

Calibration has been applied to all sources when available. A histogram of the calibrated radial velocity uncertainties along with their cumulative distribution is included in Fig. 2.4. The median uncertainty on our calibrated radial velocity measurements is about 772 km s^{-1}, but extends all the way down to below 100 km s^{-1}. Sources with a low radial velocity uncertainty tend to be either blue ($BP - RP \lesssim 0.7$) or red ($BP - RP \gtrsim 2$).
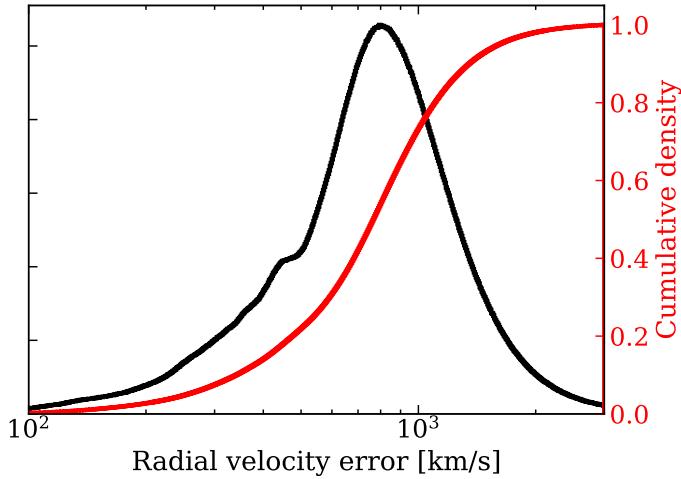
Figure 2.4: Histogram of the calibrated radial velocity uncertainties (in black) and the cumulative distribution (in red).

2.5   Random forest classifier

Although we have a general indication of the reliability of individual measurements from the outlier fraction parameter determined from our calibration model, a quality parameter determined on a source-to-source basis is important to avoid unreliable measurements in the final catalogue. We did this making use of a RFC.

The definition for a bad measurement we used is $\Delta v_r/\sigma_{vr} > 3$, with Δv_r the difference between our calibrated measurement and the reference measurement and σ_{vr} the corresponding calibrated uncertainty. For sources for which we had reference radial velocities, we used tenfold cross validation to predict the bad measurement probability. This ensures that the source for which we predicted a bad measurement probability is never part of the training set. For the remaining sources, we trained the RFC on all sources with reference radial velocity measurements. We took care to avoid information leaking from the training parameters to the radial velocities by excluding parameters like the sky coordinates and absorption. We used the `scikit-learn` RFC with 100 estimators (Pedregosa et al. 2011) and the following parameters for training:

- Reduced χ^2 of our best-fit model

- T_{eff}, $\log g$, and [Fe/H] of the best-fit model for radial velocity

- Extinction corrected BP − RP colour of the source

- Skewness of the radial velocity posterior (see Eq. 2.3)

In addition, we used the following columns provided by the *Gaia* archive (see the *Gaia* documentation[3] for column descriptions):

- `phot_g_mean_mag`

- `phot_xp_mean_mag`

- `xp_n_transits`

- `xp_n_blended_transits`

- `xp_n_contaminated_transits`

- `xp_n_measurements`

- `xp_standard_deviation`

- `xp_chi_squared/xp_degrees_of_freedom`,

where 'xp' indicates that we used the corresponding column of both BP and RP. We found that the extinction corrected colour is the most important out of these, with a feature importance of 0.13. The other columns have a similar importance (between about 0.4 and 0.8), except for the blended and contaminated transits, which have low importance at $\lesssim 0.2$. This procedure provides us with the likelihood that a particular measurement is unreliable, which we refer to as the `bad_measurement` parameter.

To verify the effectiveness of the random forest, we looked at the outlier fraction as a function of this `bad_measurement` parameter. We find good agreement with a one-to-one relation between the two, indication a successful classification.

2.6   Validation

We have already discussed the results from our spectral analysis and calibration in Sects. 2.3.1 and 2.4.4. In addition, we had access to the quality parameter `bad_measurement` described in Sect. 2.5. Using these earlier results, we focus in this section on validating that our measurements indeed measure the radial velocity and evaluate the reliability of our reported uncertainties.

To demonstrate that we are indeed measuring radial velocities from the XP spectra, we included Fig. 2.5, in which we binned the XP radial velocities
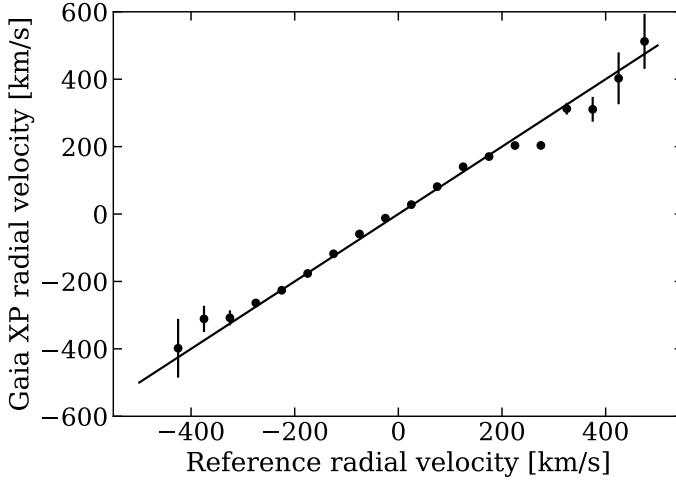
Figure 2.5: Median of the calibrated binned XP radial velocities as a function of the reference radial velocity measurements. These sources have an outlier fraction below 0.2, `bad_measurement` < 0.1, and calibrated uncertainty below 300 km/s. The solid line is the bisection.

based on their reference measurements. The uncertainties on the individual bins were calculated as

$$\sigma = \frac{1}{N} \sqrt{\frac{\pi}{2}} \left(\overline{|v_{\rm r} - \overline{v_{\rm r}}|^2} \right)^{1/2},$$ (2.7)

where the overline indicates that the mean is taken, $v_{\rm r}$ is the radial velocity, and N is the number of measurements in the bin. The measurements clearly follow the bisection with the reference radial velocity measurements, demonstrating that we indeed measured stellar radial velocities. To ensure we are not seeing the result of a correlation between radial velocity and position in the colour-magnitude diagram picked up by our calibration model, we also performed this analysis for each colour-magnitude bin separately in Fig. 2.15. To further demonstrate our ability to constrain radial velocities, we plot in Fig. 2.6 the sky projection of the median XP radial velocities compared to *Gaia* RVS. The quality cuts we applied to our catalogue are

- `rv_err` < 300 km s^{-1}

- `CMD_outlier_fraction` < 0.2

---

[3]https://gea.esac.esa.int/archive/documentation/GDR3/Gaia_archive/chap_datamodel/sec_dm_main_source_catalogue/ssec_dm_gaia_source.html

https://gea.esac.esa.int/archive/documentation/GDR3/Gaia_archive/chap_datamodel/sec_dm_spectroscopic_tables/ssec_dm_xp_summary.html#xp_summary-
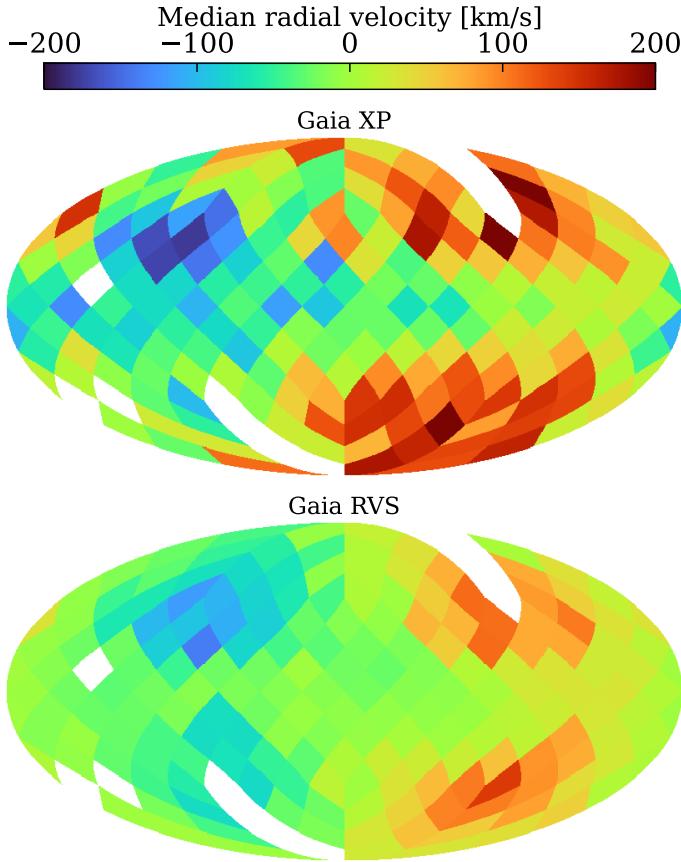
Figure 2.6: Median XP radial velocity as a function of sky position in Galactic coordinates (left) and the same but using *Gaia* RVS reference measurements (right). To highlight halo stars, we only show low metallicity stars ([Fe/H] ≤ -1). This increases the radial velocity amplitude as a function of position on the sky. In both maps we can recognise the dipole caused by the solar motion. The other selections of the sources in this figure are described in the main text. The sources are the same in both panels, and we only show colour bins with at least ten measurements.
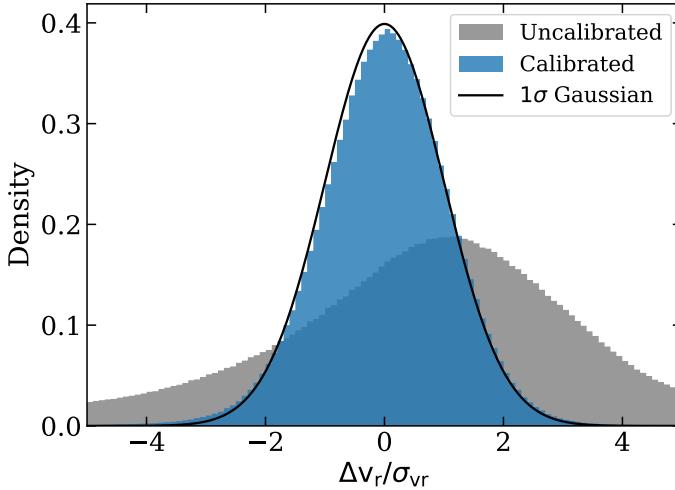
Figure 2.7: Difference in radial velocity over the uncertainty of both our calibrated and uncalibrated results compared to reference measurements. The solid line is a Gaussian distribution with a standard deviation of 1. The sample used to make this figure has calibrated radial velocity uncertainties of $< 300\ \mathrm{km\,s}^{-1}$.

- `bad_measurement` < 0.1.

In addition, we used the catalogue from Zhang et al. (2023) and selected sources with [Fe/H] ≤ -1 to mainly select halo stars and `quality_flags`$\leq$ 8. This allowed us to see the dipole caused by the solar motion in both the XP and RVS maps. The dipole disappears at the Galactic plane due to the sample being dominated by non-halo stars in that region.

To evaluate if our radial velocity uncertainties are accurate, we created a histogram of the radial velocity difference of our measurements compared to the reference measurements over the uncertainty (Fig. 2.7). We determined the standard deviation of this distribution as

$$\sigma = 1.4826 \cdot \mathrm{median}\left(\Delta v_r/\sigma_{vr} - \widetilde{\Delta v_r/\sigma_{vr}}\right), \qquad (2.8)$$

with $\Delta v_r/\sigma_{vr}$ the difference between our radial velocity and the reference one over the uncertainty and $\widetilde{\Delta v_r/\sigma_{vr}}$ the median of the same quantity. The standard deviation is about 1.03, which means that our reported uncertainties are typically accurate to a few percent. This is in contrast to the uncalibrated measurements, which we also plot in Fig. 2.7 and which show both a significant offset and a significant uncertainty underestimation.

2.7 Catalogues

We have published two catalogues along with this paper. The Main Catalogue is the catalogue we recommend for the general user. It includes only relatively precise measurements with low chance of being erroneous. For completion, we also published the Extended Catalogue, which includes all the measurements we obtained. The catalogues are available through an online table[4]. The columns included are described in Table 2.4.

2.7.1 Main Catalogue

To ensure we only published relatively high quality measurements in our Main Catalogue, we applied the following selections:

- E(B−V) < 0.5

- `rv_err` < 300 $\mathrm{km\,s^{-1}}$

- `CMD_outlier_fraction` < 0.2

- `bad_measurement` < 0.1.

The Main Catalogue contains 6 367 355 sources that pass the quality cuts. About 23% of these sources have no previous measurement in *Gaia* RVS, by far the biggest catalogue in our magnitude range. This means the Main Catalogue contains relatively accurate and precise radial velocity measurements for about 1.5M sources that have no previous measurement available. In Fig. 2.8 we show the colour-magnitude density of our Main Catalogue.

2.7.2 Extended Catalogue

For completion, we also provide our entire catalogue, without any quality cuts, which we refer to as our Extended Catalogue. The only exception is that we still only published sources with E(B−V) < 0.5, since we deemed most higher extinction measurements to be unreliable. To assist the user in making use of this catalogue, we provided additional parameters for all sources alongside those provided for the main catalogue, which is a subset of the Extended Catalogue. In the case of a star occupying a point in parameter space with insufficient reference measurements to perform calibration, we still report our XP radial velocity measurement, only without any calibration performed. In those cases we report the `CMD_outlier_fraction`, `offset`, `underestimation_factor`, and `noise_floor` parameters as NaN

---

[4]`https://doi.org/10.5281/zenodo.10043238`
  or at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via `http://cdsweb.u-strasbg.fr/cgi-bin/qcat?J/A+A/`

Table 2.4: Description of the fields included in the final published catalogue. The first five rows (above the horizontal line) are included in the Main Catalogue. The remaining rows only appear in the Extended Catalogue.

| Column	Heading	Description
1	source_id	*Gaia* DR3 source_id
2	ra	*Gaia* DR3 right ascension
3	dec	*Gaia* DR3 declination
4	rv	Calibrated XP radial velocity
5	rv_err	Calibrated XP radial velocity uncertainty
6	offset	Offset applied to the original radial velocity measurement during calibration
7	underestimation_factor	Factor (a) applied to the measured uncertainties according to Eq. 2.6
8	noise_floor	Noise floor (c) applied to the measured uncertainties according to Eq. 2.6
9	CMD_outlier_fraction	Fraction of stars in colour-mag-extinction($-\log g$) range that are considered outliers
10	bad_measurement	Probability of a bad measurement based on the RFC
11	warning	Warning flag to indicate potentially problematic radial velocity measurements
12	teff	T_{eff} of the best-fit model for radial velocity
13	logg	$\log g$ of the best-fit model for radial velocity
14	feh	[Fe/H] of the best-fit model for radial velocity
15	reduced_chi_squared	Reduced χ^2 of the best-fit model for radial velocity
16	skew	Skewness of the radial velocity posterior probability distribution
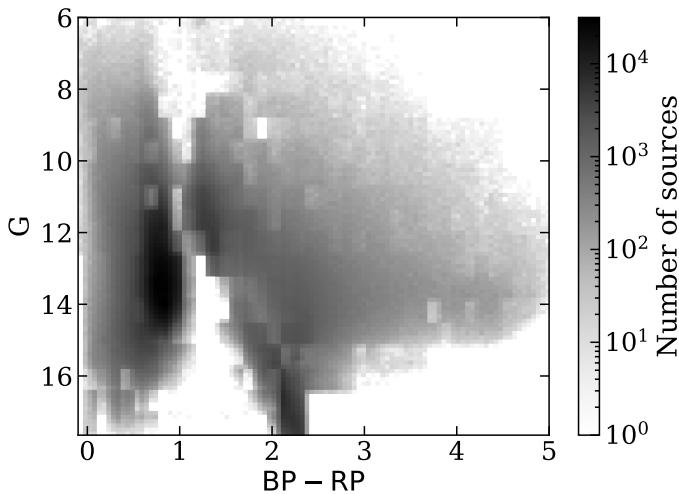
Figure 2.8: Colour-magnitude diagram for the sources in our Main Catalogue.

values. In Fig. 2.9 we show the colour-magnitude diagram of the sources appearing in our Extended Catalogue. Since there are many more caveats with this dataset compared to the Main Catalogue, we provide the user with the `warning` parameter that is supplied as a bitmask. If one of the following conditions was met, the corresponding bit was set to 1.

1. No calibration applied (0001)

2. Neighbour in *Gaia* within 2 arcsec (0010)

3. `CMD_outlier_fraction` > 0.2 (0100)

4. `bad_measurement` > 0.1 (1000)

2.8   Finding hypervelocity stars

Having presented our Main Catalogue, we now use it to investigate the science case of HVSs.

These stars can have velocities well in excess of $1000 \; \mathrm{km \, s^{-1}}$ (Koposov et al. 2020), making them much faster than stars belonging to other populations. These stars are ejected from the Galactic Centre following a dynamical encounter with our central massive black hole, Sgr A* (Brown 2015). Their identification has proven difficult with only a few dozen promising candidates (Brown et al. 2014) and a single star that can be unambiguously traced back to the centre of our Galaxy (Koposov et al. 2020). Our
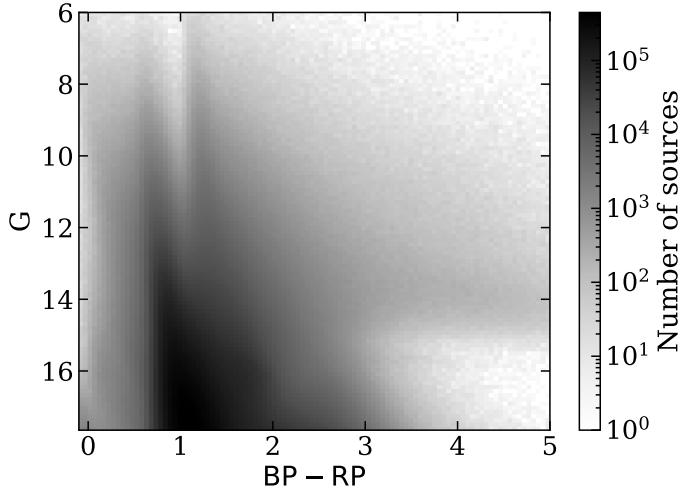
Figure 2.9: Colour-magnitude diagram for the sources in the Extended Catalogue.

new catalogue of radial velocities can facilitate blind searches for additional HVSs, helping unravel the dynamics and properties of stars in the centre of our Galaxy as well as providing valuable information about the Galactic potential (e.g. Rossi et al. 2017; Evans et al. 2022c).

For the purpose of searching for HVSs it is of interest to determine if we can still obtain reliable radial velocity measurements for extremely high velocity stars. To date, S5-HVS1 is the fastest unbound star known in our Galaxy, with a total velocity in the Galactic frame of 1755 ± 50 km s^{-1} and a heliocentric radial velocity of 1017 ± 2.7 km s^{-1} (Koposov et al. 2020). The calibrated radial velocity we measure is 799 ± 273 km s^{-1}, which is consistent with the reference radial velocity measurement of S5-HVS1 within $\sim 0.8\sigma$. The bad_measurement parameter from the RFC is 0.02 for S5-HVS1, indicating a reliable measurement. This establishes that our results are still accurate for extremely high radial velocity sources.

To evaluate the general effectiveness of the selection of high radial velocity star candidates from our Main Catalogue, we produced Fig. 2.10. In the figure we only selected stars from our Main Catalogue whose 3σ lower limit on v_r is at least 300 km s^{-1} (i.e. $v_r > 300 + 3 \cdot \sigma_{vr}$ km s^{-1}) and plot those for which reference radial velocity measurements were available. Although the distribution of our selection still peaks at 0 km s^{-1}, we can see a very significant over-density of high radial velocity sources. Most of the sources in this selection do not have reference radial velocity measurements and, as Fig. 2.10 shows, the majority of them will not have high radial veloci-
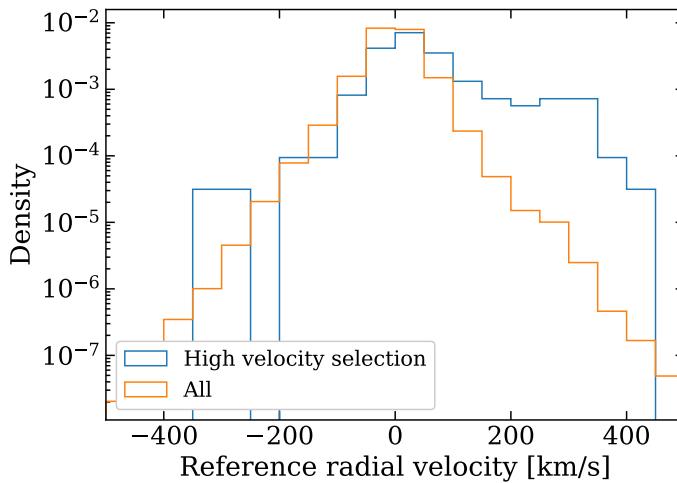
Figure 2.10: Density of the reference radial velocity distribution of all sources against those selected by $v_\mathrm{r} > 300 + 3 \cdot \sigma_\mathrm{vr}$ km s^{-1} in our Main Catalogue and for which reference radial velocity measurements are available.

ties. However, the selection of HVS candidates for follow-up radial velocity surveys can be viable. Only 3175 sources out of our Main Catalogue of 6.4M sources passed the selection of $v_\mathrm{r} > 300 + 3 \cdot \sigma_\mathrm{vr}$ km s^{-1}. The number of candidates could be further reduced by using, for example, astrometric information to constrain the orbits. Follow-up observations to precisely measure their radial velocities will be proposed for the most promising of these HVS candidates.

2.9   Discussion

Despite the challenges, we have shown that radial velocities can be obtained from *Gaia* XP measurements to a precision of better than ~ 300 km s^{-1} for stars as faint as G = 17.65. Sect. 2.9.1 discusses possible improvements to the methods presented in this paper, in Sect. 2.9.2 we provide prospects for *Gaia* DR4 and the improvements that we might expect with its release regarding radial velocities from XP spectra, and lastly in Sect. 2.9.3 we discuss science cases for XP radial velocities in *Gaia* DR4.

2.9.1   Improvements to the method

Our current approach is only viable for low to intermediate extinction sources, due to our implementation of an initial guess for T_eff. This approx-

imation breaks down for high extinction sources as mentioned in Sect. 2.3. Practically, this means that our results are not reliable for sources with $E(B-V) \gtrsim 0.5$ and we do not report our results for those sources. The issue can be mitigated by for example, using a larger range in T_{eff} during the fitting procedure for high extinction sources, or by fitting every model for all sources.

Additionally, fitting for extinction rather than relying on a 2D extinction map would allow for more accurate measurements, because for individual sources the 2D extinction map is only an estimate of the actual line-of-sight extinction. Including extinction in the fitting procedure is possible, but is also computationally very expensive, which is why we opted to use the 2D map instead.

The analysis could be further improved by choosing the fitting wavelength range on a source-to-source basis: practically, one would choose for each source the wavelength regions that hold the most spectral information. Doing this would improve the precision of the radial velocity measurements. Here, we instead used the same wavelength ranges throughout.

Alternatively to the modelling presented in this work, one could forward-model the spectral coefficients directly. Provided that the design matrix and the model of the instrument are accurate, this should give more precise results.

Improving upon the method is required if the goal is to obtain reliable radial velocity measurements for a revolutionary large set of sources. Even though we started out with XP spectra to about 220M sources, the Main Catalogue only includes around 6.4M radial velocities (or about 3% of XP sources), with an additional ~ 119M in the Extended Catalogue. Understanding and correcting for systematic effects remains the most challenging aspect.

2.9.2   *Gaia* Data Release 4

In *Gaia* DR4, XP spectra will be published for about 2B sources, in addition to individual epoch spectra of said sources. Since this is orders of magnitude higher than any current radial velocity catalogue, the potential scientific return on a well-optimised method of radial velocity analysis would be very high.

It is unknown how large the improvement will be in radial velocity accuracy and precision from *Gaia* DR3 to DR4 using the methods presented here. The reason is that systematics are a very large factor in the radial velocity uncertainty. We might consider the noise floor in Eq. 2.6 to be the intrinsic systematic uncertainty caused by imperfect calibration of XP spectra in *Gaia* DR3. If we assume that these systematics are resolved in *Gaia* DR4, we can provide an outlook for the performance of our method

when applied to *Gaia* DR4. When we ignore the noise floor, the number of sources with radial velocity uncertainties < 300 km s^{-1} in DR3 approximately doubles to ~ 17M. In addition, there would be about 1M sources with radial velocity uncertainties of < 100 km s^{-1}. The smallest uncertainties that might be achieved are expected to be of the order of 50 km s^{-1}. Although DR4 will mostly include fainter sources than the current limit of $G < 17.65$, the S/N for a given magnitude will also improve. *Gaia* DR4 will provide XP spectra to about 9 times as many sources as DR3. A rough approximation of the final number of sources with a particular quality in *Gaia* DR4 is thus 9 times the number in DR3. This would imply that the XP spectra in *Gaia* DR4 could provide us with ~ 153M and ~ 8M measurements with uncertainties better and 300 and 100 km s^{-1}, respectively. Without systematic uncertainty due to the noise floor, these stars would be mainly red (BP $-$ RP $\gtrsim 2$) and blue (BP $-$ RP $\lesssim 0.7$) in colour.

2.9.3   Science cases in *Gaia* Data Release 4

Having discussed improvements to both the methods and data with the next data release of *Gaia*, we now look at prospects for two specific science cases in *Gaia* DR4: dark companions and HVSs.

Dark companions refer to binary systems in which one of the components emits little to no light in the photometric band used to observe them. These dark companions, such as black holes, can be identified from low-resolution spectra if enough epochs are available over a sufficient time span. The photocentre of *Gaia* BH1, for instance, has a radial velocity amplitude of about 130 km s^{-1} (El-Badry et al. 2023; Chakrabarti et al. 2023), far larger than the typical uncertainty in *Gaia* RVS of only a few km s^{-1} (Katz et al. 2023). With the release of epoch XP spectra in *Gaia* DR4, searches for dark companions will become possible in the full *Gaia* catalogue of ~ 2B sources. Compared to the astrometric time series, radial velocities have the advantage of being distance independent, thus allowing for a larger search volume. Also, in comparison to *Gaia* RVS, the XP radial velocities have the advantage of being deeper and therefore covering a larger volume. *Gaia* RVS will have a limiting magnitude of $G_{\mathrm{RVS}} \sim 16$ in *Gaia* DR4, compared to the limiting magnitude of $G \sim 20.7$ for the XP spectra. We assumed the two photometric bands to be similar[5] and approximated the magnitude difference as 4.7. From the magnitude difference, we can calculate the volume ratio as

$$\frac{V_{\mathrm{XP}}}{V_{\mathrm{RVS}}} = 10^{0.6 \cdot \Delta m} \approx 660, \tag{2.9}$$

with V_{XP} the volume covered by XP spectra, V_{RVS} the volume covered by RVS radial velocities, and Δm the difference in limiting magnitude. The

---

[5] https://www.cosmos.esa.int/web/gaia/dr3-passbands

effective volume covered by XP spectra is thus about 660 times as large as that covered by RVS radial velocities. Depending on the final precision and accuracy that can be achieved, dedicated higher-resolution observations might be required to confirm systems with possible dark companions identified from *Gaia* XP radial velocities.

In addition to finding dark companions, *Gaia* DR4 XP radial velocities could support the search for HVSs. Because of the high intrinsic velocities of these stars, large uncertainties are less problematic. As demonstrated in Sect. 2.8, the contamination of a selection of extremely high XP radial velocity sources is substantial in our Main Catalogue. With the improved analysis suggested in Sect. 2.9.1, in combination with a reduction in systematics that we expect in *Gaia* DR4, the contamination will decrease. This will allow for more effective follow-up campaigns to identify new HVSs. As mentioned above, the advantage of using *Gaia* XP spectra is that the effective volume is much larger than that of the *Gaia* RVS catalogue.

Both for dark companions and HVSs, XP radial velocities will be most effective in identifying them for red ($BP - RP \gtrsim 2$) and blue ($BP - RP \lesssim 0.7$) sources, since these sources have the lowest uncertainties in XP radial velocity. This is not expected to change from *Gaia* DR3 to DR4, since it is inherent to the radial velocity information contained within the XP spectra.

2.10   Conclusion

As a proof of concept, we have clearly demonstrated that *Gaia* XP spectra can be used to measure radial velocities. Along with this paper we publish the Main Catalogue, which contains reliable and precise radial velocity measurements for about 6.4M sources, 23% of which have no previous radial velocity measurements in *Gaia*. In addition, we publish the Extended Catalogue, which contains all ~ 125M sources for which we have obtained a radial velocity measurement. This constitutes $\sim 84\%$ of sources with *Gaia* XP spectra and $E(B-V) < 0.5$. The extended catalogue, however, contains a significant number of unreliable measurements and should therefore only be used with caution.

In general, sources with $BP - RP \gtrsim 2$ and $BP - RP \lesssim 0.7$ tend to give the most precise radial velocity measurements in our catalogue, down to uncertainties of ~ 100 km s^{-1}. In the future, we expect the most precise radial velocity measurements from *Gaia* XP spectra to have uncertainties of the order of 50 km s^{-1}.

Critically, this work has demonstrated the potential of measuring radial velocities for over 10^9 sources in *Gaia* DR4 using XP spectra. This would constitute an orders-of-magnitude increase compared to the largest current catalogue. However, the methods presented here should be further improved

to fully exploit the scientific content available to us.

Acknowledgements

---

[6]`https://extinction.readthedocs.io/en/latest/`

Appendix

2.A T_{eff} initial guess

For computational efficiency, we made an initial guess for the T_{eff} of each source and only considered models within 500K of this initial guess, as described in Sect. 2.3.

We made the initial guess for T_{eff} based on the BP $-$ RP colour and 2D extinction along the line of sight. First we corrected the colour of the source for extinction using the extinction law provided by $Gaia$[7]. This is only an approximation, since we do not know the intrinsic colour of a source affected by extinction and use the observed colour instead. By analysing test sources for the entire T_{eff} grid range, we used an empirical exponential fit to the relation between BP $-$ RP and T_{eff}. Only for colours BP $-$ RP $\lesssim 0$ did we find a turn-off from the exponential relation, for which we fitted a linear function. The relation is given by

$$T_{\mathrm{eff}}(K) = \begin{cases} 6721 \cdot \exp\left(-0.95 \cdot c\right) + 2617 & \text{if } c > 0.01 \\ -18796 \cdot c + 9431 & \text{if } c \leq 0.01 \end{cases}, \qquad (2.10)$$

where c denotes the extinction corrected BP $-$ RP colour. If the initial guess for T_{eff} was outside the temperature range of our models (see Table 2.1), we did not attempt to fit the source and no radial velocity was obtained.

2.B $\log g$ systematics

For red sources we observed strong systematics in the radial velocity we measured, based on the surface gravity of the stars. We demonstrate this in Fig. 2.11, where we used the surface gravity measurements from Zhang et al. (2023) to separate dwarfs from giants. To investigate the origin of this systematic offset we used $\log g$ and T_{eff} measurements from APOGEE in comparison to our best-fit values. In Fig. 2.12 we show how the radial velocity offset relates to errors in the parameter estimation for dwarfs. We can see that dwarfs tend to get assigned radial velocities that are too high. This also varies as a function of $\Delta \log g$ and ΔT_{eff}. In particular, if the $\log g$ of our best-fit model labels a dwarf as a giant a positive offset is introduced, which can be seen from the colour gradient towards negative $\Delta \log g$.

To make this figure and have sufficient sources, we used a much larger range of colour and magnitude than we do for single bins in the calibration. The spread in offsets for individual bins is smaller and therefore less problematic. Fig. 2.13 shows the same, but now only for giants. In general we

---

[7]https://www.cosmos.esa.int/web/gaia/edr3-extinction-law
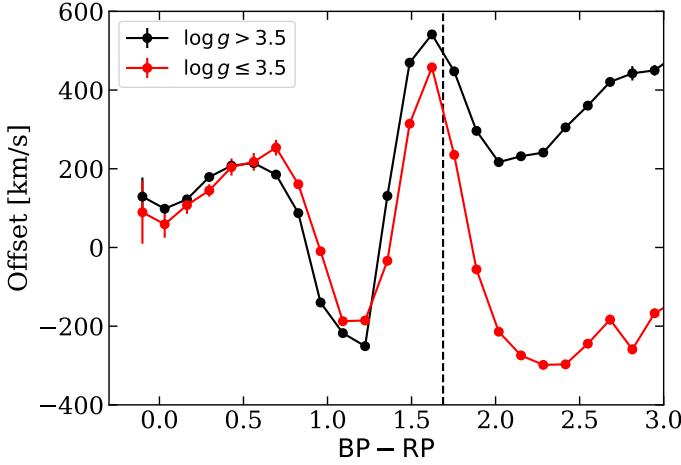
Figure 2.11: Offset between the uncalibrated *Gaia* XP radial velocities and reference ones as a function of BP − RP colour. We split the dwarfs and the giants using the log g measurements from Zhang et al. (2023). The dashed vertical line shows the limit above which we used two bins in log g for our calibration model. The figure only includes sources with E(B−V) < 0.1 and 12 < G < 13.
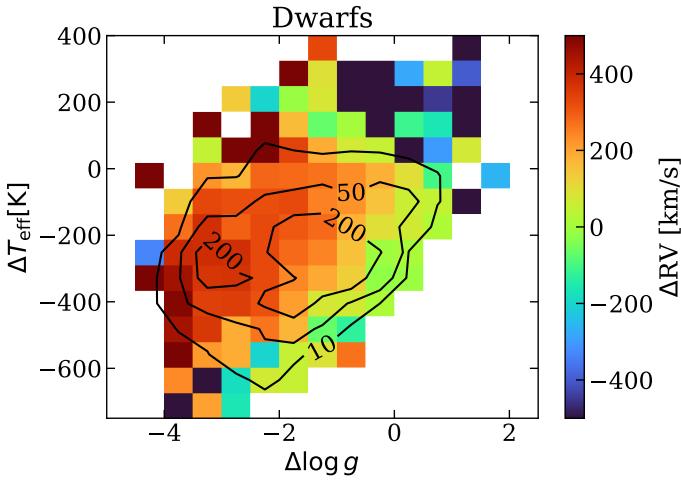
**Figure 2.12:** Radial velocity offset for dwarfs (log g > 3.5 according to Zhang et al. (2023)) as a function of the difference in the best-fit log g and T_{eff} we obtain versus those from APOGEE. The contour lines give the underlying source density. The stars in the sample have 10 < G < 16, 2 < BP − RP < 2.5, and E(B−V) < 0.1.
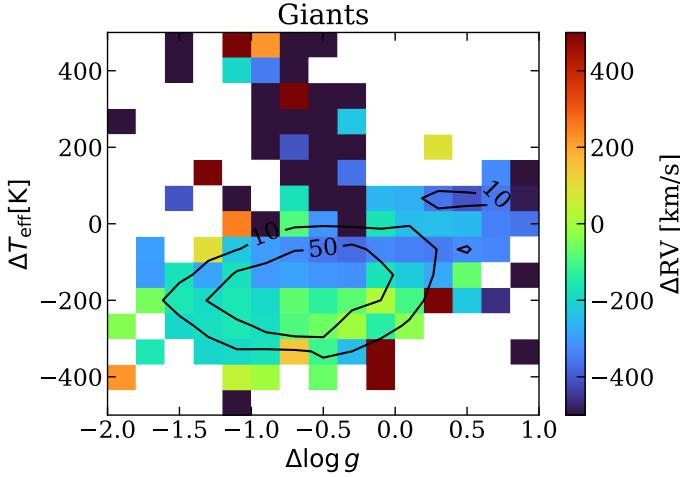
Figure 2.13: Same as Fig. 2.12, but now only for giants.

can see that the offset tends to be much lower for giants. It is possible that it is caused by the basis function representation in *Gaia*, which undergoes optimisation and might lead to systematic differences in the translation of giant and dwarf spectra. Fortunately, we could effectively mitigate the effects of this bias, whatever its origin. To evaluate our treatment of the observed offset described in Sect. 2.4.2, we recreated Fig. 2.11 for our calibrated sample, which we show in Fig. 2.14. The figure clearly demonstrates that the offset between dwarfs and giants for very red sources has been effectively removed. Colour-dependent systematics are still visible in the offset, however, particularly around $BP - RP \sim 1.5$. This effect is explained in Sect. 2.D and is related to the `CMD_outlier_fraction`.

2.C   Further validation

To demonstrate that the calibration we performed does not introduce an apparent radial velocity sensitivity, we included Fig. 2.15. This figure gives the median XP radial velocity for bins in reference radial velocity within the Main Catalogue, where each curve is constructed using stars within one colour-magnitude diagram bin. The one-to-one slope demonstrates that within individual colour-magnitude bins we are sensitive to radial velocity. This excludes the possibility that a correlation between colour-magnitude and radial velocity is introducing an apparent radial velocity sensitivity.
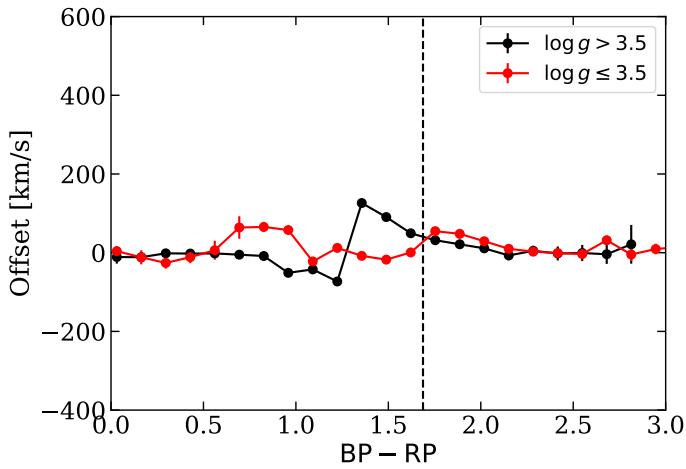
Figure 2.14: Same as Fig. 2.11, but after calibration is applied to the bin $11.9575 < G < 12.59$.
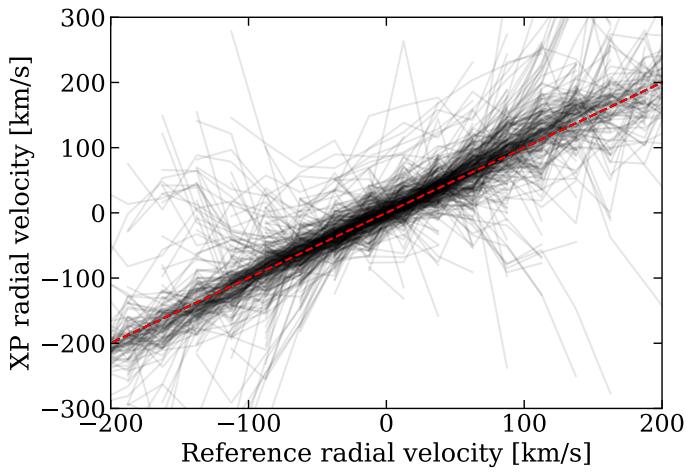

Figure 2.15: Median XP radial velocity as a function of bins in reference radial velocity. The dashed red line gives the bisection and the transparent black lines the data for individual bins in the calibration.
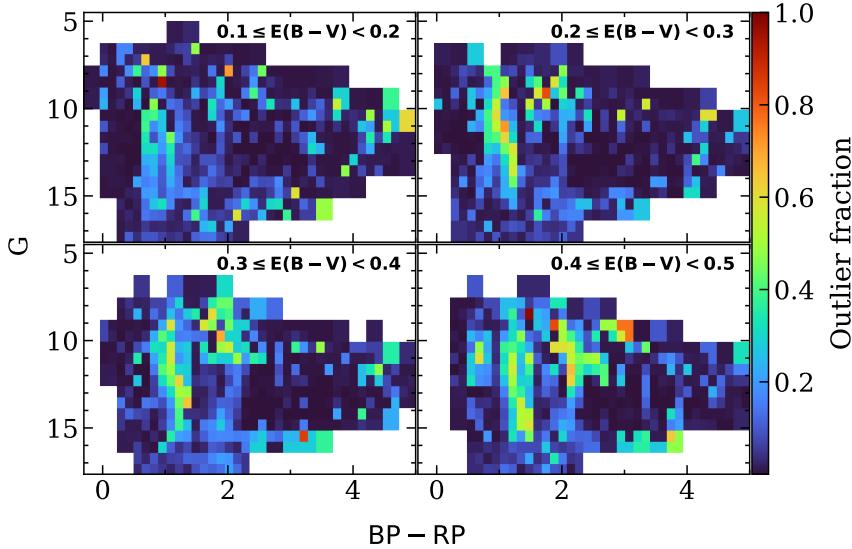
Figure 2.16: Outlier fractions as a function of colour and magnitude for the remaining E(B−V) bins not shown in the main text.

2.D   Markov chain Monte Carlo examples and results

In Fig. 2.16 we show the outlier fraction of the model described in Sect. 2.4.2 (Eq. 2.5) as a function of colour and magnitude for the higher E(B−V) bins and not shown in the main text.

In Fig. 2.17 we give the underestimation factor, offset, and noise floor for sources with E(B−V) < 0.1. The definition of the offset can be found in equation 2.5 and the definition of the underestimation factor and noise floor in equation 2.6.

To provide more insight into what our calibration model does, we show the MCMC results for a number of bins. We show the 11.9575 < G < 12.59 magnitude bin for three different colour bins indicated in Fig. 2.18. We note that the method does not work equally well for all bins, which affects the calibration. If we look at the middle and right panels, we see that the overall distribution is skewed. Correcting for the offset in that sample will still result in a median that is significantly higher or lower than 0, because of the skew. This is the residual effect we see in Fig. 2.14. Part of this can be explained by log g effects that are still present in this colour region. We chose not to attempt to correct for these effects, because for these intermediate colour sources, there is no clear split in offset between dwarfs and giants.
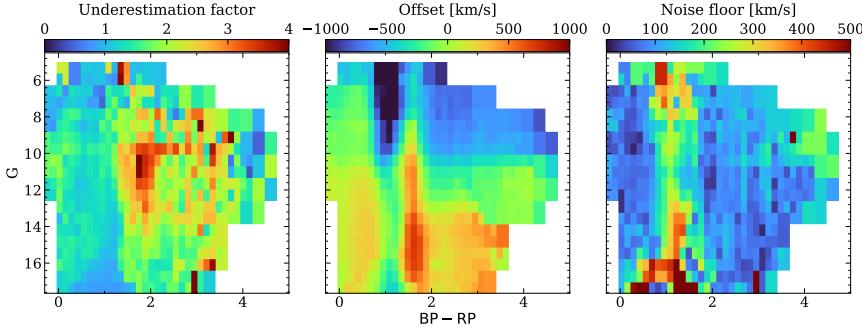
Figure 2.17: From left to right: Underestimation factor, offset, and noise floor for sources with E(B−V) < 0.1.
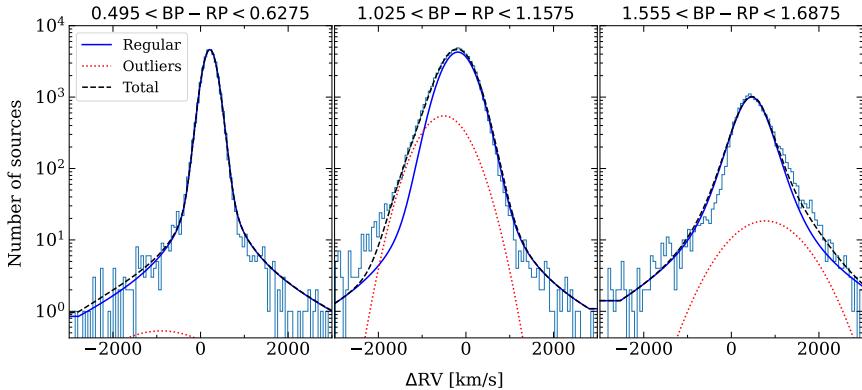
Figure 2.18: Histograms of the radial velocity difference between the raw XP measurement and the reference values. The three lines show the parameters found by the fit of the predicted distributions of the regular, outlier, and combined populations. The different panels show different bins in BP − RP colour, as indicated above the panels.

Instead a more elaborate strategy would have to be employed to correct for the remaining systematics.
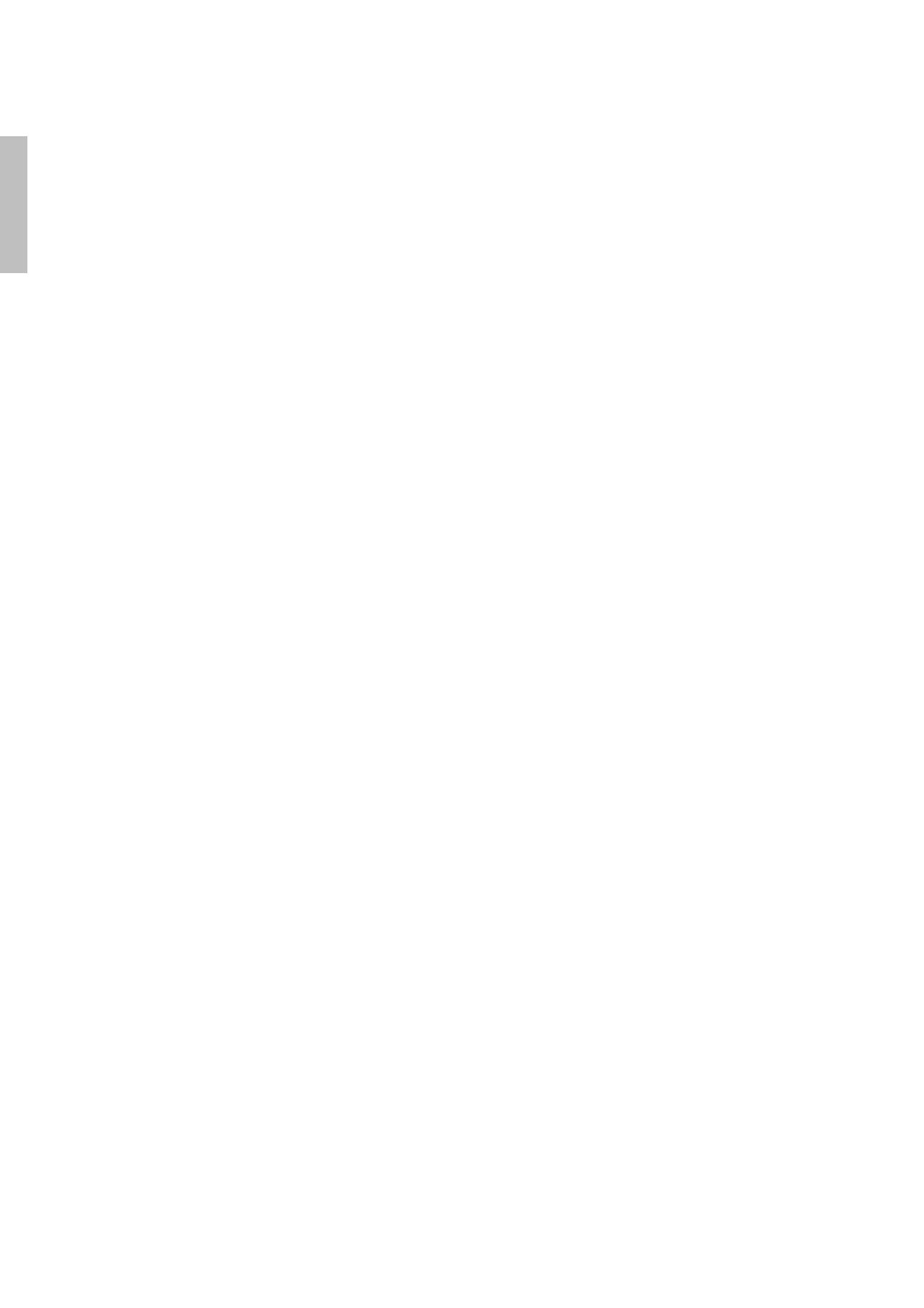
3 | IMPROVED CONSTRAINTS ON GALACTIC CENTRE EJECTION OF HYPERVELOCITY STARS BASED ON NOVEL SEARCH METHOD

Abstract

Hypervelocity stars (HVSs) are stars which have been ejected from the Galactic Centre (GC) at velocities of up to a few thousand $km\,s^{-1}$. They are tracers of the Galactic potential and can be used to infer properties of the GC, such as the initial-mass function and assembly history. HVSs are rare, however, with only about a dozen promising candidates discovered so far. In this work we make use of a novel, highly efficient method to identify new HVS candidates in *Gaia*. This method uses the nearly radial trajectories of HVSs to infer their distances and velocities based on their position and *Gaia* proper motion alone. Through comparison of inferred distances with *Gaia* parallaxes and photometry we identified 600 HVS candidates with G < 20 including the previously discovered S5-HVS1, out of which we obtained ground-based follow-up observations for 196 stars. As we found no new HVSs based on their radial velocity, we used detailed HVS ejection simulations to significantly improve previous HVS ejection rate constraints. In particular, the ejection rate of HVSs more massive than 1 $M_\odot$ cannot be higher than 10^{-5} yr^{-1} at 2σ significance. Additionally, we predict that there are 5–45 unbound HVSs in the complete *Gaia* catalogue (1σ interval), most of which will be main-sequence stars of a few $M_\odot$ at heliocentric distances of tens to hundreds of kpc. By comparing our results to literature HVS candidates, we find an indication of either a time-dependent ejection rate of HVSs or a non-GC origin of previously identified HVS candidates.

3.1   Introduction

The Galactic Centre (GC) is a highly complex environment, hosting both a Nuclear Star Cluster and a super-massive black hole called *Sagittarius A** (Sgr A*; e.g. Genzel et al. 2010). The combination of these two factors gives rise to a unique, high-energy environment within our Galaxy, which challenges our understanding of key physical processes. The origin of the S-star cluster, for instance, remains unknown to this day, because the strong tidal force in this region inhibits standard star formation from molecular clouds (Genzel et al. 2010). In addition, the initial-mass function (IMF) of stars near the GC has been a long-standing point of debate, whose resolution can give us an important clue as to the mass assembly hisotry of the GC. Some studies find it is consistent with the canonical IMF from Kroupa (2001) (Maness et al. 2007; Löckmann et al. 2010), while others find evidence for a top-heavy IMF in disc structures near Sgr A* (e.g. Paumard et al. 2006; Klessen et al. 2007; Bartko et al. 2010; Lu et al. 2013; von Fellenberg et al. 2022). The combination of high line-of-sight extinction and source crowding make the study of the GC challenging, requiring specialised instruments such as *GRAVITY* (Eisenhauer et al. 2011).

One promising means of better understanding our GC comes from hypervelocity stars (HVSs), which are stars ejected from the GC at extremely high velocities. They are believed to be ejected from the vicinity of Sgr A* (order of 10 AU), but can be observed in parts of the sky more accessible to study (Hills 1988; Yu & Tremaine 2003). This in turn allows for detailed stellar parameter measurements with existing observatories, and the study of wavelength ranges entirely inaccessible near the GC. It has also been suggested that HVSs could explain the existence of the S-star cluster, because the companion of the HVS in the progenitor binary is deposited at radii consistent with the S-star cluster (Gould & Quillen 2003). This migration of already-formed stars into the S-star cluster could therefore solve the issue of in-situ star formation in the strong tidal field near Sgr A* (e.g. Ghez et al. 2003, see, however, Habibi et al. (2017b)).

In recent years, the term HVS has been used inconsistently. For clarity, we define an HVS to be characterised exclusively by having been ejected from the vicinity of Sgr A*. Since ejections can occur at a range of velocities, some of these HVSs will still be bound to the Galaxy.

The discovery of HVSs has proved challenging in spite of their significant scientific potential. Only a few dozen promising candidates (e.g. Brown et al. 2014) and a single star that can be confidently traced back to the GC (Koposov et al. 2020) have been identified. The main challenges include their rarity and the difficulty in disentangling these stars from e.g. hyper runaway stars from the Galactic disc (e.g. Przybilla et al. 2008b; Kreuzer et al. 2020; Irrgang et al. 2021).

In this work we perform a targeted survey of HVS candidates. The candidates are identified using a novel, highly efficient selection method with *Gaia* Data Release 3 (DR3; Gaia Collaboration et al. 2016, 2023a), which means we have excellent sensitivity to HVSs from relatively few candidates. We perform ground-based follow-up observations for the most promising HVS candidates and use these observations in combination with sophisticated models to constrain the GC environment and the ejection of HVSs. We use these constraints to investigate the total population of HVSs accessible to *Gaia* in current and future data releases. We furthermore evaluate previously identified candidates using our observational constraints.

In Section 3.2 we describe how we select HVS candidates from *Gaia* DR3. In Section 3.3 we present our follow-up observations. In Section 3.4 we discuss the model and simulations, in addition to how we apply the observational selections to the models. In Section 3.5 we present constraints on the GC environment and HVS from our observations in combination with the simulations. In Section 3.6 we discuss our results and implications for existing HVS candidates as well as what we might expect from *Gaia* DR4. Finally, in Section 3.7 we provide closing remarks.

3.2   HVS candidate selection

The main challenge when searching for HVSs is that they are extremely rare; there are for instance no HVSs with velocity greater than 700 km s^{-1} in the 34M sources in *Gaia* DR3 with radial velocities, that could be robustly identified (Marchetti et al. 2022). HVS candidates selected from observational catalogues for follow-up observations are therefore typically strongly contaminated by non-HVSs. Here we present our new method to identify HVS candidates originating from the GC. In Section 3.2.1 we present the general principle we use to find HVSs, and in Section 3.2.2 we provide the precise selections on the *Gaia* catalogue for our observational campaign.

3.2.1   Radial trajectory

Let $\vec{R}$ be the vector from the GC to the star and $\vec{V}$ the star's velocity vector relative to the GC. We can write these down in terms of the observables

- $\vec{R_0}$: Vector from the GC to the Sun

- $\hat{n}$: Unit vector from the Sun to the source

- D: Distance from the Sun to the source

- $\vec{V_0}$: Velocity of the Sun in the Galactic frame of rest
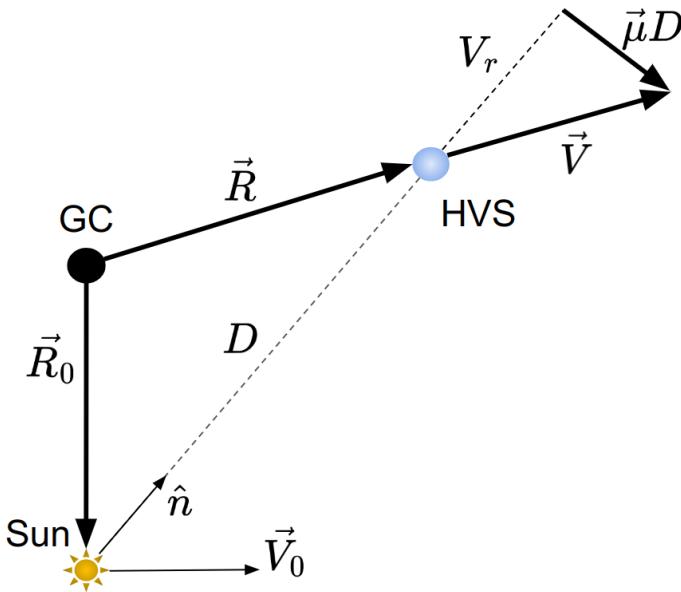
Figure 3.1: A diagram of the position ($\vec{R}$) and velocity ($\vec{V}$) of an HVS relative to the GC. This can be expressed in terms of a combination of the star's observed position ($\hat{n}$) and proper motion ($\vec{\mu}$) and measurements of the Sun's position relative to the GC ($\vec{R_0}$) and velocity ($\vec{V_0}$) along with the distance to the star (D) and its radial velocity (V_r).

- V_r: Radial velocity of the source relative to the Sun

- $\vec{\mu}$: Proper motion of the source in the Heliocentric frame (always $\perp$ to $\hat{n}$)

giving

$$\vec{R} = \vec{R_0} + D\hat{n} \tag{3.1}$$

and

$$\vec{V} = \vec{V_0} + V_r\hat{n} + D\vec{\mu}. \tag{3.2}$$

We provide a sketch of the configuration in Fig. 3.1. We use the measurement of $\vec{R_0}$ from GRAVITY Collaboration et al. (2018), which is 8.122 kpc, and $\vec{V_0}$ from Drimmel & Poggio (2018), which is $[12.9, 245.6, 7.78]$ km s^{-1}. The unit vector $\hat{n}$ is set by the sky coordinates of a star which are provided by *Gaia*, as is the proper motion vector $\vec{\mu}$. For most sources in *Gaia* DR3, approximately 1.5 billion, we have measurements for all but D and V_r. It is worth noting here that *Gaia* does measure parallaxes, but only for nearby

sources do these result in precise distance measurements. The way of circumventing this lack of knowledge of D and V_r is to only focus on objects that have radial trajectories pointing out from the GC. As we will see later, this enables us to calculate the distance and radial velocity from proper motion and sky position alone. For a star on a radial trajectory from the GC, its velocity aligns with the vector pointing from the GC to the star. This means the position $\vec{R}$ and velocity $\vec{V}$ are parallel and thus obey

$$\vec{R} \times \vec{V} = \vec{0}. \tag{3.3}$$

We can expand this using equations 3.1 and 3.2 leaving the terms

$$\vec{R_0} \times \vec{V_0} + V_r(\vec{R_0} \times \hat{n}) + D(\vec{R_0} \times \vec{\mu} + \hat{n} \times \vec{V_0}) + D^2(\hat{n} \times \vec{\mu}) = \vec{0} \tag{3.4}$$

and we can pull out the distance and radial velocity by taking the dot products with $\hat{n}$ and $\vec{\mu}$ and rearranging.

Taking the dot product of equation 3.4 with $\hat{n}$ reveals

$$D = \frac{\hat{n} \cdot \left(\vec{R_0} \times \vec{V_0} \right)}{\vec{R_0} \cdot (\hat{n} \times \vec{\mu})} = \frac{V_0}{\mu} \left(\frac{\hat{n} \cdot \left(\hat{R_0} \times \hat{V_0} \right)}{\hat{R_0} \cdot (\hat{n} \times \hat{\mu})} \right) \tag{3.5}$$

and similarly the dot product of equation 3.4 with $\vec{\mu}$ gives

$$\begin{aligned} V_r &= \frac{D\vec{V_0} \cdot (\hat{n} \times \vec{\mu}) - \vec{\mu} \cdot \left(\vec{R_0} \times \vec{V_0} \right)}{\vec{R_0} \cdot (\hat{n} \times \vec{\mu})} \\ &= V_0 \left(\frac{\frac{D}{R_0} \hat{V_0} \cdot (\hat{n} \times \hat{\mu}) - \hat{\mu} \cdot \left(\hat{R_0} \times \hat{V_0} \right)}{\hat{R_0} \cdot (\hat{n} \times \hat{\mu})} \right). \end{aligned} \tag{3.6}$$

This means that for a star moving on a radial trajectory, the distance and radial velocity are set by the observables available for ~ 1.5B stars in *Gaia* DR3.

For HVSs, the assumption that their trajectory is purely radial can be reasonable, since stars moving much faster than the escape velocity are not deflected significantly by the Galactic potential (Kenyon et al. 2018; Boubert et al. 2020). In this work, we apply the above presented method to determine the distance and radial velocity to all stars in *Gaia* DR3. Since those solutions are only physical for stars on radial trajectories, which most stars in *Gaia* are not, we refer to them as the implied distance, D_I, and implied radial velocity, $V_{r,I}$.

We convert D_I to implied parallax (ϖ_I) to allow for convenient comparison to measurements by taking the inverse of D_I. We calculate the uncertainty on ϖ_I assuming that the proper motion measurement is the

only source of uncertainty. We make the simplifying assumption that the proper motion uncertainties are uncorrelated and since $\varpi_{\mathrm{I}} \propto \mu$, we can approximate the uncertainty on ϖ_{I} as

$$\sigma_{\varpi_{\mathrm{I}}} = \frac{\sqrt{\left[\vec{R}_0 \cdot (\hat{n} \times \hat{\mu}_\delta)\, \sigma_{\mu_\delta}\right]^2 + \left[\vec{R}_0 \cdot (\hat{n} \times \hat{\mu}_{\alpha*})\, \sigma_{\mu_\alpha}\right]^2}}{\hat{n} \cdot \left(\vec{R}_0 \times \hat{V}_0\right)}, \qquad (3.7)$$

with $\hat{\mu}_\delta$ and $\hat{\mu}_\alpha$ the basis vectors for the proper motion in declination (Dec) and right ascension (RA) respectively, and σ_{μ_δ} and σ_{μ_α} the uncertainties on the proper motions in Dec and RA respectively. For a more accurate uncertainty analysis one would sample over the uncertainty on all observables.

3.2.2   HVS candidate selection criteria

The ϖ_{I} and $V_{\mathrm{r,I}}$ we derived in the previous section can be used to identify HVS candidates. Here we give a short conceptual overview of our approach. We first compare the implied parallax to the *Gaia* parallax and reject candidates for which they are inconsistent. In addition, we determine the location in the Hertzprung-Russel (HR) diagram according to the implied distance and remove any candidates for which we consider this solution to be unphysical. Furthermore, we only look at HVS candidates with a high enough 3D implied velocity for our radial trajectory assumption to be reasonable. Lastly, we apply additional cuts that are aimed to reduce the contamination of our sample. We base some of these cuts on catalogues of mock HVSs obtained from simulations described in Section 3.4.2. These simulations model the orbits of ejected HVS populations and use the characteristics of *Gaia* to predict their observational properties.

   Having discussed the general principle of how we select HVS candidates in *Gaia*, we now provide the detailed requirements set for a particular star to be observed in our survey. We categorise these into four groups, each discussed in detail below and summarised in Table 3.1. The goal of the selections presented here is to increase the purity of the sample; i.e. reduce the number of non-HVSs while retaining the real HVSs as much as possible. This is motivated by the necessarily limited number of sources we can provide follow-up observations for.

3.2.2.1   Astrometric and kinematic selections

We start by using the astrometric and kinematic properties of stars to select potential HVSs. HVSs need to be very fast to prevent significant deviation from their otherwise radial trajectories by the torque on the orbit generated by the non-sphericity of the Galactic potential. For this reason, we only consider stars with total implied velocity, V_{I}, in the range $[800, 3500]$

km s^{-1}, independent on Galactocentric distance. The lower limit ensures that (excluding the inner GC) all stars would be unbound. The upper limit includes most of the extremely fast predicted HVSs (Rossi et al. 2014), while limiting contamination by non-HVSs (which can have extremely high implied velocities). Secondly, we compare the implied parallax to the parallax measured by *Gaia*. We only select sources with implied parallax consistent with measured parallax within 2σ, considering both the uncertainty on the measured and the implied parallaxes. We do this comparison to the parallax and not the distance since the parallax has Gaussian uncertainties.

The source of uncertainty on ϖ_I is the measured uncertainty on $\vec{\mu}$ from *Gaia*. Because comparing parallaxes is not constraining when both are highly uncertain, we only consider sources where ϖ_I over its associated uncertainty is larger than five. In addition, most parallaxes in the *Gaia* catalogue have large fractional uncertainties. This also means that requiring the implied parallax and measured parallax to be consistent is often not a stringent requirement. To alleviate this, we additionally compare the implied distances to the photo-geometric ones determined in Bailer-Jones et al. (2021). The photo-geometric distances make use of the colour-magnitude information from stars by incorporating a colour-magnitude dependent prior on the extinction corrected absolute magnitude, which varies as a function of sky position. Similar to the parallax selection, we require that the implied distances are consistent with the photo-geometric ones. As a threshold we use 2σ, where we use the 16th and 84th percentiles from Bailer-Jones et al. (2021) as the negative and positive 1σ estimates for the posterior respectively.

3.2.2.2   Photometric selections

The implied position of many stars in the HR diagram (according to the implied distance) will be unphysical, because the implied distances are only appropriate for stars moving on radial trajectories. To remove unphysical solutions, we only consider sources within part of the HR diagram.

We first correct the G, G_{RP}, and G_{BP} magnitudes for extinction using the 2D extinction map from Schlegel et al. (1998) and the recalibration from Schlafly & Finkbeiner (2011), assuming every source is behind the extinction layer. We use an $R_V = A_V/E(B-V)$ of 3.1 with the Fitzpatrick (1999) extinction law. We use the values provided by *Gaia* to convert the extinction at 550 nm to the relevant bands[1]. Since those corrections are formally only valid when the intrinsic colour is known, we perform 10 iterations over the extinction coefficients, adjusting the colour of the source at each step. We only select sources which are implied to be within $\Delta M_G = 2$ of the main-sequence, which we define to be $M_G = 4.3 \times (G_{BP} - G_{RP}) + $

[1] https://www.cosmos.esa.int/web/gaia/edr3-extinction-law
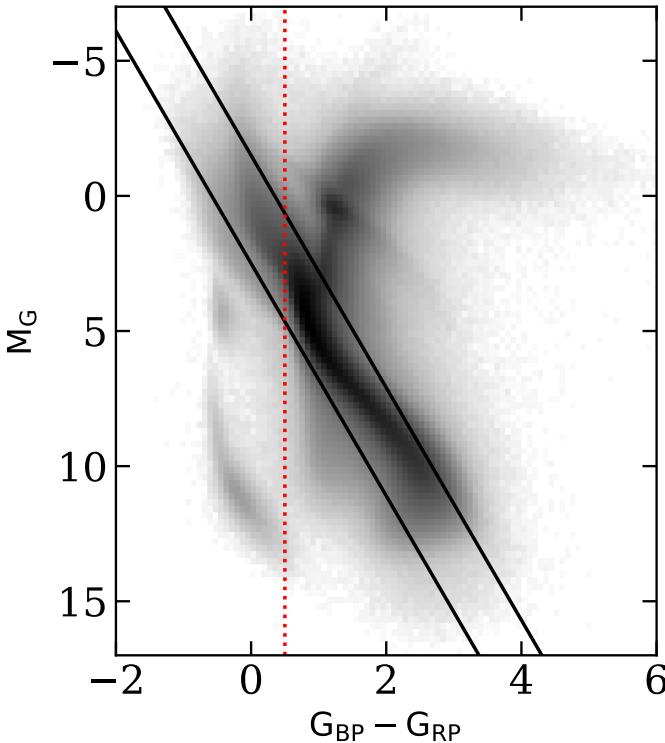
Figure 3.2: Reference HR diagram created from sources in *Gaia*. The dotted line gives our colour limit, only stars to the left of which we consider. The solid lines indicate the main-sequence region within which we consider HVS candidates. The colour scale is logarithmic.

0.5. This is a fairly simplistic selection, which can be improved upon in future surveys by, for instance, taking the relative contamination of our HVS candidate catalogue by field stars into account as a function of position on the HR diagram. To exclude highly extincted regions, we only consider HVS candidates with a G-band extinction coefficient $A_G < 1.5$.

We noticed that the relative contamination of our HVS candidate catalogue by field stars was lowest for blue sources, based on the simulations we discuss in Section 3.4. For this reason we select only sources with an extinction corrected $BP - RP$ colour below 0.5, which corresponds to a mass $\gtrsim 1.3$ $M_\odot$. We show these selections in Fig. 3.2 in combination with a reference HR diagram for sources in *Gaia*. To provide perspective, we also overlay the density of HVS candidates in *Gaia* in Fig. 3.14 in the appendix.

Table 3.1: Overview of the HVS candidate selections used in this study.

Selection	Section for reference
$\varpi_{\mathrm{I}} - \varpi_{\mathrm{Gaia}} < 2\sigma$	3.2.2.1
$800 < V_{\mathrm{I}} < 3500 \ \mathrm{km\,s^{-1}}$	3.2.2.1
$\varpi_{\mathrm{I}}/\sigma_{\varpi_{\mathrm{I}}} > 5$	3.2.2.1
RUWE < 1.4	3.2.2.1
$D_{\mathrm{I}} - D_{\mathrm{BJ}} < 2\sigma$	3.2.2.1
$-1.5 < M_{\mathrm{G}} - 4.3 \times (\mathrm{G_{BP}} - \mathrm{G_{RP}}) < 2.5$	3.2.2.2
$A_{\mathrm{G}} < 1.5$	3.2.2.2
$\mathrm{G_{BP}} - \mathrm{G_{RP}} < 0.5$	3.2.2.2
$N_{\mathrm{HVS,sim}}/N_{\mathrm{HVS,candidates}} > 5$	3.2.2.3
Separation from LMC $> 8\mathrm{deg}$	3.2.2.3
Separation from SMC $> 3\mathrm{deg}$	3.2.2.3

3.2.2.3 Sky coordinate selections

We know that the distribution of HVSs on the sky should be anisotropic, even for an isotropic ejection mechanism in the GC due to on-sky projection. Moreover, we expect that the contamination of our HVS candidate sample is highest near the GC and anti-centre, where our selections are relatively ineffective. In addition, in the Galactic plane we expect the contamination to be higher than at high Galactic latitudes due to the high relative number of field stars to expected HVSs. In order to limit contamination by field stars we define the expected number of HVSs to number of candidates for different directions in the sky. For this purpose we use *HEALPix*[2] (*nside* = 3; Górski et al. 2005; Zonca et al. 2019) to divide the sky into equal area pixels. We determine the expected density of HVSs across the sky using the simulations described in Section 3.4. We only observe sources in *HEALPix* pixels where the ratio of the density of expected HVSs to candidates (normalised by the number of candidates) is larger than five[3].

To prevent contamination from stars in the Large Magellanic Cloud (LMC) or Small Magellanic Cloud (SMC), we additionally require any HVS candidate to be more than 8 and 3 deg from their respective centres[4].

The selection steps up to this point leave us with only 600 HVS candidates, the previously identified S5-HVS1 being one of the top candidates (Koposov et al. 2020). We provide an overview of the selections in Table 3.1. The complete source list can be found online[5]. We provide the first

---

[2]http://healpix.sf.net/

[3]The `HEALPix` pixels we use are 0-19, 22-31, 35-42, 46-55, 58-62, 65, 66, 69-73, 78, 82-85, 93-96, 101, 102, 103 and 107 in the ring scheme based on RA and Dec.

[4]The LMC centre we use has $l = 280.4652$, $b = -32.8884$ deg. The SMC centre we define to be at $l = 302.8084$, $b = -44.3277$ deg.

[5]https://zenodo.org/doi/10.5281/zenodo.12179452

Table 3.2: Here we provide the first five sources, ordered by `source_id`, from our catalogue of 600 HVS candidates with their *Gaia* DR3 `source_id`, implied heliocentric radial velocity, and implied heliocentric distance.

Gaia DR3 `source_id`	$V_{\mathrm{r,I}}$ [km s^{-1}]	$D_{\mathrm{I}}[kpc]$
16647293239608960	805	6.6
57537894455885184	1429	15.2
292010016092347648	829	21.1
307220183209400448	638	12.2
331500110074917120	1092	18.7
...		

five sources in Table 3.2.

3.2.2.4   Instrument-specific selections

We used two observatories for follow-up observations of our candidates; the Isaac Newton Telescope (*INT*) and the New Technology Telescope (*NTT*). The *INT* is a 2.54-meter telescope located on the island of La Palma and the *NTT* is a 3.58-meter telescope sited at La Silla in Chile. We used two observatories to gain both northern and southern hemisphere coverage of the sky. The instrument-specific selection for the *INT* is

- G < 19

- Dec > −20 deg

- RA < 60 deg or RA > 230 deg.

Our instrument specific selection for the *NTT* is

- G < 19.3

- Dec < 0 deg.

We find a total of 284 HVS candidates in *Gaia* that match all of the listed selections. We performed spectroscopic follow-up observations for these sources, which are described in Section 3.3.

3.3   Observations

In the previous section, we discussed how we select our HVS candidates. In this section we describe the follow-up observations we performed for those candidates. We describe the instrument set up in Section 3.3.1, the data reduction in Section 3.3.2, the spectral analysis in Section 3.3.3, and lastly the observed stars in Section 3.3.4.

An overview of our observations is given in Table 3.3.

Table 3.3: Overview of the observing dates for our survey.

Instrument	Observing dates	Run number
INT	2022 Aug 20-21	ING.NL.22B.002
INT	2022 Sep 19-30	ING.NL.22B.002
NTT	2022 Nov 25-28	110.23SU.001 & 110.23SU.002
NTT	2023 June 3	111.24MP.001
NTT	2023 July 14-15	111.24MP.002

3.3.1   Instrument set up

To confirm or reject HVS candidates, we do not require high radial velocity precision. We therefore set up the instruments to allow efficient radial velocity measurements with uncertainties of a few tens of $km\,s^{-1}$.

At the *INT*, we use the Intermediate Dispersion Spectrograph (IDS) with the EEV10 detector and the R400V grating. This gives us an approximate resolution of R $= \frac{\lambda}{\Delta\lambda} = 1600$ at 4500 Å[6]. We were allocated a total of twelve nights of observing at the *INT*.

At the *NTT*, we use the ESO Faint Object Spectrograph and Camera (v.2; EFOSC2) with grism Gr#07. This setup gives us a resolution of about 500 at 3800 Å[7]. We were allocated seven nights of observing at the *NTT*.

3.3.2   Data reduction

In the following, we describe the data reduction steps applied to the raw data from the *INT* and *NTT*. We use the same pipeline for the two telescopes, with minor adjustments where needed to account for the different instruments.

We perform standard bias subtraction and flat-field correction to all science spectra. We then mask out any cosmic rays using `ccdproc` (Craig et al. 2022). We trace and subsequently extract the spectra using `PyRAF`, which is the Python implementation of `IRAF` (Tody 1986, 1993). Background subtraction is performed by selecting a region around the science spectrum. Through visual inspection, we make sure there are no sources in the background region and make adjustments where needed. In the case of poor seeing during the observation, for instance, the background region might need to be further separated from the centre of the science spectrum. In the case of the *INT*, wavelength calibration is performed using the CuAr+CuNe lamp. For the *NTT* we make use of the He and Ar lamps. An arc spectrum is taken after every science exposure to account for instrument

---

[6]https://www.ing.iac.es/astronomy/instruments/ids/idsgrat_tables.html
[7]https://www.eso.org/sci/facilities/lasilla/instruments/efosc/inst/Efosc2Grisms.html

flexures. No flux calibration is applied, since it is not required for radial velocity measurements.

We additionally calculate the signal-to-noise (S/N) as a function of wavelength. We start by taking the square root of the sum of squares of the photon noise and the readout noise. We remark that this is an approximation, because in the regime of low photon counts, the constraint on the flux is not Gaussian (see e.g. Guy et al. 2023, section 4.2.9). We use the same trace used in extracting the science spectrum to extract the noise spectrum and perform error propagation in subtracting the background.

3.3.3 Spectral analysis

Having obtained wavelength calibrated spectra, we now describe our analysis. We determine the radial velocities from our reduced data using `RVSpecFit` (Koposov 2019b). This software relies on direct pixel fitting using interpolated stellar templates from the PHOENIX spectral library (Husser et al. 2013). For the spectra from the *NTT* we use the wavelength interval $3500 < \lambda < 5240$ Å, while for the *INT* we use the wavelength interval $3600 < \lambda < 6800$ Å. We perform barycentric correction on the resulting radial velocities using `Astropy`.

3.3.4 Observed HVS candidates

We obtained radial velocity measurements for 196 out of our 284 HVS candidates. This translates to a completeness factor of about 69%. We only consider spectra with a S/N larger than two and three for the *INT* and *NTT* respectively. For lower S/N spectra we were not able to recover reliable radial velocity measurements. For another five stars, we found previous radial velocity measurements in the *SDSS* DR14 low-resolution stellar catalogue (Abolfathi et al. 2018) or in the *LAMOST* DR8 low-resolution catalogue (Zhao et al. 2012). We give the data on these stars in the appendix Table 3.5. We will also consider these five sources in our observational sample, bringing our total completeness to about 71%. In Fig. 3.3 we plot the sky distribution of our HVS candidate catalogue, indicating which ones have been observed successfully. We provide an observational log of our observations in an online table[8], the first five rows of which are shown in Table 3.4.

In Fig. 3.4 we plot the implied radial velocity against the measured radial velocity. We can see that only the previously identified S5-HVS1 falls on the bisection, clearly identifying it as an HVS. However, not all HVSs will be on the bisection, because either their trajectories are not purely radial or due to observational uncertainty. In practice, this means we can not state with

---

[8]https://zenodo.org/doi/10.5281/zenodo.12179452
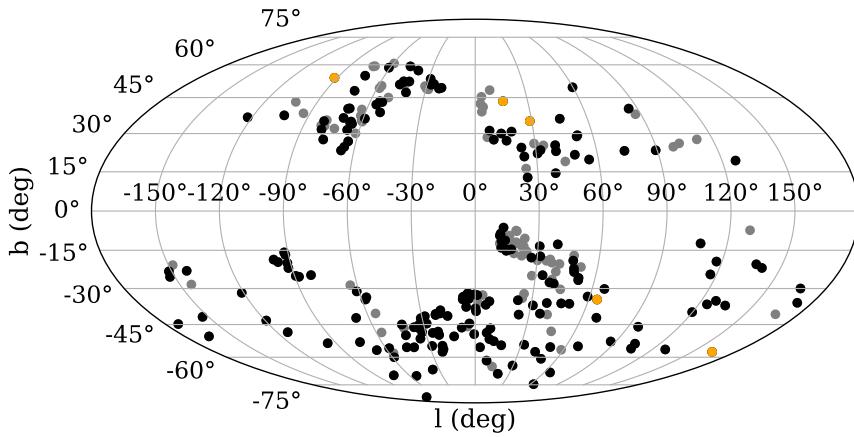
Figure 3.3: Sky distribution of our HVS candidate sample in Galactic coordinates. Black points indicate the stars with follow-up observations and the orange points the ones with either *SDSS* or *LAMOST* radial velocity measurements. For the grey stars we do not have radial velocity measurements.
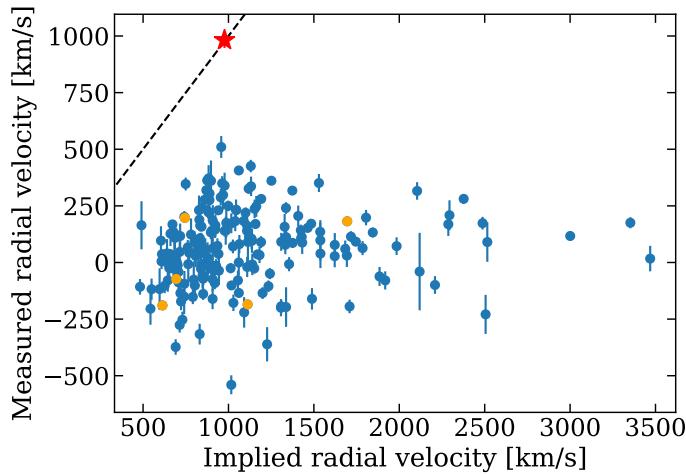

Figure 3.4: Implied radial velocity against the measured radial velocity for the HVS candidates we observed. The dashed line gives the bisection. The blue points indicate our measurements, the red star highlights our measurement of S5-HVS1, and the orange points show the stars with radial velocity measurements from *SDSS* and *LAMOST*.

Table 3.4: Five sources from our observed HVS candidates in no particular order. The full table can be accessed online (see main text). The radial velocities are in the barycentric frame.

Gaia DR3 source_id	radial velocity [km s^{-1}]	radial velocity uncertainty [km s^{-1}]	date of observation	instrument
6408940170344856320	220.6	38.0	2022-11-28	NTT
1664729323960896160	42.4	40.6	2022-09-23	INT
4080886923180284928	32.8	35.4	2023-06-04	NTT
3755045965083116160	281.0	21.2	2022-11-26	NTT
4195180362312873984	-75.7	80.2	2023-07-16	NTT
...				

absolute certainty that the other observed sources are not HVSs, because of observational uncertainty and orbits which are not purely radial. We will return to this point in Section 3.4.3.

3.4 HVS mock catalogues

In this work we use a comparison between observations and simulations to infer properties of HVSs. Section 3.4.1 describes the HVS ejection mechanism and Section 3.4.2 discusses the simulation suite in which it is applied. We use these simulations to make informed decisions for our HVS candidate selection, mentioned in Section 3.2.2. Most importantly, the simulations allow us to provide constraints on the ejection of HVSs and the GC environment as presented in Section 3.5.

3.4.1 Hills mechanism

We assume HVSs are ejected through the Hills mechanism proposed in Hills (1988) (for a review see Brown 2015). In this mechanism, a binary star system approaches a massive black hole (MBH) within the tidal radius, where the tidal force from the MBH exceeds the binary self gravity. The result is that the binary is separated; one star is captured by the MBH in a tight, elliptical orbit and the other star is ejected as an HVS. If the progenitor binary is on a parabolic orbit, which is an excellent approximation in the scenario assumed here (Kobayashi et al. 2012), both stars in the binary have an equal likelihood of being ejected, independent of mass ratio (Sari et al. 2009a). The velocity at which the HVS is ejected from the sphere of influence of the MBH is given by

$$V_{\mathrm{ej}} = \sqrt{\frac{2Gm_{\mathrm{c}}}{a}} \left(\frac{M_{\mathrm{MBH}}}{M}\right)^{1/6}, \qquad (3.8)$$

where we neglected a multiplicative constant factor of order unity which depends on the geometry of the encounter (see, e.g., Fig.10 in Sari et al. 2009a). In equation 3.8, m_{c} is the mass of the captured companion, M_{MBH} in the case of the Galaxy is the mass of Sgr A* (4×10^6 M$_\odot$, Eisenhauer et al. 2005; Ghez et al. 2008; Schödel et al. 2009; Gillessen et al. 2009; Boehle et al. 2016; Gillessen et al. 2017; Do et al. 2019; GRAVITY Collaboration et al. 2019), and M the total mass of the progenitor binary (Sari et al. 2009a; Kobayashi et al. 2012; Rossi et al. 2014). Most of the stars ejected through this mechanism are still bound to the Galaxy, but the ejections can occur at velocities in excess of 1000 km s^{-1}, putting stars on orbits unbound to the Galaxy.

3.4.2   Simulations

The simulation framework we use is `Speedystar`[9] (Contigiani et al. 2019; Evans et al. 2022b). This code is able to create mock observational catalogues of Hills mechanism HVSs, and of stars ejected by a binary massive black hole (Evans et al. 2023). Here we focus exclusively on the Hills mechanism. Below we describe the most important steps to understand the model, for details we refer to Evans et al. (2022b).

We start by creating an HVS progenitor population of binaries. These binaries are characterised by the primary zero-age main sequence mass ($m_{\rm p}$), mass ratio (q), and the orbital semi-major axis (a) at the moment of disruption. The mass of the primary is drawn from a mass function (MF) which follows a power law distribution of the form $f(m_{\rm p}) \propto m_{\rm p}^{-\kappa}$ in the range $[0.1, 100]$ M$_\odot$. The secondary is then assigned a mass according to a mass ratio distribution of the form $f(q) \propto q^\gamma$, with $0.1 \leq q \leq 1$. Lastly, the orbital semi-major axis are assigned assuming a binary orbital period distribution of the form $f(\log P) \propto (\log P)^\pi$. Different values for the three parameters κ, γ, and π are explored with uniform priors to account for uncertainty in the HVS progenitor population. For κ we use the interval $[0.3, 3.3]$ to allow for both very top-heavy and bottom-heavy MFs. For γ we use the interval $[-2, 2]$ and for π the interval $[0, 2]$. We allow for the wide range in MF index because of the uncertainty on the IMF near the GC, alluded to in the introduction. IMF indexes of 0.45, 1.7, and a canonical 2.3 have all been claimed near Sgr A* (Bartko et al. 2010; Lu et al. 2013; Löckmann et al. 2010, respectively). The parameter range on π allows for a log-uniform period distribution (also known as Öpik's law), or distributions with increasing fractions of wide binaries. The parameter range on q encapsulates the results from different studies into massive binaries in star forming regions (Sana et al. 2012, 2013; Moe & Stefano 2013, 2015; Dunstall et al. 2015).

For each star we assume solar metallicity, with $Z_\odot = 0.0142$ (Asplund et al. 2009). Additionally, we performed tests with super-solar metallicity values of [Fe/H] = 0.3, since stars near the GC are mostly metal rich (Do et al. 2015; Feldmeier-Krause et al. 2017a; Nandakumar et al. 2018; Schultheis et al. 2019).

From this progenitor population, binaries are drawn at a constant rate. One of the binary components is selected at random and assigned an ejection speed according to equation 3.8. We assume the ejection is isotropic and assign the star a random age within the total lifetime of the shorter-lived star of the progenitor binary, where we implicitly assume a constant star formation rate. The ejected stars are subsequently integrated in the potential of the Galaxy (see Marchetti et al. 2018, table 1). The result

---

[9]`https://github.com/fraserevans/speedystar`

is a simulated current population of HVSs throughout the Galaxy. These HVSs are evolved using single stellar evolution models from Hurley et al. (2000) in AMUSE (Portegies Zwart et al. 2009, 2013; Pelupessy et al. 2013; Portegies Zwart & McMillan 2018).

Lastly, mock observations are performed incorporating the visual extinction along the line of sight, effective temperature, surface gravity, and metallicity of the simulated HVSs. This is done using the MESA Isochrone and Stellar Tracks model grid (Dotter 2016a; Choi et al. 2016a). This allows us to determine the magnitude and colour of these stars as they would be observed by various observatories. Astrometric uncertainties in parallax and proper motion, as observed by *Gaia*, are determined using the DR3 astrometric spread function of Everall et al. (2021).

3.4.3   HVS candidate selection applied to simulations

Now that we have described both our simulations and observations, we will describe how we apply the selections discussed in Section 3.2 to our simulated HVSs.

Firstly we have to consider which stars are observed by *Gaia* in the first place. Although the magnitude limit of *Gaia* is G = 20.7 (Gaia Collaboration et al. 2016), not all sources down to that depth appear in the published *Gaia* catalogue for various reasons. A description of which stars do and do not appear in the published *Gaia* catalogue is referred to as the *Gaia* selection function, which has been empirically determined in Cantat-Gaudin et al. (2023)[10]. Although *Gaia* is virtually complete down to our magnitude limit of G = 19.3 except in crowded regions, we nonetheless incorporate the empirical *Gaia* selection function for completeness. We also convolve the 'true' parallaxes and proper motions with their uncertainties. No uncertainties are considered on the photometric measurements.

Most of the selections described in Section 3.2.2 can be directly applied to the resulting mock HVS catalogue. We can calculate the implied distance and radial velocity for the mock HVSs in the same way we did for the *Gaia* catalogue and apply our selections. The same can be done for the photometric, sky coordinate, and instrument-specific selections. An exception is the photo-geometric distance from Bailer-Jones et al. (2021). To apply this selection to the mock HVS catalogue, we resort to the Bailer-Jones (2015) source code (C.A.L. Bailer-Jones, private communication) and calculate the photo-geometric distances for the simulated HVSs. After obtaining the photo-geometric distances for all our mock HVSs in this way, we again apply the same selections to our mock catalogue as to the *Gaia* catalogue.

---

[10]https://github.com/gaia-unlimited/gaiaunlimited?tab=readme-ov-file
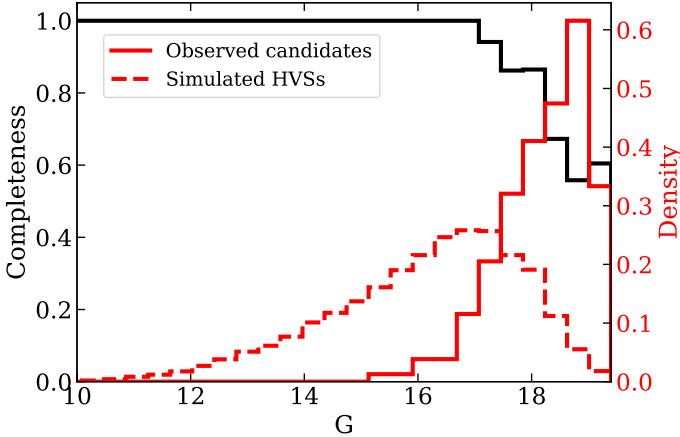
Figure 3.5: Observational completeness of our sample of HVS candidates as a function of apparent G magnitude. We also show the distribution of simulated HVSs and our observed candidates in G magnitude in red on the right axis.

The only exception is the `RUWE`, which we do not model for our simulated HVSs. We expect HVSs to be single stars and for single stars fewer than one in a million have `RUWE` > 1.4 (Penoyre et al. 2022).

Since our observational catalogue does not cover every candidate from the selections introduced in Section 3.2, we need a prescription to quantify our observational completeness. In Fig. 3.5 we show our follow-up observational completeness as a function of the apparent magnitude of the HVS candidates. For each bin in G magnitude, we calculate the fraction of sources we observed over the number of sources in our catalogue (see Section 3.2.2.4). As can be seen, the observational completeness decreases for fainter sources, since we prioritised bright sources. In addition, we have lower completeness for regions with many HVS candidates that were not accessible for a significant amount of time during the observational runs. This expresses itself mainly in a lower completeness near the GC. In Fig. 3.6 we show our follow-up observational completeness as a function of the angle to the GC. Since the observational completeness shows a stronger bias in G-magnitude than in the sky position, we decide to model the observational completeness only on G-magnitude. We bin the magnitudes of our candidates using 50 bins between $G = 0$ and $G = 19.4$ and calculate the fraction of observed stars within each bin. If there are no candidates in a bin, we consider the observations to be complete. The model of the selection function of our observations is then the completeness fraction per bin in G-magnitude, which we can apply to our simulations.

Lastly, we noted in Section 3.3.4 that we cannot reject with absolute
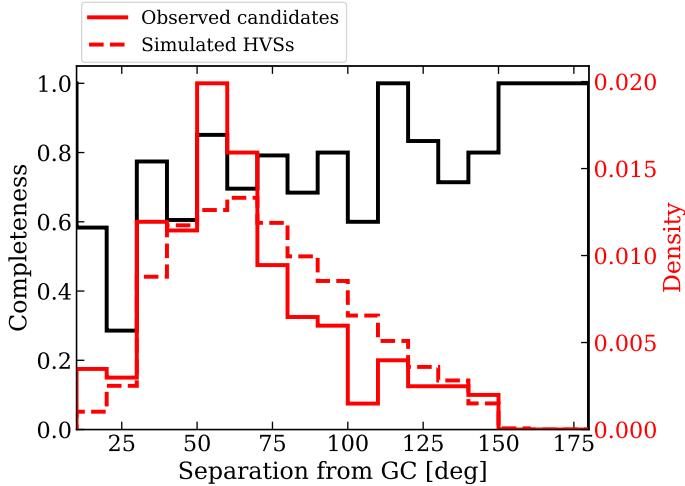
Figure 3.6: Observational completeness of our sample of HVS candidates as a function of the angle to the GC. In addition, we show the distribution of simulated HVSs and our observed candidates as a function of their angular separation from the GC in red on the right axis.

certainty the HVS candidates that have significantly different observed and implied radial velocities. To account for this, we only consider simulated HVSs for which $V_{r,I} - V_r < 300 \text{ km s}^{-1}$, which holds for about 86% of HVSs in our simulations. None of the sources we observed, except S5-HVS1, meet this criterion.

By combining all observational selections and applying them to the simulations, we can predict the number of HVSs that should be in our observational sample, which we use to provide constraints on the GC in Section 3.5.2.

3.5 Constraints on the Galactic Centre

Having described the observations, we can now provide physical constraints on the GC environment and HVS ejection process. Before we do so in Section 3.5.2, we will first discuss the properties of HVSs that can be observed with *Gaia* in Section 3.5.1.

3.5.1 HVSs observable by *Gaia*

Not all HVSs can be observed by *Gaia*. Low-mass stars, for instance, can only be observed out to short heliocentric distances, due to their intrinsic
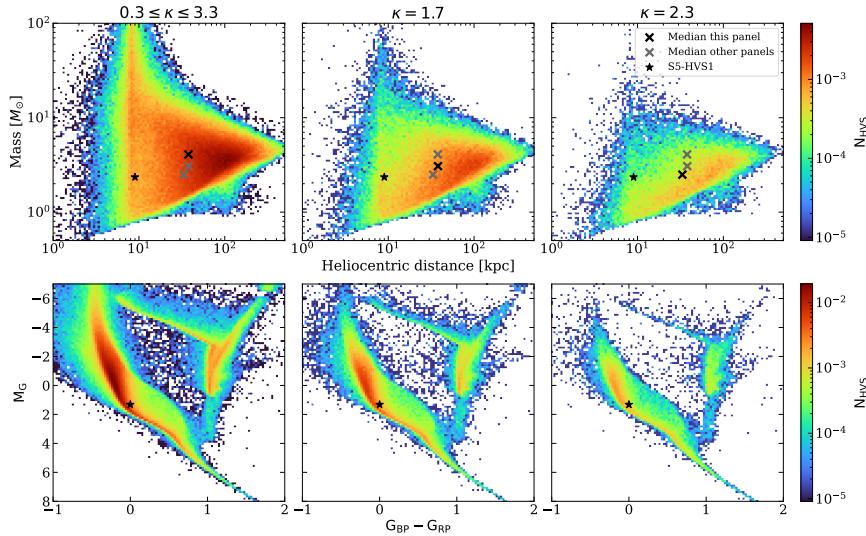
Figure 3.7: Predicted population of unbound HVSs in the *Gaia* catalogue. The three columns display different assumptions on the MF of the primary of the HVS progenitor binary population, as indicated above the respective columns. The numbers per bin are calculated for a fiducial ejection rate of 10^{-6} yr^{-1}. *Top row*: the distance-mass distribution for predicted HVSs in *Gaia*. The black crosses indicate the median values of the respective panels, while the grey crosses indicate those for the other panels for easy comparison. The black star indicates the position of S5-HVS1. *Bottom row*: the extinction corrected colour-magnitude distribution for predicted HVSs in *Gaia*.

low brightness. To establish the properties of HVSs that can be observed by *Gaia*, we make use of the simulations described in Section 3.4.2. We require that to be observed by *Gaia* and be considered an HVS, the star must have an apparent G magnitude below 20.7, in addition to travelling on an unbound orbit. We additionally consider the *Gaia* selection function mentioned in Section 3.4.3.

For every set of model parameters we can now find the stars that would appear in the *Gaia* catalogue. We caution that due to observational uncertainties, not all of these real HVSs could be trivially identified as such; in practice, it can be challenging to determine if a star is in fact unbound or not. In Fig. 3.7 we plot the distance against the mass for HVSs that can be observed by *Gaia*, along with the colour-magnitude diagram. In the left column, we add the contributions from all MFs within our prior range (see Section 3.4.2). In the middle and right columns we show the distribution for two specific MFs (a top-heavy one, $\kappa = 1.7$, and one close to the canonical Salpeter MF, $\kappa = 2.3$). The resulting distribution should be seen as a pre-

diction for the overall population distribution of HVSs present in the *Gaia* catalogue, discovered or undiscovered. We can see that most of the HVSs in the *Gaia* catalogue will be at heliocentric distances between about 10 and 200 kpc, be on the order of a few solar masses, and be main-sequence stars. Below heliocentric distances of ~ 5 kpc, the volume covered by *Gaia* is too small which makes discovering any HVSs within this region highly unlikely. The vertical feature in the top row at ~ 8 kpc is the GC, where, due to projection effects, an over density of HVSs on the sky is present. The cutoff feature towards low-mass/distant stars is caused by the detection limit of *Gaia* in apparent magnitude. The stars below this boundary are giants, which are visible out to further distances than main-sequence stars. The upper right of the figure is not populated, because extremely massive stars do not live long enough to travel far from the GC. In Appendix Fig. 3.15, we additionally show the apparent magnitude distribution for am MF index $\kappa = 2.3$.

Interestingly, the MF has little effect on the ratio of evolved to main-sequence stars we expect to be within the *Gaia* catalogue. The fraction of evolved HVSs is expected to be about 0.17, changing by less than 0.03 between different MFs.

Using these results, we can also determine the completeness of *Gaia* for HVSs as a function of mass. In other words, what fraction of HVSs is expected to be in the *Gaia* catalogue as a function of mass. For this, we only consider HVSs that have not evolved into remnants. We show the results in Fig. 3.8 for varying distance limits. We can see that the completeness for HVSs in *Gaia* approaches zero for masses below ~ 1 $M_\odot$ even when only considering HVSs within 50 kpc from the Sun, which are the ones our observations are mainly sensitive to. For high mass HVSs ($M \gtrsim 5$ $M_\odot$) the completeness converges to nearly one, because those stars do not live long enough to travel far from the GC.

3.5.2   HVS parameter constraints

Having established the types of HVSs that *Gaia* is sensitive to, we will now describe our constraints on the ejection of HVSs.

By applying the same selections to the simulations described in Section 3.4 as the observations, we can predict the number of HVSs that should be observed in our catalogue given a set of model parameters. By rejecting models that are incompatible with the observed number of HVSs, we provide constraints on the model parameters, and as such, the physical environment of the GC and the process of HVS ejection.

We use uniform priors within the parameter ranges given in Section 3.4.2. The posterior probability of the model with parameters M is therefore given
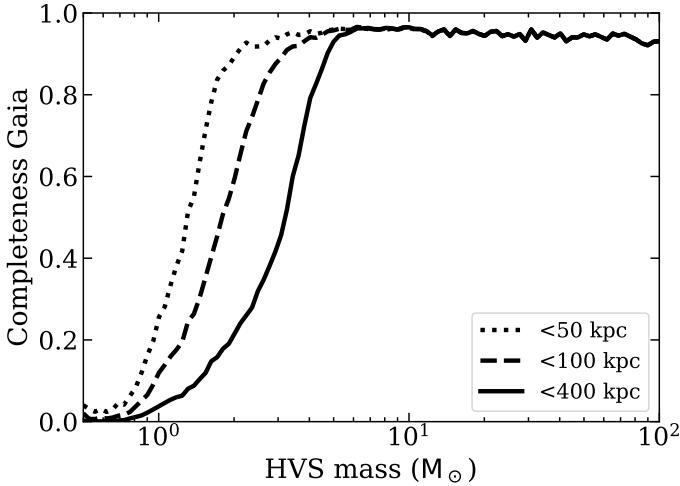
Figure 3.8: Completeness of unbound HVSs in the main *Gaia* catalogue as a function of mass. The different lines correspond to different limits on the heliocentric distance as indicated in the legend. Our observations are mostly sensitive to HVSs within 50 kpc.

by

$$P(M|N_{\mathrm{HVS}}) \propto \frac{\lambda^{N_{\mathrm{HVS}}} \exp{-\lambda}}{N_{\mathrm{HVS}}!}, \tag{3.9}$$

with N_{HVS} the number of observed HVSs and λ the average number of expected HVSs according to the simulations for a given set of parameters M.

We find that the main parameters we can constrain based on the number of observed HVSs are the ejection rate and MF (as is also the case in Evans et al. 2022a,b). Both the log-period distribution and mass-ratio distributions are significant in their effect on the observed number of HVSs, but affect the results much less than the MF and ejection rate.

In Fig. 3.9 we provide the posterior on the MF and ejection rate, marginalised over the binary period and mass ratio distributions. Compared to previous studies, we see a remarkable improvement in the upper limit on the combined ejection rate and MF. Even though we only obtained 196 radial velocity measurements, our results are about one dex more constraining than the $\sim 34M$ radial velocity measurements in *Gaia* DR3 (also considering S5-HVS1 was discovered). This demonstrates the effectiveness of our targeted survey versus blind surveys when identifying HVSs. It is also worth noting that we use a more conservative prior on the log-period distribution compared to earlier work, since the prior used in Evans et al. (2022a,b)
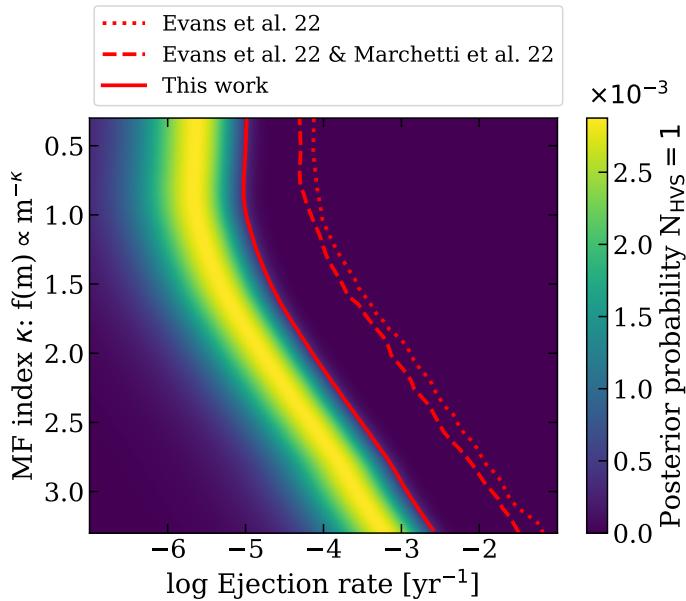
Figure 3.9: Posterior probability on the log of the HVS ejection rate and MF of the primary in the progenitor binary. The solid line gives the 2σ upper limit of the combined ejection rate and MF index. For comparison, the dotted and dashed lines give the 2σ upper limit on the same parameters from blind searches in the 6D phase-space *Gaia* eDR3 and DR3 samples respectively. The eDR3 constraint was calculated in Evans et al. (2022b) and the DR3 constraint is calculated by multiplying the posteriors in Fig. 3 of Marchetti et al. (2022) with the top-right panel of Fig. 5 in Evans et al. (2022b).

results in higher ejection velocities to which our observations are most sensitive. We only highlight the upper limit and not the lower limit, because the upper limit is better constrained. The lower limit is highly sensitive to the prior used (e.g. uniform or Gamma distribution). In Fig. 3.16 we show the posterior if we use a log-uniform prior of $1/\lambda$. The posterior with the log-uniform prior shows that the upper limit remains, while allowing lower ejection rates with equal probability.

For $\kappa \gtrsim 1.5$ we can see in Fig. 3.9 that the ejection rate is a strong function of the MF index. This is caused by our limited sensitivity to low-mass stars, as demonstrated in Fig. 3.7. A more bottom-heavy MF would cause a lower fraction of HVSs to be observable by *Gaia*, thus necessitating a relatively higher ejection rate. For extremely top-heavy MFs, the ejection rate and MF index become relatively independent due to two competing effects:

- a larger fraction of HVSs is sufficiently massive to be observed with increasingly top-heavy MFs, and

- the finite lifetime of these increasingly massive stars causes a significant fraction of HVSs to evolve into stellar remnants before exiting the sphere within which they would be visible were they main sequence stars.

Overall, our upper limit ejection rate constraints are consistent with the process where stellar binaries' centre of mass trajectories have undergone gravitational scatterings off other (single) stars or giant molecular clouds: once every $10^4 - 10^6$ yr a binary star is sent on a highly eccentric orbit such that its pericenter becomes equal or smaller than the tidal separation radius (a.k.a. "loss cone orbit") and an HVS is ejected (Yu & Tremaine 2003; Perets et al. 2007). For a recent review on rates see Stone et al. (2020, especially section 3.6.3). Robust rate predictions to compare with our observational constraints should include realistic descriptions of the phase space and properties of binary stars, as well as those of single stars and giant molecular clouds out to ~ 100 pc from SgrA* in our GC. Such calculations and comparisons are beyond the scope of this paper and will be presented in a follow up investigation.

We can re-parameterise the ejection rate, such that it is no longer a function of the MF index by only considering HVSs massive enough to be observable by *Gaia*. The total ejection rate defines the ejection rate of HVSs in the mass range $[0.1, 100]$ M$_\odot$, as explained in Section 3.4.2. We can transform this to the ejection rate in a different mass range by integrating the MF, which allows us to determine the number of stars in a mass interval $[M_1, M_2]$

$$N(M_1 < M < M_2) \propto \frac{1}{\kappa + 1} \left(M_2^{\kappa+1} - M_1^{\kappa+1} \right). \qquad (3.10)$$
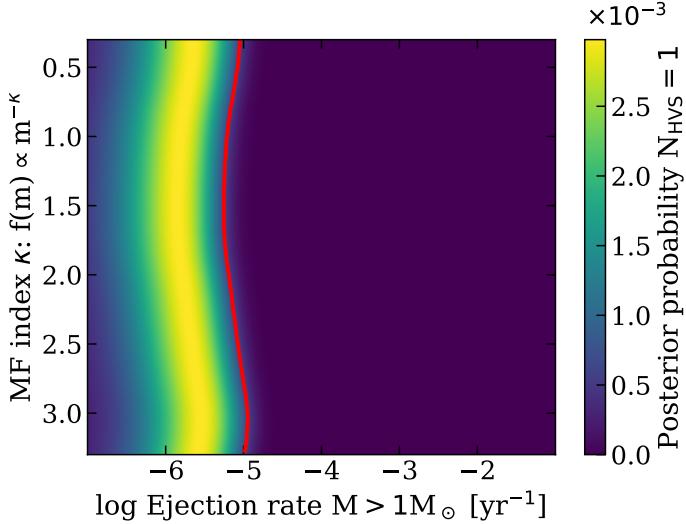
Figure 3.10: Same as Fig. 3.9, but now only for HVSs with $M > 1M_\odot$.

The ejection rate for stars in the mass range $[M_1, M_2]$ is then the total ejection rate times the fraction of stars in the mass interval $[M_1, M_2]$. In Fig. 3.10 we give the resulting posterior for HVSs more massive than $1M_\odot$, making them potentially accessible to *Gaia*. The figure demonstrates that irrespective of the MF, the ejection rate of HVSs more massive than $1M_\odot$ cannot be higher than about 10^{-5} yr^{-1} at 2σ confidence.

In Fig. 3.11 we show the flight time (i.e. time since disruption) distribution of HVSs that our observations are sensitive to. We can see that the observations presented in this work are sensitive to ejections over the past $\sim 50 - 100$ Myr. Our ejection rate constraint therefore applies to ejections within the last $\sim 50 - 100$ Myr. HVS ejections that occurred more than 100 Myr ago would not produce any observable HVSs in our sample and can therefore not be constrained with our observations.

In the hypothetical case we observed two HVSs instead of one, the effect on our constraints is mainly that the posterior (Figs. 3.9 and 3.10) becomes more narrow (the 2σ confidence interval by ~ 0.5 dex). In addition, the 2σ upper limit constraint would increase by about ~ 0.1 dex.

3.5.3   How many HVSs are in *Gaia*?

Using our constraints from Section 3.5.2, we can predict how many HVSs are in the complete *Gaia* DR3 catalogue. Along with the predicted mass-distance relation, shown in Fig. 3.7, this gives us a perspective on the dis-
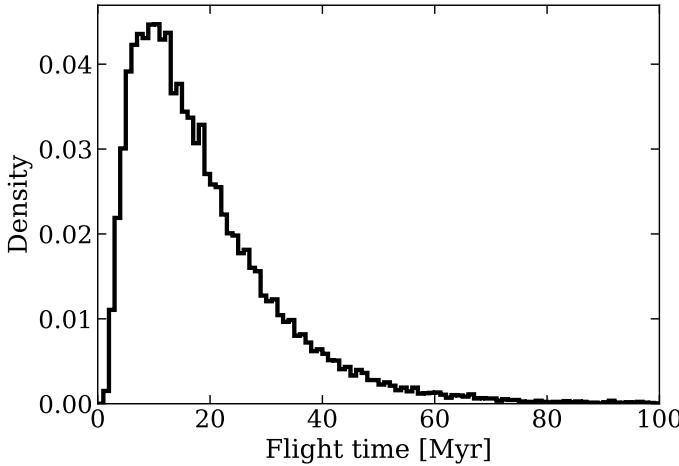
Figure 3.11: Predicted distribution of flight times for HVSs to which our observations are sensitive.

covery space of HVSs in *Gaia*. In Fig. 3.12 we show the expected number of HVSs in *Gaia* as a function of the MF. Within 1σ, we expect there to be between about 5 and 45 HVSs ($1.7 < \kappa < 2.3$), with the mode of the distribution around 18. The significant uncertainty in this number is caused by the single confirmed HVS within our observational sample. Importantly, the number of predicted HVSs is about one to two orders of magnitude less than predicted in earlier work (Marchetti et al. 2018). This difference is mostly caused by a much lower ejection rate than previously expected. Theoretical prospects relying on *Gaia* discovered HVSs will have to take take this into account.

3.6 Discussion

Having presented our main results, we will now provide some discussion. We start by reviewing our assumptions in Section 3.6.1, secondly we discuss the influence of the LMC on our results in Section 3.6.2, then we discuss the previously discovered HVS S5-HVS1 in light of our results in Section 3.6.3, we discuss some previously identified HVS candidates in Section 3.6.4, and lastly provide prospects for the discovery of additional HVSs in Section 3.6.5.
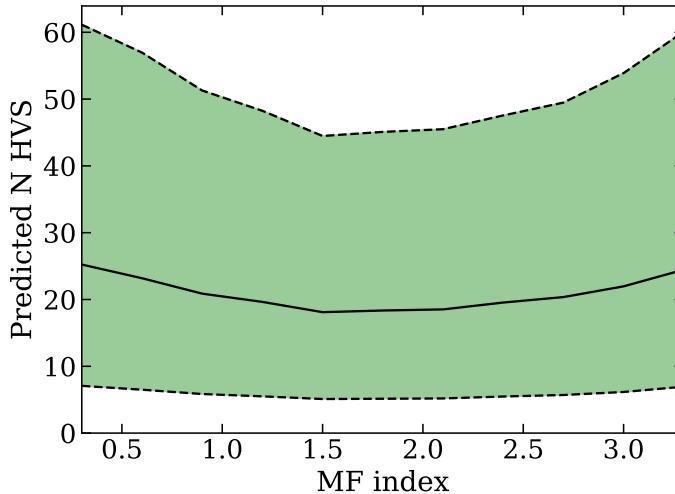
Figure 3.12: Number of unbound HVSs in the complete *Gaia* catalogue as a function of the MF of the progenitor binary population.

3.6.1   Assumptions

In this work we rely on simulations to interpret our observational results. Evaluating the assumptions the simulations rely on is therefore important. Most of these have been discussed in Evans et al. (2022b), including alternative MF and mass ratio parameterisations. We focus here on differences.

The MF we use in the simulations is different from the IMF for HVS progenitor binaries. The simulation assigns a random ejection time within the lifetime of a star, which makes the MF we constrain in this work more analogous to the present day mass function. In addition, the use of a single, time-independent MF in the simulations implicitly assumes a constant star formation rate. This is a simplifying assumption, which we know does not hold in detail (e.g. Nogueras-Lara et al. 2020).

Regarding the log-period distribution, we use a more stringent prior compared to Evans et al. (2022b). Power-law indexes in α (see Section 3.4.2) below zero would result in higher ejection velocities because of the lower binary separation (see equation 3.8). Since our observations are most sensitive to the fastest HVSs, our prior is more conservative.

We assumed a solar metallicity for the progenitor binaries throughout this work. We evaluate the effect of metallicity on our results by running identical simulations with [Fe/H] = 0.3. The difference on the posterior presented in Fig. 3.9 is about 7% of the statistical uncertainty. We therefore conclude that metallicity is currently not an important consideration for our

work.

We also tested the influence of the adopted Galactic potential on our results. To compare, we ran the simulations for the potential determined in McMillan (2017). This potential is a more massive commonly used potential compared to our default one. The difference on our posterior from Fig. 3.9 changes by $0.01 - 0.1$ dex, which is well below the statistical uncertainty.

3.6.2   Influence of the Large Magellanic Cloud

Our search for HVS candidates is based on pure radial trajectories. In reality, the non-spherical potential of the Galaxy deviates HVSs from exclusively radial trajectories. Because we compare our observations with simulations that are performed with a realistic, non-spherical potential and an identical selection procedure, these deviations from radial trajectories are not a concern for our results. However, the LMC is not incorporated in the potential model of the Galaxy in which the simulated HVSs are propagated. This might be a concern if the LMC significantly deviates HVSs from their trajectories (see Kenyon et al. 2018; Boubert et al. 2020). In that case, simulations over-predict the number of HVSs we would be able to find and therefore bias the ejection rate constraint towards values which are too small.

We can evaluate the influence of the LMC by a comparison between integrating orbits of simulated HVSs with and without the LMC. We reran the simulations for $\kappa = 2.3$ including the LMC with a mass of 1.5×10^{11} $M_\odot$, which is near the total mass most studies find and close to half the mass of the Galaxy within 50 kpc (Erkal et al. 2019; Vasiliev et al. 2020; Shipp et al. 2021; Koposov et al. 2023). In the run with the LMC, we found an increase of about 2.8% in the number of recovered HVSs in the simulation, which is about half the Poisson noise for those simulations. Our simulation with the LMC did not include the reflex motion of the Galaxy, but since this effect is of the same order as the deflection (Boubert et al. 2020), we conclude that the LMC has no discernible influence on our results.

3.6.3   S5-HVS1

The serendipitous discovery of S5-HVS1 in the S^5 survey (Li et al. 2019a) appears curious. It is the only HVS that can be unambiguously traced back to the GC and is, by far, the fastest out of the stars that are suggested to originate from the Hills mechanism, in addition to being the closest (Koposov et al. 2020; Brown et al. 2014, 2018). An interesting question to pose is then: what is the chance that the S^5 survey included an HVS? To understand this likelihood, we again make use of the simulations described in Section 3.4.2.

We consider the footprint of the S^5 survey to be any stars with *Gaia* parallax $\varpi < 3\sigma_\varpi + 0.2$, mock DECam (Dark Energy Survey Collaboration et al. 2016) photometry $15 < g < 19.5$, $-0.4 < (g - r) < 0.1$, and falling within 2 deg from an S^5 pointing (see Li et al. 2019b, table 2) following Evans et al. (2022b). Only HVSs with heliocentric radial velocities of $V_r > 800$ km s^{-1} are considered, since these were inspected in Koposov et al. (2020). We find that within the 1σ posterior probability on the MF and ejection rate from Fig. 3.9, the number of expected HVSs in S^5 is about $0.14^{+0.36}_{-0.11}$. Although unlikely, the discovery of S5-HVS1 in S^5 is consistent with our constraints on the ejection rate of HVSs.

The radial velocity we measure for S5-HVS1 is 995 ± 12 km s^{-1} when considering the wavelength range 3850 to 5240 Å. We used this wavelength range in particular because the spectrum we obtained for S5-HVS1 was much higher signal-to-noise than the other sources, and we noticed the wavelength calibration below 3850 Å showed systematics. This is most likely caused by the lack of calibration lines being available for that part of the spectrum. Our measurement is consistent with the earlier found 1017 ± 2.7 km s^{-1}, so we find no indication of binarity.

Throughout this work, we assume that S5-HVS1 has been ejected by the Hills mechanism. Other mechanisms have, however, not been ruled out. A three-body interaction between a hypothetical intermediate mass black hole and Sgr A*, for instance, could have ejected S5-HVS1 and is consistent with the S-star eccentricity distribution (Generozov & Madigan 2020). However, Evans et al. (2023) find that the lack of further detected HVSs in *Gaia* would necessitate a very specific configuration of any hypothetical intermediate mass black hole and Sgr A*.

3.6.4   Distant HVS candidates

In general, the predicted HVSs in *Gaia* (see Fig. 3.7) appear similar to those found in Brown et al. (2014), being dominated by distant stars of a few M$_\odot$. The difficulty with these stars is that due to their distance, their orbits can not be confidently traced back to the GC (Irrgang et al. 2018; Kreuzer et al. 2020). Confusion with other types of sources, such as disc runaways and halo stars is therefore an issue (e.g. Brown et al. 2018).

We use the posterior on the ejection rate from Fig. 3.9 to make a prediction on the number of HVSs within the MMT HVS Survey (Brown et al. 2014). We reproduce the same colour selections outlined in section 2.1 of Brown et al. (2012). For simplicity, we consider the footprint of *SDSS* to be Dec > 0 deg for the Northern Galactic cap, and Dec > -15 deg for the Southern Galactic cap with a Galactic latitude $|l| > 30$ deg (for the sky coverage see Aihara et al. 2011). Brown et al. (2018) only studies stars with a heliocentric radial velocity transformed to the Galactic frame greater than
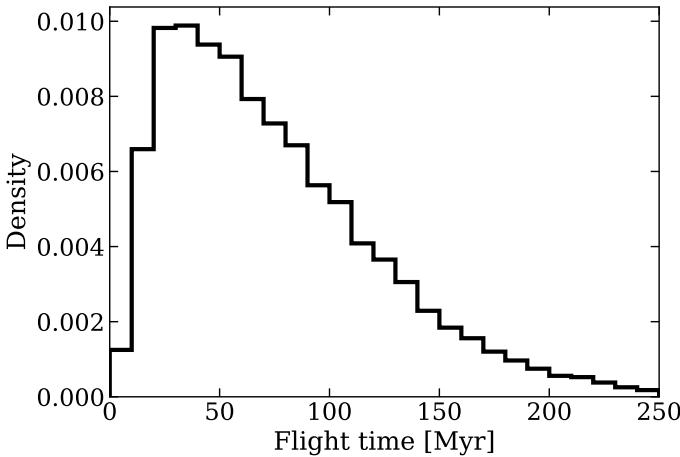
Figure 3.13: Flight time distribution for simulated HVSs matching the MMT survey selections.

$275 \ \mathrm{km\,s^{-1}}$, which we follow. In their study, Brown et al. (2018) identify seven stars with a probable origin in the GC. We use the posterior from Fig. 3.9, in combination with the above selections to calculate how many HVSs there are predicted to be within this footprint. We sample the MF index in the range $[1.7, 2.3]$ and multiply each sample by the posterior on the ejection rate for that MF index. We find that within a 68% confidence interval, we expect there to be between 0.3 and 2.8 HVSs in the MMT survey footprint. For the 95% confidence interval, we expect there to be between 0.1 and 5.1 HVSs respectively. It is thus likely that a number of the stars identified in Brown et al. (2018) are genuine HVSs. Interestingly however, there is thus a $> 2\sigma$ tension between our results and the number of GC origin HVSs found in Brown et al. (2018). Apart from a statistical deviation, we have two main hypotheses as to the origin of this tension: the ejection rate of HVSs has not been constant over the flight time of the stars identified in Brown et al. (2018), or (a fraction of) the stars identified in Brown et al. (2018) have a different origin (supported by Generozov & Perets 2022). To investigate the first hypothesis, we show the distribution of flight times for predicted HVSs in the MMT footprint in Fig. 3.13 . By comparison to Fig. 3.11, we can clearly see that the simulated HVSs matching the MMT survey selections tend to have been ejected longer ago relative to those our survey presented in this work is sensitive to. The unbound HVS candidates reported to have a probable origin in the GC have flight times between about 50 and 150 Myr (see Brown et al. 2014, table 1), which is consistent with our predictions.

Brown et al. (2018) estimated the ejection rate of unbound HVSs in the

mass range $2.5 \leq M_\odot \leq 4$ to be 1.5×10^{-6} yr^{-1}. Using our observations, we can provide an independent constraint on the ejection rate in this mass range. We do this following the same procedure as for Fig. 3.10, but now with the above mentioned mass range. Our constraints on this ejection rate are relatively insensitive to the MF varying by a factor of about 3. The 2σ upper limit peaks at about 1.3×10^{-6} yr^{-1} for an MF index $\kappa \sim 2.5$, while the mode of the posterior peaks at about 2.8×10^{-7} yr^{-1} (at the same MF index). Our maximum likelihood prediction is thus more than five times lower compared to the findings in Brown et al. (2018).

3.6.5   Prospects

Gaia DR4 will include radial velocities for sources with a limiting magnitude of $G_{RVS} < 16.2$ (Katz et al. 2023). If we consider all unbound HVSs with our ejection rate constraints from Section 3.5, we predict there to be about 3^{+4}_{-2} HVSs within the *Gaia* DR4 radial velocity catalogue if the MF index is between 1.7 and 2.3. It is, however, unlikely that all hypothetical HVSs in the footprint of the radial velocity catalogue of *Gaia* DR4 could be recognised as such. If we additionally require an accurate parallax measurement ($\varpi/\sigma_\varpi > 5$), we only expect there to be 1^{+1}_{-1} HVSs. If on top of that, we only consider HVSs with radial velocities > 500 km s^{-1}, we are left with fewer than one expected HVS.

A potentially promising means to extend to fainter stars is using radial velocities from the low-resolution *Gaia* G_{BP}/G_{RP} spectra (Verberne et al. 2024b). These spectra will be published for all *Gaia* sources in DR4. The radial velocity measurements from these spectra are most accurate for blue sources ($G_{BP} - G_{RP} \lesssim 0.7$), which is the expected main discovery space for HVSs in *Gaia* (see Fig. 3.7). Because the nominal uncertainties are high for radial velocity measurements using the G_{BP}/G_{RP} spectra, we only consider HVSs with radial velocities greater than 1000 km s^{-1}, in addition to only selecting stars with $G_{BP} - G_{RP} < 0.7$. Within this parameter range, we expect there to be 7^{+9}_{-5} HVSs in *Gaia* DR4. However, significant improvements would be required in the technique of obtaining these radial velocities from the G_{BP}/G_{RP} spectra to reliably identify these HVSs (Verberne et al. 2024b).

In addition to *Gaia* DR4, large upcoming spectroscopic survey instruments such as *DESI* (Cooper et al. 2023), *WEAVE* (Dalton et al. 2014) and *4MOST* (de Jong et al. 2019) provide the opportunity to observe much larger numbers of sources. As part of the low-resolution high-latitude survey, *WEAVE* will observe about 20 000 HVS candidates in the northern hemisphere. This is two orders of magnitude larger than the observational sample described in this work. Moreover, *WEAVE* will allow us to observe

stars down to the detection limit of *Gaia*. The precise survey strategy for *WEAVE* will be discussed in upcoming work.

3.7   Conclusion

In this work, we presented a novel selection method of HVS candidates. Utilising ground-based follow-up observations, we were able to reject all newly identified candidates. By combining this null-detection of new HVSs with sophisticated simulations, we were able to significantly improve the constraints on both the ejection rate of HVSs and the MF of the HVS progenitors. We show that the MF and ejection rate are degenerate because *Gaia* is only sensitive to HVSs more massive than about $1M_\odot$. We are therefore able to provide robust constraints on the ejection rate for HVSs more massive than $1M_\odot$ independent on the MF. Over the past $\sim 50 - 100$ Myr, the average ejection rate for these stars cannot have been higher than about 10^{-5} yr^{-1} at 2σ significance.

We use our constraints on the ejection rate and MF to evaluate previously identified HVS candidates. We find evidence that either the ejection rate of HVSs has varied significantly over the past ~ 150 Myr, or these previously identified candidates include stars that do not originate in the GC.

Our constraints predict that within a 68% confidence interval, there are between about 5 and 45 unbound HVSs in the complete *Gaia* catalogue. The majority of these stars will be main-sequence stars with a mass of a few $M_\odot$ and be at Heliocentric distances between $10 - 100$ kpc.

Data Availability

The *Gaia* data underlying this article were accessed from `https://gea.esac.esa.int/archive/`. The HVS candidate catalogue used in this work and the results from our analysis of the calibrated spectra can be accessed from `https://zenodo.org/doi/10.5281/zenodo.12179452`.

Acknowledgements

Appendix

3.A   HR density of HVS candidates

As mentioned in Section 3.2.2.2, the implied position for a star that is not moving on a radial trajectory can place it in a unphysical position in the HR diagram. In Fig. 3.14 we show where HVS candidates in *Gaia* that match all but our colour-magnitude selections end up in the HR diagram.

---

[11]`https://doi.org/10.5281/zenodo.6867957`
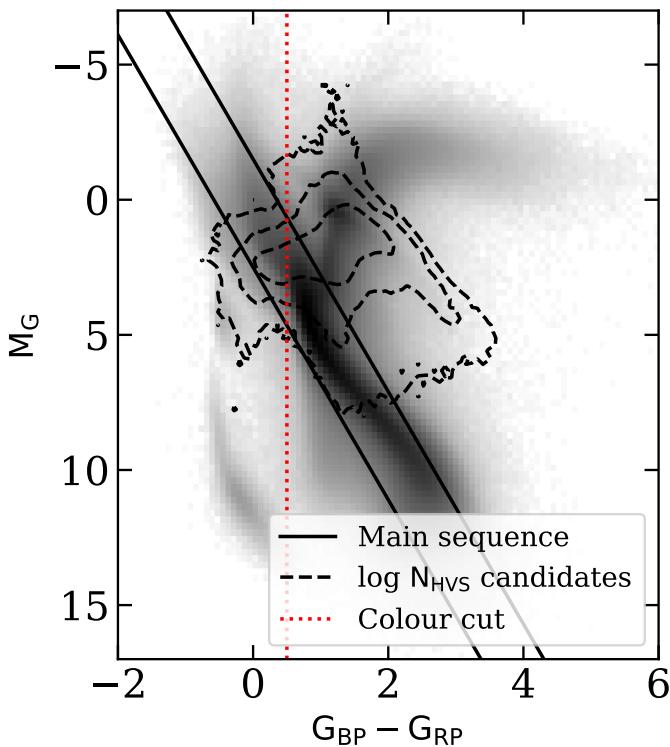
Figure 3.14: Same as Fig. 3.2, but now the overlain dashed line shows the density of HVS candidates that meet all our selections, except the colour-magnitude based ones.

Table 3.5: *Gaia* DR3 *source_id*'s and radial velocities for the five stars for which we used literature radial velocity measurements.

Gaia DR3 *source_id*	Radial velocity [km s^{-1}]	Source
2494761131458383872	-185 ± 13	*SDSS*
2700705779669486848	-73 ± 10	*LAMOST*
3792400463887610368	182 ± 13	*LAMOST*
4430178054799074688	197 ± 8	*SDSS*
4459380675615776640	-189 ± 5	*SDSS*

3.B   Radial velocities from *SDSS* and *LAMOST*

In Table 3.5 we list the five sources for which we found literature radial velocity measurements in either *SDSS* or *LAMOST*.

3.C   G-magnitude distribution *Gaia* observable HVSs

In Fig. 3.15 we show the apparent G-magnitude distribution for predicted HVSs that are observable by *Gaia* assuming an MF-index $\kappa = 2.3$. The two vertical features correspond to the main-sequence and red-giant branch. We can see that most HVSs in *Gaia* will be faint, as expected, since the observed volume increases towards faint magnitudes.

3.D   Lower limit on the ejection rate and MF

The lower limit on the ejection rate and MF is sensitive to the prior used. Throughout the main text, we used a uniform prior in λ. In Fig. 3.16 we instead show the posterior if we use a log-uniform prior equal to $1/\lambda$. This prior might be more appropriate, since we do not know the magnitude of the rate, but do know it should be positive. This prior effectively removes the lower limit, since the posterior peaks for the limit of the ejection rate approaching zero.
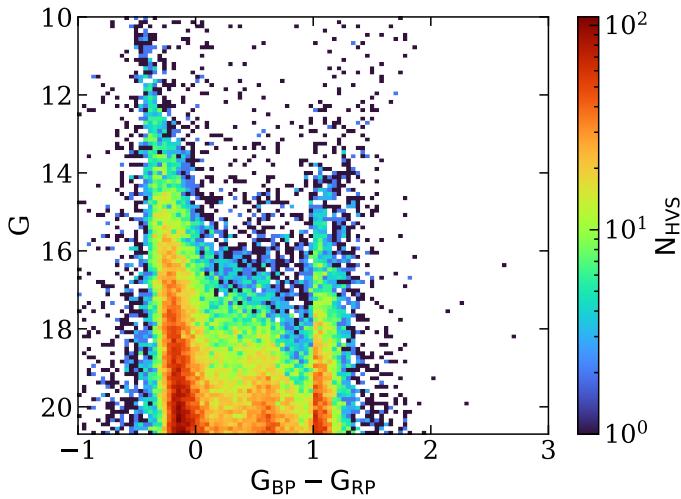
Figure 3.15: Apparent G-magnitude distribution of simulated HVSs predicted to be observable by *Gaia*.
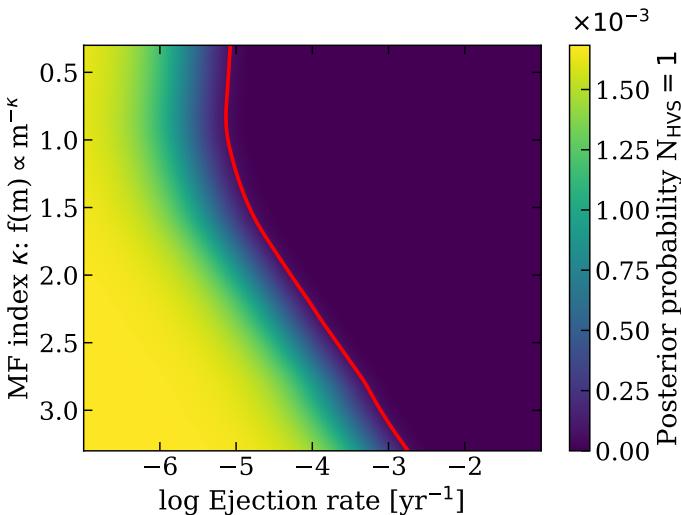

Figure 3.16: Posterior on the ejection rate and MF. Same as Fig. 3.9, but with a log-uniform prior on the ejection rate of $1/\lambda$. The 95% upper limit is indicated by the solid red line.

4 | Double progenitor origin of the S-star cluster

Abstract

The origin of the cluster of S-stars located in the Galactic Centre is tied to the supermassive black hole Sagittarius A*, but exactly how is still debated. In this paper, we investigate whether the Hills mechanism can simultaneously reproduce both the S-star cluster's properties and the observed number of hypervelocity stars. To do so, we forward-modelled the capture and disruption of binary stars originating from the nuclear star cluster (NSC) and the clockwise disc (CWD). We find that the ratio of evolved to main-sequence S-stars is highly sensitive to the origin of the binaries, and that neither the injection of binaries from the CWD nor from the NSC exclusively can reproduce all observations. However, when considering the injection of binaries from both locations, we are able to reproduce all the observations simultaneously, including the number of observed hypervelocity stars, the evolutionary stage of the S-stars, their luminosity function, and the distribution of their semi-major axes. The implications of our findings include that $\sim 90\%$ of hypervelocity stars ejected over the past ~ 10 Myr should originate from the CWD, that the main-sequence S-stars originated in the CWD, and that the evolved S-stars originated in an old stellar population such as the NSC.

4.1 Introduction

The Galactic Centre (GC) is unique for its combination of relative proximity and extreme stellar dynamics. It is, however, also a challenging environment to study due to high line-of-sight extinction and source crowding. The GC is host to the super-massive black hole Sagittarius A* (Sgr A*) with a mass of 4.3×10^6 M$_\odot$ (Eisenhauer et al. 2005; Ghez et al. 2008; Schödel et al. 2009; Gillessen et al. 2009; Boehle et al. 2016; Gillessen et al. 2017; Do et al. 2019; GRAVITY Collaboration et al. 2019; Gravity Collaboration et al. 2024) and a nuclear star cluster (NSC). In addition, Sgr A* is orbited by a cluster of mainly B-type main-sequence stars out to ~ 0.04 pc called the S-star cluster (see review by Genzel et al. 2010). The origin of this cluster remains unknown, since the strong tidal force from Sgr A* inhibits standard star formation from molecular clouds in this region (Genzel et al. 2010). Nonetheless, the S-star cluster has been intensely studied, due to the possibility of tracing individual orbits of stars and thereby constraining the mass and distance to Sgr A* (e.g. Ghez et al. 2008; Gillessen et al. 2009; Do et al. 2013; Gillessen et al. 2017; GRAVITY Collaboration et al. 2018).

Within the highly complex environment of the GC, exotic gravitational interactions can occur. Of particular interest to this work is the Hills mechanism, in which a stellar binary approaches a massive black hole (MBH) to within the tidal radius where the tidal force of the MBH overcomes the self gravity of the binary (Hills 1988; Yu & Tremaine 2003). The result is that one star is captured into a tight elliptical orbit, while the other is ejected as a hypervelocity star (HVS). HVSs can travel at $\gtrsim 10^3$ km s^{-1}, making them unbound to the Galactic potential (Hills 1988; Bromley et al. 2006; Rossi et al. 2014). The captured stars on the other hand have been suggested as a possible formation mechanism for the S-star cluster, since they should get deposited at radii similar to those of the S-stars (e.g. Gould & Quillen 2003; Ginsburg & Loeb 2006). In addition, recent observations have provided strong indirect evidence that the Hills mechanism operates in the GC (Koposov et al. 2020).

In this study, we examine whether the observed properties of both the S-star cluster and HVSs are consistent with predictions by the Hills mechanism. This will tell us whether the entire S-star cluster might have formed through the Hills mechanism or whether a more complex assembly history is required to explain the observed properties of the S-stars. What makes this study unique compared to earlier studies (e.g. Madigan et al. 2009; Zhang et al. 2013; Generozov & Madigan 2020) is that we simultaneously simulated both the S-star and HVS populations and compared both with state-of-the-art observations.

In Section 4.2, we summarise our observational knowledge of the S-star cluster. Section 4.3 describes our methods for forward modelling the

population of S-stars and HVSs. In Section 4.4, we investigate two possible origins for progenitor binaries. The discussion on our results and potential future work are presented in Section 4.5. Finally, in Section 4.6 we give our conclusions.

4.2   The S-star cluster

We start by giving an overview of the observed properties of the S-star cluster in Section 4.2.1, before discussing the effects of observational selection in Section 4.2.2.

4.2.1   Overview

The S-star cluster has isotropically distributed stellar orbits with a radial extent of roughly ~ 0.04 pc from Sgr A*; it is formed by mainly B-type massive main-sequence stars, and it is surrounded by at least one disc of massive young stars containing many O/WR stars out to ~ 0.5 pc (Levin & Beloborodov 2003; Paumard et al. 2006; Bartko et al. 2009; von Fellenberg et al. 2022). No O/WR stars are detected in the inner 0.04 pc (Genzel et al. 2010).

The S-star cluster has been observed for nearly three decades (Eckart & Genzel 1996; Ghez et al. 1998). During this time, a range of instruments have been used to study its properties with techniques including adaptive optics-assisted astrometry, integral field spectroscopy, and optical/infrared interferometry. Due to the high line-of-sight extinction ($A_V \sim 50$; Fritz et al. 2011), the study of the S-star cluster is done around 2 μm, where extinction is lower ($A_{Ks} \sim 2.5$; Schödel et al. 2010; Fritz et al. 2011).

The number of parameters needed to fully describe an orbit around the GC is six. Usually, these are sky position, proper motion, radial velocity, and acceleration in the plane of the sky (Gillessen et al. 2017). A catalogue containing orbits for 40 stars is published in Gillessen et al. (2017), most of which are in the central arcsecond. Furthermore, a catalogue of 36 young stars with orbital solutions is given in von Fellenberg et al. (2022). More recently, orbital parameters were determined for 20 stars at 2–7 arcsec from Sgr A* in Young et al. (2023).

Observations indicate that S-stars follow a thermal eccentricity distribution ($n(e) \propto e$) and travel on orbits with periods as short as decades, with the shortest observed period being about 13 yr (Gillessen et al. 2017). The most studied S-star is S2: a bright ($m_K = 14$, without extinction correction), massive star ($M_{ZAMS} \sim 14$ M$_\odot$) on a 16 year orbit (e.g. Habibi et al. 2017a). The metallicity of the S-stars is currently unknown (Habibi et al. 2017a).
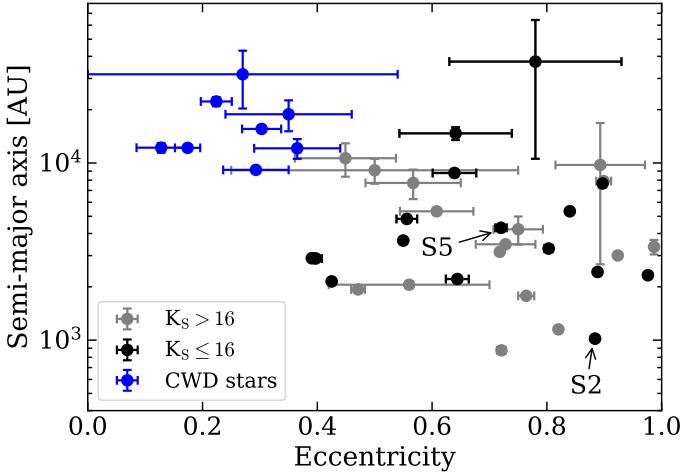
Figure 4.1: Orbital solutions for stars at the GC found in Gillessen et al. (2017), with the exception of S5, which is from von Fellenberg et al. (2022). Stars that are considered part of the young disc are coloured blue, which are S66, S67, S83, S87, S91, S96, S97, and R44. S111 is omitted, because of the hyperbolic solution. The grey points show S-stars fainter than our completeness limit, while black points show those that are brighter.

The young disc surrounding the S-star cluster, often referred to as the clockwise disc (CWD), has a top-heavy initial mass function (IMF) with a slope of around -1.7 (Lu et al. 2013; Gallego-Cano et al. 2024). The CWD is consistent with a single star formation episode about 2–6 Myr ago (Paumard et al. 2006). There is also evidence for additional discs at varying radii (e.g. von Fellenberg et al. 2022).

In Fig. 4.1, we show orbital solutions in the 2D space of eccentricity against semi-major axis, for stars belonging to either the CWD or the S-star cluster (data from Gillessen et al. 2017). Our figure clearly shows that these two stellar structures occupy different ranges of this parameter space.

4.2.2   Observational selection function

The stars for which orbital parameters have been determined cannot be seen as an unbiased dataset for the properties of the S-star cluster: accelerations are more difficult to measure for faint stars with relatively long orbital periods. To understand this bias, we need a prescription of the observational selection function. Using mock observations, Burkert et al. (2024) finds an empirical relation that connects a star's pericentre distance and eccentricity

to the chance that an orbital solution would have been found in Gillessen et al. (2017). However, Burkert et al. (2024) did not consider the apparent magnitude of a star to be part of the selection function. Since they base their selection on our ability to measure the acceleration in the plane of the sky within their mock observations, it is implicitly assumed that a radial velocity measurement was also available. To account for this additional factor, we assumed the orbital solutions to be complete down to $Ks = 16$, based on the detection limit of *SINFONI* for a typical run (Gillessen et al. 2009). Since the Gillessen et al. (2017) catalogue, only a single additional orbital solution for a $Ks < 16$ star within the Burkert et al. (2024) completeness limit has been published (S5; von Fellenberg et al. 2022). This star was reported as being consistent with the CWD, but given its high eccentricity and small semi-major axis (typical for S-stars) we instead treated it as an S-star. For $Ks < 16$, there are respectively twelve and four orbital solutions for main-sequence and evolved S-stars in our dataset. We did not consider S111, since it has a hyperbolic orbital solution (Gillessen et al. 2017). Throughout this work, our stellar magnitudes are not corrected for extinction.

4.3 Method

Now that we have discussed the properties and observations of the S-star cluster, we will describe how the Hills mechanism could explain the formation of both the S-star cluster and HVSs. We briefly review the stellar dynamics underlying this model in Section 4.3.1, and discuss the implementation of this theory in our simulations in Section 4.3.2.

A summary of our approach is as follows. For a given progenitor binary star formation history, IMF, and Hills mechanism rate, we predicted the number of observed main-sequence and evolved S-stars, the semi-major axes of their orbits around Sgr A*, and their luminosity function. Additionally, we predicted the number and properties of the related HVSs observable in the Verberne et al. (2024a, hereafter V24) survey. Fig. 4.2 shows a schematic overview of our modelling procedure. Finally, these predictions are simultaneously compared to i) the observed properties of the S-star cluster, and ii) the number of observed HVSs in the V24 survey.

4.3.1 Hills mechanism

In the Hills mechanism (Hills 1988), a stellar binary approaches a massive black hole within the tidal radius, where the tidal force of the black hole separates the binary, ejecting one star as an HVS and capturing the other star in a tight, elliptical orbit (for a review, see Brown 2015). The captured stars have been suggested as the possible origin of the S-stars (e.g. Gould & Quillen 2003; Ginsburg & Loeb 2006).
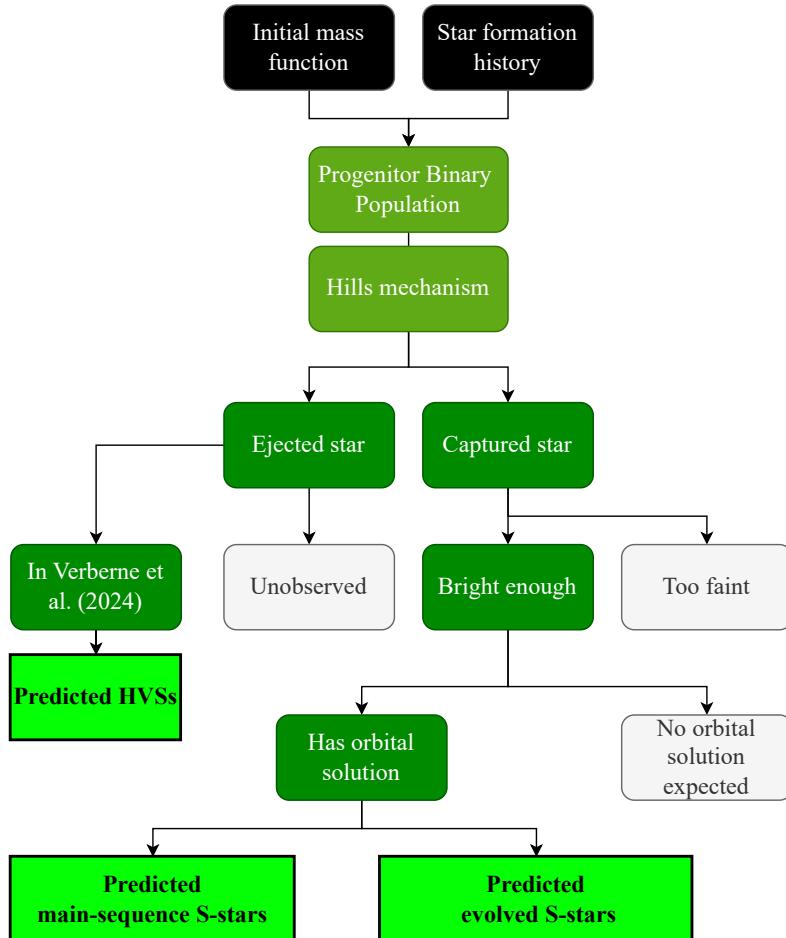
Figure 4.2: Schematic overview of the steps of our forward model that predicts the numbers of observed HVSs, main-sequence S-stars, and evolved S-stars.
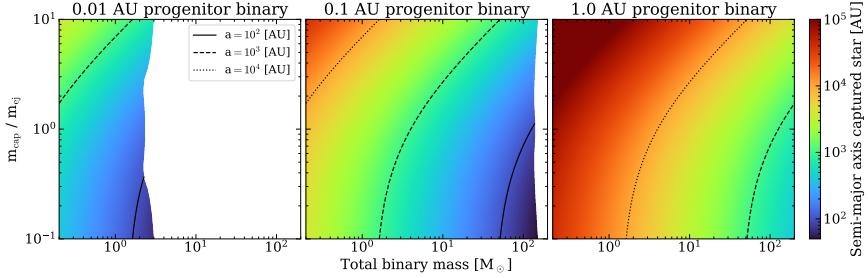
Figure 4.3: The semi-major axis of the captured star's orbit around Sgr A* as a function of the progenitor binary's total mass and mass ratio, for three different binary separations; the mass ratio is defined as the mass of the captured star over that of ejected star. We only show solutions for which the progenitor binary separation is larger than the sum of the stellar radii at zero-age main-sequence.

Progenitor binaries are expected to be on approximately parabolic orbits (Kobayashi et al. 2012), for which the ejection probability is equal for both stars and does not depend on the mass ratio (Sari et al. 2009a). In this three-body encounter, there is an energy exchange, where this amount

$$\Delta E = \alpha^2 \frac{m_{\mathrm{ej}} m_{\mathrm{cap}} G}{a_{\mathrm{b}}} \left(\frac{M_{\mathrm{bh}}}{m_{\mathrm{b}}} \right)^{1/3} \quad \text{and} \quad m_{\mathrm{b}} = m_{\mathrm{cap}} + m_{\mathrm{ej}}, \qquad (4.1)$$

is gained by the ejected star at the expense of the captured one. In the expression for ΔE, α is a prefactor of order unity that depends on the geometry of the encounter (the inclination of the binary orbital plane, the binary phase at closest approach, the pericentre radius, and the binary eccentricity), m_{ej} and m_{cap} are the mass of the ejected and captured star respectively, G the gravitational constant, and a_{b} the semi-major axis of the progenitor binary. Thus the ejection velocity of the ejected star is

$$V_{\mathrm{ej}} = \sqrt{\frac{2\Delta E}{m_{\mathrm{ej}}}} = \alpha \sqrt{\frac{2 G m_{\mathrm{cap}}}{a_{\mathrm{b}}}} \left(\frac{M_{\mathrm{bh}}}{m_{\mathrm{b}}} \right)^{1/6} \qquad (4.2)$$

(e.g. Rossi et al. 2014). The captured star, on the other hand, will travel on an orbit with a semi-major axis equal to

$$a_{\mathrm{cap}} = \frac{G M_{\mathrm{bh}} m_{\mathrm{cap}}}{2\Delta E} = \frac{a_{\mathrm{b}} M_{\mathrm{bh}}}{2\alpha^2 m_{\mathrm{ej}}} \left(\frac{m_{\mathrm{b}}}{M_{\mathrm{bh}}} \right)^{1/3} \qquad (4.3)$$

after the separation of the binary system. In Fig. 4.3, we show the semi-major axis of the captured star as a function of the total binary mass and the mass ratio of the captured over the ejected star, where we take $\alpha = 1$. This figure clearly demonstrates that Hills mechanism disruptions of binaries will

result in stars captured into orbits with semi-major axes similar to those of
the observed S-stars.

The initial pericentre distance (r_p) of the captured star is equal to the
pericentre distance of the binary around the MBH and is set by the pene-
tration factor (β) and tidal radius (r_t) defined as

$$\beta = \frac{r_\mathrm{t}}{r_\mathrm{p}} \quad \text{and} \quad r_\mathrm{t} = a_\mathrm{b} \left(\frac{M_\mathrm{bh}}{m_\mathrm{b}} \right)^{1/3}. \tag{4.4}$$

We can thus write the initial eccentricity of the captured star as

$$e_\mathrm{init} = 1 - \frac{r_\mathrm{t}}{\beta a_\mathrm{cap}} = 1 - \frac{a_\mathrm{b}}{\beta a_\mathrm{cap}} \left(\frac{M_\mathrm{bh}}{m_\mathrm{b}} \right)^{1/3}. \tag{4.5}$$

We refer to this pericentre distance and eccentricity as initial, because the
angular momentum of a star deposited in the S-star cluster is expected to
change rapidly due to dynamical resonant relaxation processes (e.g. Rauch
& Tremaine 1996; Hopman & Alexander 2006; Binney & Tremaine 2008;
Antonini & Merritt 2013; Generozov & Madigan 2020). The timescale on
which the energy of the orbit changes is much longer and happens on the
two-body relaxation timescale. The precise timescales on which the angular
momentum and energy change are still an active field of research (see review
by Alexander 2017). Additionally, stellar collisions might play an important
role near Sgr A* (e.g. Sari & Fragione 2019; Rose et al. 2023; Rose & Mockler
2024; Ashkenazy & Balberg 2025).

4.3.2   Simulations

Having discussed the theory behind the Hills mechanism and its potential
to explain the S-star cluster, we now describe our simulations.

In our model, progenitor binaries are characterised by mass, age, metal-
licity, mass ratio, and binary separation. We assumed stars with solar metal-
licity ($Z_\odot = 0.0142$; Asplund et al. 2009) for simplicity and evaluated the
impact of this assumption in Section 4.5.3. Other properties of our progen-
itor binaries were drawn from probability distributions. In particular, the
distribution of the mass ratio is $f(q) \propto q^\gamma$, where $0.1 \leq q \leq 1$ and we used
an orbital period (P) distribution of the form $\log(P/[1 \text{ sec}]) \propto (\log \mathrm{P}/\mathrm{P}_0)^\pi$,
where P_0 is 1 sec. Our prior ranges are $-2 \leq \gamma \leq 2$ and $0 \leq \pi \leq 2$. The
mass and age of the binaries were determined given a star formation history
and an initial mass function (IMF). Using the star formation history and
the IMF, we can compute the 'present-day' mass function at any look-back
time. For a Hills mechanism disruption at look-back time t, we calculated
the number of stars alive per unit mass m and unit age A from

$$\frac{dN}{dm dA}(m, \ A \mid t) = \mathrm{SFR}(A + t) \ C(A) \ P(m|A), \tag{4.6}$$
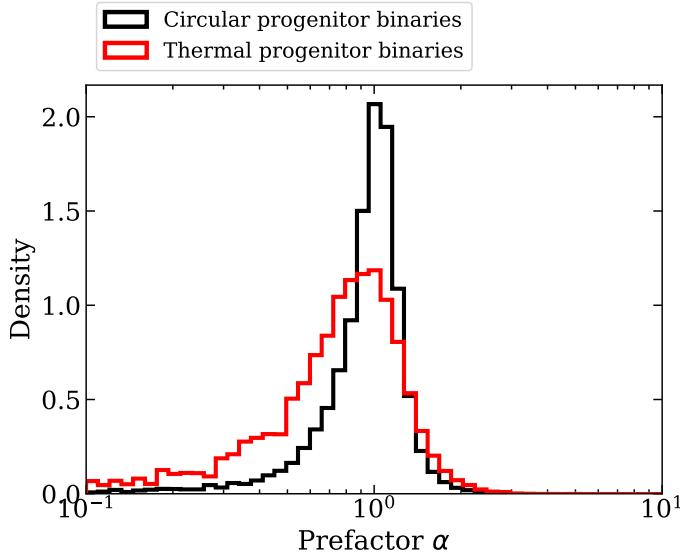
Figure 4.4: Prefactor α that we sample over to account for random orbital phases and inclinations at the time of disruption (Sersante et al. in prep.) The black and red lines show the distribution for circular and thermally distributed progenitor eccentricities respectively.

where $\mathrm{SFR}(A+t)$ is the star formation rate at time $A+t$, $C(A)$ the fraction of stars alive at age A given an IMF, and $P(m \mid A)$ the mass function for stars alive of age A, which is the IMF truncated at the maximum mass of a star alive with age A. If a star survives till the present day, it has an age $A + t$. The fraction of stars alive at age A is calculated by

$$C(A) = \int P(m) \, (m < m_{\mathrm{max}}(A)) \, dm, \tag{4.7}$$

where $P(m)$ is the normalised IMF, and $m_{\mathrm{max}}(A)$ the maximum mass star alive at age A. We assumed the rate of binary disruptions at a point in time is proportional to the number of stars alive at that time. The Hills mechanism rate we report is always the current rate.

Unless stated otherwise, we assumed that the progenitor binaries are disrupted isotropically, which means that the resulting ejected and captured stars have random orientations. We calculated the energy that is exchanged in the tidal separation from equation 4.1. We used the results from the simulations in Sersante et al. (in prep.) and randomly sampled over the prefactor α, using the distribution shown in Fig. 4.4. We assumed here that the progenitor binaries are on circular orbits in the full lose-cone regime

(Merritt 2013) such that β (equation 4.4) is distributed log-uniformly. We also investigated a thermal eccentricity distribution in Section 4.5.5.

4.3.2.1 S-star predictions

After the binary is disrupted, we calculated if the captured star is within the stellar tidal disruption radius of Sgr A* (Rees 1988). If so, the star was removed from the simulation, if not we assumed that the eccentricity is rapidly thermalised such that $n(e) \propto e$ while the orbital energy remains constant. The reason for this is that simulations point out that resonant relaxation can thermalise the S-star eccentricity distribution over $\sim 10^7$ yr (e.g. Perets et al. 2009b; Antonini & Merritt 2013; Generozov & Madigan 2020), which is of the order of the age of the CWD. The extended mass required for resonant relaxation to be fast enough is similar to the 1σ upper limit on the enclosed mass within the S2 star orbit (Generozov & Madigan 2020; Gravity Collaboration et al. 2024). After the orbits were thermalised, we again removed any stars whose pericentre lies within the stellar tidal disruption radius of Sgr A*. On the other hand, we kept the semi-major axes of stars fixed, because we assumed negligible changes in the orbital energy over the timescales of interest (see Section 4.5.4). We followed the stellar evolution of the captured stars using `MIST` until the present epoch and determined the Ks magnitude (Dotter 2016b; Choi et al. 2016b). We assumed a uniform foreground extinction of $A_{K\mathrm{s}} = 2.5$ (Schödel et al. 2010; Fritz et al. 2011) towards the GC and corrected the photometry accordingly.

We only investigated stars with $K\mathrm{s} < 16$, because of observational completeness (see Section 4.2.2). We assigned to each remaining star a probability of having a known orbital solution following Burkert et al. (2024). We used the equivalent evolutionary point (EEP) number in `MIST` to separate main-sequence stars from evolved stars at the terminal-age main-sequence (EEP = 454). This leaves us with a number of predicted main-sequence and evolved S-stars that we can compare to observations in Gillessen et al. (2017).

4.3.2.2 HVS predictions

Simultaneously to the analysis of the captured stars presented above, we tracked the ejected stars using `Speedystar`[1] (Rossi et al. 2017; Marchetti et al. 2018; Contigiani et al. 2019; Evans et al. 2022a,b). This code simulates the ejection, propagation, and evolution of HVSs and provides synthetic photometry and *Gaia* observables for simulated HVSs. Our procedure is identical to that presented in V24, section 4.3, except that we now considered realistic star formation histories as described in Section 4.4. This was incorporated through the prescription in equations 4.6 and 4.7.

---

[1] `https://github.com/fraserevans/speedystar`

V24 performed a dedicated survey of HVSs. They rely on the near-radial trajectories of HVSs to identify candidates and provide follow-up observations. The survey has a clearly defined selection function, allowing us to assess the detectability of any given simulated star. V24 identify a single previously known HVS, S5-HVS1 (Koposov et al. 2020): a 1700 $\mathrm{km\,s^{-1}}$ star at about 9 kpc from us that originates in the GC, where it was ejected about 4.8 Myr ago. This is the only star unambiguously linked to a GC ejection and provides strong indirect evidence of the Hills mechanism.

For a given star formation history, we forward modelled the expected number of HVSs in the dedicated survey presented in V24. The forward modelling leaves us with a predicted (simulated) population of HVSs that we can compare to the one HVS found observationally in V24. Therefore, when we mention the predicted number of HVSs, we are referring to the predicted number of HVSs in the V24 survey. The main constraining power of these observations is in providing an upper limit to the Hills mechanism rate, since the detection of only one HVS is very informative.

We consider HVSs to be the best comparison point since, following their definition in V24, they are almost certainly formed through the Hills mechanism. Although different formation mechanisms might contribute to the observed S-star population, we know that Hills mechanism disruptions will put stars into S-star-like orbits (Section 4.3.1).

4.4   Progenitor populations

In this section we will investigate two specific progenitor populations for binaries disrupted through the Hills mechanism.

4.4.1   First scenario: binaries originated from the NSC

We start by considering if binaries originating from parsec scales in the NSC (half light-radius around $\sim 5-10$ pc) can explain the observed properties of the S-star cluster and HVSs.

The origin of progenitor binaries on which the Hills mechanism operates is uncertain. If stellar relaxation dominates the influx of disrupted binaries, the binaries will tend to originate at or beyond the sphere of influence (~ 3 pc) of the MBH (Perets et al. 2007). Additionally, Penoyre et al. (in prep.) find that the axi-symmetry of the central potential of the Milky Way produces a collisionless flux of binaries from around the sphere of influence radius that dominates over the collisional rate. Thus, it appears unavoidable that stellar binaries from the NSC are set on orbits that result in their tidal separation, although the rate is still uncertain. We therefore assumed here
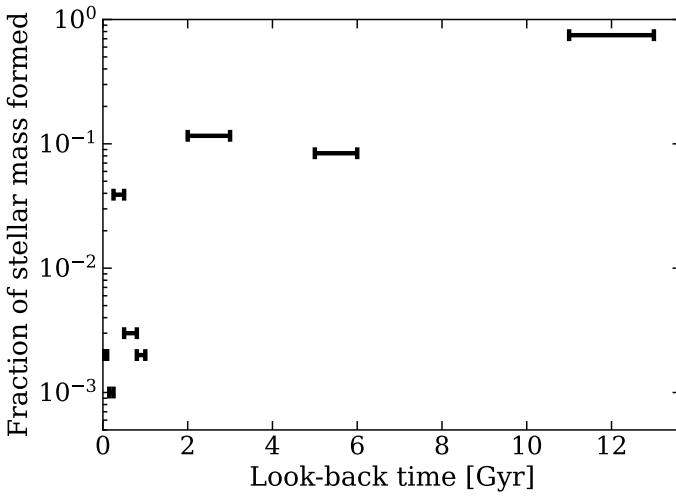
Figure 4.5: Look-back time against the fraction of stellar mass formed over each star formation epoch, which encapsulates the star formation history of the nuclear star cluster as determined in Schödel et al. (2020) for a $2Z_\odot$ model. The bars indicate the bin sizes of Schödel et al. (2020) within which we assume the star formation rate was constant.

that all binaries originate from the NSC and have the same star formation history as the NSC.

The NSC stellar population is consistent with having a canonical IMF (Kroupa 2001). On the other hand, its star formation history is still debated, but it is known that it has not been constant (e.g. Pfuhl et al. 2011; Schödel et al. 2020; Chen et al. 2023), with most of the population having formed billions of years ago. The model of the star formation history we used is the super-solar metallicity ($2Z_\odot$) model excluding the very young stars in the CWD from Schödel et al. (2020, table 3). In Fig. 4.5, we plot the fraction of the total stellar mass formed as a function of look-back time. We can see that the NSC is dominated by old stars, with only a few percent having been formed in the last billion years. We assumed that the star formation rate was constant within the intervals reported in Schödel et al. (2020).

We only considered captured stars with $M > 0.8$ $M_\odot$, since lower mass stars are not bright enough to be observed near the GC at $Ks < 16$. Evolved 1 $M_\odot$ stars, on the other hand, might still be observed at those magnitudes according to MIST. The sample is dominated by relatively low-mass binaries, since the bulk of the population is very old.

In the left panel of Fig. 4.6, we show our results for binaries originating from the NSC. Since the predicted numbers linearly scale with the Hills mechanism rate, their ratio remains constant: for every observed HVS we
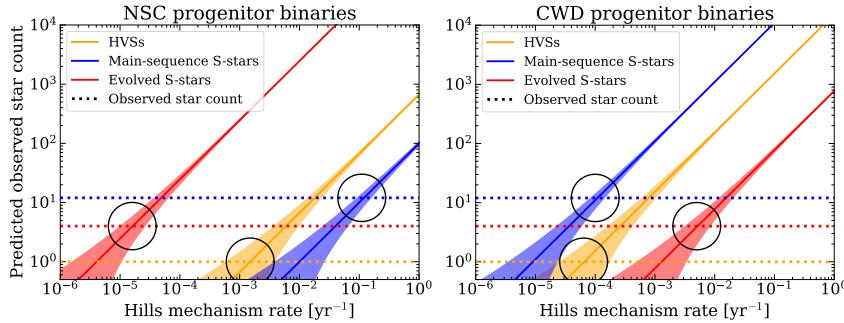
Figure 4.6: Hills mechanism rate against the predicted number of observed HVSs, main-sequence S-stars, and evolved S-stars. The dashed horizontal lines correspond to the observed number in each of these categories. The black circles highlight where the model and data agree. **Left:** the model predictions when progenitor binaries for the Hills mechanism originate from the NSC. **Right:** same as the left panel, but for binaries originating from the young CWD.

expect 0.15 and 3.7×10^2 main-sequence and evolved S-stars respectively within our brightness limit. This is incompatible with observations because

- for every observable main-sequence S-star, around 2.5×10^3 evolved S-stars should be visible unless they are somehow depleted (see Section 4.5.2),

- to account for the number of observed main-sequence S-stars, the Hills mechanism would deplete the NSC in ~ 100 Myr ($M_{NSC} \sim 2 \times 10^7 M_{\odot}$; Feldmeier-Krause et al. 2017b),

- and there should be $\sim 10^2$ HVSs in the V24 survey footprint.

4.4.2 Second scenario: binaries originated from the CWD

We showed in the previous section that Hills mechanism disruptions of binaries originating from the overall population of the NSC cannot simultaneously explain the observed properties of the S-star cluster and HVSs. Here we will investigate an alternative progenitor binary population, originating from the CWD (see also Madigan et al. 2009, 2014; Koposov et al. 2020; Generozov & Madigan 2020; Generozov 2021). This alternative scenario is largely inspired by the overlap between the orbital plane of the progenitor binary of S5-HVS1 around Sgr A* and that of the CWD (see Fig. 10 in Koposov et al. 2020), which is suggestive since HVSs are ejected in the plane of

the progenitor binary around the MBH. The age of the CWD is uncertain, but likely less than 10 Myr (Lu et al. 2013).

We assumed a single star formation episode 10 Myr ago, and we modeled it with a Gaussian distribution centred at that epoch, with a standard deviation of 1 Myr, and additionally a top-heavy IMF with a slope of -1.7 (Lu et al. 2013; Gallego-Cano et al. 2024). Only stars more massive than about 7.5 $M_\odot$ are bright enough to pass our brightness limit given the age of the cluster. A clear difference with respect to the NSC scenario is the relative abundance of massive binaries, due to the younger ages of the stars and their top-heavy IMF.

Because HVSs are ejected in the plane of the progenitor binary around the MBH, we included this selection effect in our analysis of the expected number of HVSs in the V24 survey. This was done by only considering HVS ejections in the plane of the CWD, described by an inclination and longitude of the ascending node of $(i, \Omega) = (130°, 96°)$, and a half-width at half-maximum of about $15°$ (Yelda et al. 2014). For a description of this coordinate system, see Lu et al. (2009).

In the case of the NSC, we assumed that the progenitor binaries were on parabolic orbits around the MBH. Because the CWD is much closer to Sgr A*, the centre of mass of the progenitor binary might approach SgrA* on a trajectory with significantly lower total (negative) energy (i.e. lower eccentricity). To account for this, we sampled from the radial surface density of the CWD ($\sim r^{-2}$; e.g. von Fellenberg et al. 2022) between 1 and 8 arcseconds from Sgr A*. The sampled semi-major axis sets the initial energy of the binary. We then calculated the semi-major axis of the captured star around Sgr A* taking into account this initial energy. In practice, this has a limited impact, since the initial energy is much smaller than the exchanged energy by a factor of $\sim 10^{-5}$ for observable S-stars in our simulations.

In the right panel of Fig. 4.6, we show the Hills mechanism rate against the predicted number of observed HVSs, main-sequence S-stars, and evolved S-stars. Given this star formation history, we expect, within our brightness limit, 7.2 and 0.05 main-sequence and evolved S-stars, respectively for every HVS in the V24 survey. This means that the number of observed HVSs in V24 is consistent with the twelve observed main-sequence S-stars within our brightness limit. If S5-HVS1 was ejected from a binary originating in the CWD, it is thus likely that most if not all main-sequence S-stars in the S-star cluster were deposited through the Hills mechanism. On the other hand, this model has trouble explaining the evolved S-stars, since for every evolved S-star it predicts around 140 main-sequence S-stars. This further demonstrates that if we assume the S-star cluster is exclusively formed through the Hills mechanism, the ratio of evolved to main-sequence S-stars is sensitive to the star formation history of the progenitor binary population. However, there is no obvious single progenitor population of binaries that can explain the number of HVSs, main-sequence S-stars, and evolved

S-stars simultaneously.

4.4.3   Combined CWD and NSC progenitors

We argued in Section 4.4.1, that progenitor binaries exclusively originating from the NSC cannot explain the observed properties of HVSs and main-sequence/evolved S-stars. However, we would like to stress that theoretically, one would expect there to be a finite disruption rate > 0 yr^{-1} of binaries originating from the NSC. Furthermore, a young progenitor population is needed to produce sufficient main-sequence S-stars. A natural interpretation of the observed properties of the S-star cluster compared to our simulations is that the main-sequence S-stars were recently deposited on their current orbits by disruptions from a young population of stars, such as the CWD, while the evolved S-stars were deposited through disruptions of mainly old, low-mass binaries originating in, for example, the NSC. Here we investigate if this scenario is able to explain all the observables simultaneously.

If progenitor binaries are disrupted from both the NSC and CWD, we expect there to be two rates that make up the total Hills mechanism rate: a constant rate caused by the inflow of binaries from the NSC (Γ_{NSC}) and a transient rate from the CWD (Γ_{CWD}). Our model is a linear combination between these two rates. The three observables we fitted to are the number of HVSs, main-sequence S-stars, and evolved S-stars that are observed. We show the resulting log-likelihood of Γ_{CWD} and Γ_{NSC} in Fig. 4.7 in combination with the confidence regions of the 68th and 95th percentile. We find that the rate of the Hills mechanism for binaries originating from the NSC is $1.4^{+0.9}_{-0.6} \times 10^{-5}$ yr^{-1} and that from the CWD is $1.0^{+0.3}_{-0.2} \times 10^{-4}$ yr^{-1}. The current rate of Hills mechanism disruptions from the CWD would thus have to be about an order of magnitude higher than the constant background rate of binary disruptions from the NSC in order to fit the observables. Over the past ~ 10 Myr, we therefore predict that $\sim 90\%$ of ejected stars are young ($\lesssim 10$ Myr). Our derived rates are comparable to old full and empty loss-cone rate estimates (Hills 1988; Yu & Tremaine 2003). In this scenario, S5-HVS1 is about 150 times more likely to have been ejected from the young progenitor population; this, combined with the age estimate for S5-HVS1 of less than 10^8 yr (Koposov et al. 2020) and the alignment of the CWD plane with that of its progenitor binary, leads us to argue that S5-HVS1 came from the CWD and shares the same age.

Now that we have a combined CWD and NSC model, we evaluate if this model can reproduce all the observables of both the S-star cluster and HVSs in Fig. 4.8. Firstly, we recount that neither progenitor population, when considered individually, was able to simultaneously predict the observed numbers of HVSs, main-sequence S-stars, and evolved S-stars. However, as
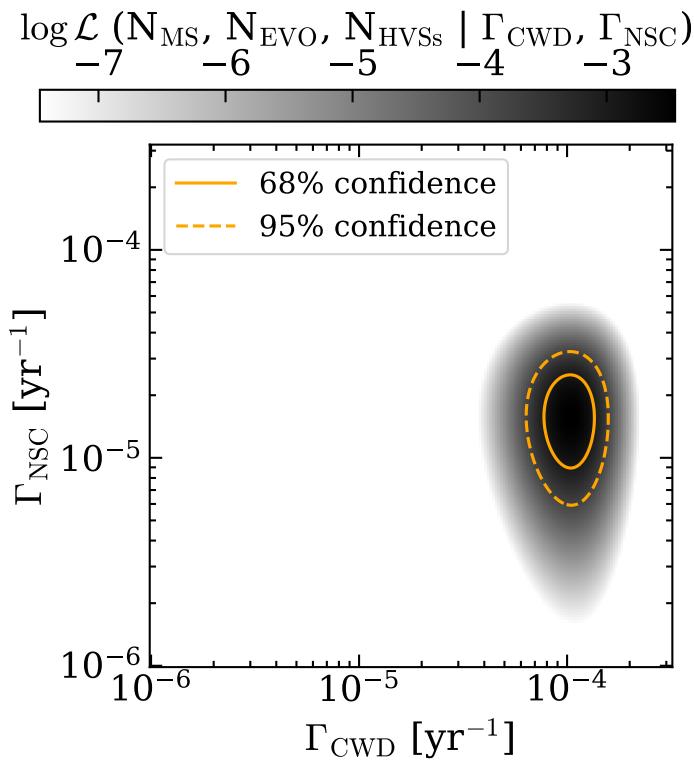
Figure 4.7: Log-likelihood on the Hills mechanism rate for our combined CWD and NSC model. In orange we show the 68th and 95th percentile confidence regions of our fit.
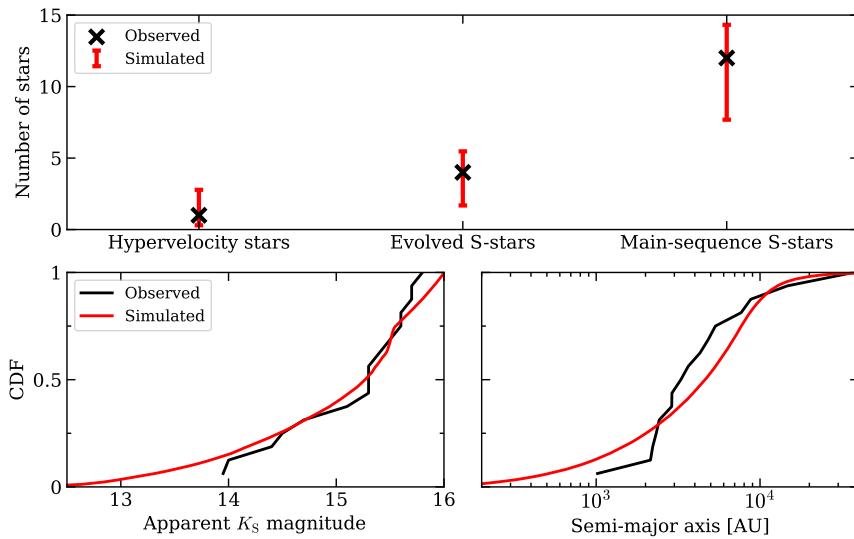
Figure 4.8: Comparison of our best fit model, including Hills mechanism disruptions from both the NSC and CWD, with observations. **Top:** number of stars from our simulation compared to the observations for HVSs, evolved S-stars, and main-sequence S-stars. The error bars show the Poisson error on the predicted numbers from the model. **Bottom left:** cumulative distribution function of the visual Ks magnitude of S-stars for our simulations compared to the observations. **Bottom right:** cumulative distribution function of the semi-major axis of S-stars for our simulations compared to the observations.

shown in the top panel of Fig. 4.8, the combined model successfully accounts for all three of these quantities at once. Moreover, both the Ks-band and semi-major axis distributions are consistent between our model and the observations, with KS-test p-values of 0.88 and 0.55 respectively.

Stellar tidal disruption events

We find that for both our progenitor populations about 20% of Hills mechanism disruptions result in a stellar tidal disruption event (TDE). The resulting TDE rate of the order 10^{-5} to 10^{-6} yr^{-1} is consistent with extragalactic TDE rates (Stone & Metzger 2016). However, this estimate is strongly influenced by our assumption of a full loss-cone. In the empty loss-cone regime, the Hills mechanism contribution to the TDE rate would be about two orders of magnitude lower. Our estimated Hills mechanism contribution to the TDE rate of 10^{-5} yr^{-1} should thus be seen as an upper limit.

4.5   Discussion

In this work we have investigated whether the Hills mechanism can explain both the observed properties of the S-star cluster and HVSs, and we found a plausible formation scenario that is consistent with data (see Section 4.4.3 and Fig. 4.8). In this Section we critically review our model assumptions in the first five subsections. We investigate further the comparison between the semi-major axis distribution of our simulated population and the observed one in Section 4.5.6. We discuss future prospects in Section 4.5.7.

4.5.1   Star formation history

In this work, we have investigated two progenitor binary populations. The most important parameter that sets the ratio of main-sequence to observed S-stars is the star formation history. To demonstrate this, we calculated the amount of time a star in the GC can be above our brightness limit as a function of its initial mass, and we show this in Fig. 4.9. First of all, we notice that only main-sequence stars more massive than about 4 $M_\odot$ can be observed out to the GC, which have a main-sequence lifetime of $\lesssim 170$ Myr. Secondly, we notice that evolved, low-mass stars can be brighter than our threshold for about an order of magnitude longer than massive main-sequence stars. For an old population of stars, such as the NSC, the result will thus be that most of the stars above the brightness limit will be evolved, low-mass stars. For a young population, such as the CWD, most of the stars above the brightness limit will instead be massive main-sequence stars. The ratio of evolved to main-sequence stars therefore traces the star formation history.
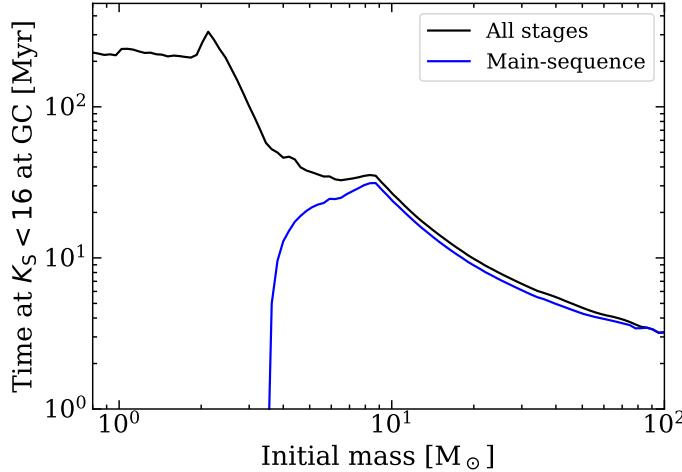
Figure 4.9: Initial stellar mass against the time spent with a magnitude of less than 16 in the Ks band, when the star is located at the GC. We utilise the stellar evolution code `MIST`.

Our results show that a young progenitor population is needed to produce sufficient main-sequence S-stars to match observations. Our conclusion here is in line with Generozov et al. (2025), who, through a different argument, recently found the need for a young population of stars to explain the main-sequence S-stars. However, we argue that binary disruptions exclusively from a young population such as that of the CWD cannot explain the presence of the evolved S-stars.

Our conclusions most importantly relate to the star formation history of the progenitor populations. Alternative populations of possible old progenitor binaries, such as the nuclear stellar disc (Schödel et al. 2023), result in similar conclusions. This is particularly relevant, since massive perturbers might increase the disruption rate of binaries originating at these radii by several orders of magnitude (Perets et al. 2007). In the case of the young progenitor binary population, we consider the CWD to be the most likely origin, given its close proximity to Sgr A* and its young age.

As a more agnostic star formation history, we also investigated a constant star formation rate over the past 13 Gyr and a canonical IMF slope of -2.3 for our progenitor population. Based on these simulations, we expect there to be 0.11 and 1.5×10^2 main-sequence and evolved S-stars respectively within the brightness limit for every HVS in the V24 survey. This is qualitatively similar to the results for the NSC progenitor origin and also incompatible with observations.

In the context of star formation history, we would also like to point out that the commonly used single power-law slope for the (present-day) mass function implicitly assumes a constant star formation rate with an IMF that is not explicitly defined. Since the stellar lifetime roughly scales with $M^{-2.5}$ (e.g. Ryan & Norton 2010), the IMF can be approximated by subtracting about -2.5 from the power-law slope of the current mass function if star formation was constant over the last several billion years. For a population of stars with a current mass function slope of -2.3, the IMF would thus be about $\frac{dN}{dM} \propto M^{0.2}$, unless star formation was not constant. Such a highly top-heavy IMF is not realistic even for the top-heavy CWD (e.g. Bartko et al. 2010; Lu et al. 2013; Gallego-Cano et al. 2024).

4.5.2   Stellar collisions

So far, we have ignored the effects of stellar collisions. In general, in a dense cluster of stars and black holes with a high velocity dispersion, like that occupied by the S-stars, stellar collisions might become important. There is currently no consensus as to the importance of collisions for shaping the S-star cluster. Collisions may shape the density profiles of NSCs on a scale similar to that of the S-star cluster, depleting stars at small radii and potentially even leading to mergers (Rose et al. 2023; Rose & Mockler 2024; Ashkenazy & Balberg 2025). Generozov et al. (2025), however, find that collisions only have a minor impact on their results.

It should also be noted that there is no consensus as to the net effect of collisions. Depending on the mass ratio, evolutionary stage of each component involved, impact parameter, and impact velocity the result of a collision will vary drastically (e.g. Benz & Hills 1987; Trac et al. 2007; Mastrobuono-Battisti et al. 2021; Rose et al. 2023; Rose & Mockler 2024).

In Fig. 4.10, we show the collisional timescale as a function of eccentricity and semi-major axis for a Bahcall & Wolf (1976) cusp following Sari & Fragione (2019). We can see that stellar collisions might be important for a significant part of the parameter space. Particularly for low-mass stars that could otherwise reside in the S-star cluster for several billion years during their main-sequence, collisions might play an important role. The massive main-sequence S-stars will not be significantly affected, since their main-sequence lifetimes tend to be shorter than the collisional timescale shown here.

Collisions are thus relevant for our NSC progenitor population model. To ensure our conclusions are robust, we also calculated our NSC progenitor results for ejections only over the past 100 Myr. The effect of only considering recent ejections is that the predicted number of evolved S-stars decreases. However, even if we only consider these recent binary disruptions, we would still expect about 24 evolved S-stars for every main-sequence S-
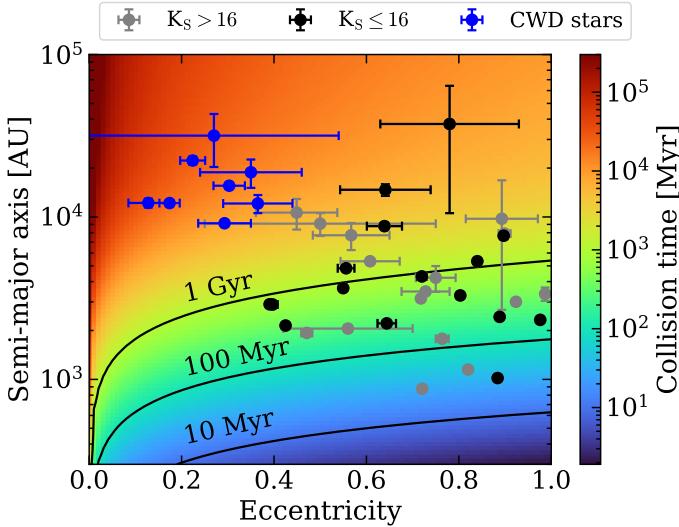
Figure 4.10: Collisional timescale of a 10 $M_\odot$ star with a 5.6 $R_\odot$ radius as a function of stellar eccentricity and semi-major axis. The stars with orbital solutions (see Fig. 4.1) are overplotted.

star. This means that even when only considering recent disruptions, an old progenitor binary population is inconsistent with observations. Because possible depletion will remove a fraction of the simulated S-stars, we expect that the Hills mechanism rate constraint for the NSC in our combined model is a lower limit. The true rate would need to be higher to account for the depletion rate over time.

4.5.3   Metallicity

The metallicity of the S-stars is, as of yet, unknown (Habibi et al. 2017a). For this reason, it is unclear what metallicity to assume for the progenitor binary population.

In the case of the NSC, most studies find a super-solar mean metallicity of [Fe/H] $\sim$ 0.2 to 0.3 (e.g. Schultheis et al. 2019; Schödel et al. 2020) with a minor population of a few percent being metal poor (e.g. Do et al. 2015; Feldmeier-Krause et al. 2017b; Gallego-Cano et al. 2024). The metallicity of the CWD, like that of the S-stars, is unknown, though in situ star formation in the GC is generally expected to be (super) solar (e.g. Do et al. 2015; Feldmeier-Krause et al. 2017b; Schödel et al. 2020).

Fortunately, the effect of the assumed metallicity in this study is minimal, since we only consider if stars are on the main-sequence or evolved, in
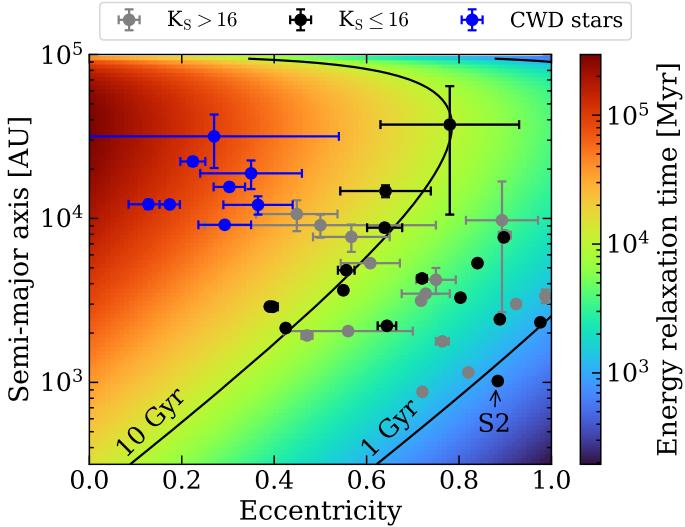
Figure 4.11: Two-body energy relaxation timescale as a function of eccentricity and semi-major axis assuming a Bahcall & Wolf (1976) cusp. The stars with orbital solutions are overplotted (see Fig. 4.1). The two solid lines mark the 1 and 10 Gyr relaxation times.

combination with a brightness limit. If we assume a metallicity of [Fe/H] = 0.3, we find that the number of main-sequence S-stars to evolved stars changes from 7.0 to 7.5 and the number of evolved S-stars from 0.05 to 0.04 for the CWD binary origin. Also for the HVS observations, metallicity is unimportant, because of the large coverage of colour-magnitude space (V24).

4.5.4 Energy diffusion

In this paper, we assumed that the energy of any captured stars does not evolve after the tidal interaction. In reality, the energy diffuses on the two-body relaxation timescale (e.g. Bar-Or & Alexander 2016; Sari & Fragione 2019). If we use the formalism from Sari & Fragione (2019), while assuming a Bahcall & Wolf (1976) cusp of stars, we get the two-body relaxation timescale in energy shown in Fig. 4.11. We can see that only S2 has a shorter energy relaxation time than 1 Gyr, which is still much longer than its main-sequence lifetime (Habibi et al. 2017a). We noted in Section 4.5.1, that only main-sequence stars with masses above ~ 4 M$_\odot$ are expected to pass our brightness limit, for which the main-sequence lifetime is about 170 Myr. This is much shorter than the typical two-body relaxation timescale

for the energy shown in Fig. 4.11. We conclude that for the assumed density profile, our assumption that the captured stars' energy is constant is safe for main-sequence stars. If, instead, a star was deposited > 1 Gyr ago, the energy might have changed significantly. This is the case for very old disruptions from the NSC progenitor population discussed in Section 4.4.1.

4.5.5   Progenitor binary eccentricity

The internal eccentricity of progenitor binaries influences the energy that is exchanged in the Hills mechanism through the prefactor that we plot in Fig. 4.4. If we assume a thermal eccentricity distribution for our progenitor binaries originating in the CWD, the number of main-sequence and evolved S-stars per observed HVS change from 7.2 and 0.05 to 7.1 and 0.06 respectively. Moreover, if we compare the cumulative distribution functions of the simulated semi-major axes of the S-stars, we find a maximum difference of about 0.02, which is well below statistical uncertainties given the number of observed S-stars. Assuming a thermal eccentricity distribution or a circular one makes thus little difference for our results. This is explained by considering that disruptions that result in observable S-stars tend to be of high energy, which means high values for the prefactor. For high prefactors, we can see from Fig. 4.4 that the thermal eccentricity distribution is very similar to the circular one.

4.5.6   The semi-major axis distribution

So far, we have not discussed the semi-major axis distribution of observed and modelled S-stars in-depth. The reason is that it is the most difficult observable to interpret. One needs to model the impact of collisions in combination with energy and angular momentum diffusion and observational biases, before attempting to compare a simulated semi-major axis distribution with an observed one.

We showed in Fig. 4.8 how our model compares to the observed semi-major axis distribution. We found the two to be consistent, but the relatively limited number of S-stars for which we can assume the observations to be complete limits our sensitivity. The lack of predicted S-stars at semi-major axes greater than about 2×10^4 AU is not surprising, since this is caused by the observational selection function (Burkert et al. 2024). However, our model is able to naturally explain the lack of S-stars at semi-major axes below about 10^3 AU, without the need for stellar collisions. The limiting factors for producing stars at such radii are the energy that can be exchanged in the Hills mechanism and the high initial eccentricity that can cause TDEs to occur for small semi-major axes.
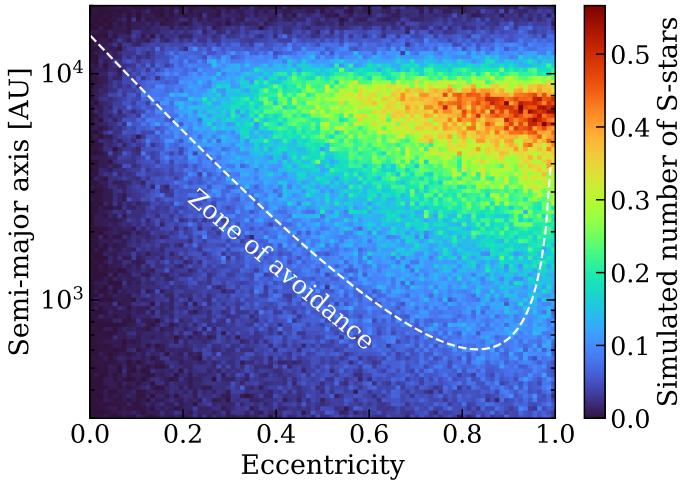
Figure 4.12: Eccentricity against semi-major axis for our simulation of the CWD origin binaries. The colour bar shows the number of S-stars predicted for an ejection rate of 10^{-2} yr^{-1}. In white we show the boundary of the zone of avoidance identified in Burkert et al. (2024). The drop in the number of simulated S-stars at large semi-major axes is due to the selection function from Burkert et al. (2024).

In the literature, a 'zone of avoidance' was identified in pericentre distance against eccentricity where no S-stars have been found, even though observations should have identified them if they exist. Burkert et al. (2024) defines the region by $\log(r_{\mathrm{p}}) < 1.57 + 2.6 \times (1 - e)$, where r_{p} and e are the pericentre distance and eccentricity respectively. In Fig. 4.12, we show the zone of avoidance in semi-major axis and eccentricity space (which are independent quantities), in combination with the predicted S-star cluster distribution of our CWD origin simulations. The figure demonstrates that our simulations can naturally explain the lack of observed stars in the zone of avoidance. The fundamental reasons being the limit on the energy that can be exchanged in the Hills mechanism (equation 4.1), captured stars ending up as TDEs at small semi-major axis (since their initial eccentricity is close to 1), and subsequent relaxation of the eccentricity distribution. Our results therefore agree with the conclusion of Generozov et al. (2025), that the Hills mechanism can naturally explain the lack of orbital solutions in the zone of avoidance.

4.5.7   Prospects

Definitively identifying the formation origin of the S-star cluster is very challenging. Key observables of the S-stars that could assist in this are in

particular the individual ages, metallicities, and masses of the S-stars. These quantities could help distinguish if the S-stars were formed from multiple populations of stars, or a single star formation episode, which is especially useful for evaluating scenarios. Of particular interest for the Hills mechanism origin hypothesis is the age and metallicity distribution of the evolved S-stars. If the evolved S-stars are old, low-mass stars it would suggest that the main-sequence stars have a different origin. Either the Hills mechanism efficiently disrupts binaries from young structures such as the CWD or alternative scenarios are needed to explain their presence near Sgr A*, such as those explored by Levin (2007), Rantala & Naab (2024), and Akiba et al. (2024).

A deeper completeness limit of S-stars observed near Sgr A* would also help eliminate formation scenarios, since it would allow for more detailed comparisons using the Ks-band luminosity function and semi-major axis distributions in particular. This deeper completeness limit will be allowed by new instruments such as GRAVITY(+) (Eisenhauer et al. 2011; GRAVITY+ Collaboration et al. 2022a), and an increased temporal baseline for the observations.

Specifically for relating the Hills mechanism to the S-star cluster, a larger sample of confirmed HVSs originating in the GC is vital. This would allow for much lower uncertainties in the Hills mechanism rate, even for a single new HVS discovery. Additionally, measurements of the metallicities, ages, and masses of HVSs would provide strong constraints on their progenitor population. Large spectroscopic surveys such as *Gaia* DR4, *DESI* (Cooper et al. 2023), *WEAVE* (Dalton et al. 2014), and *4MOST* (de Jong et al. 2019) will provide unprecedented numbers and depth of observations and are expected to discover new HVSs (V24).

4.6   Conclusions

In this work, we investigated if binary disruptions through the Hills mechanism can explain both the observed population of S-stars and state-of-the-art HVS observations. We forward modelled the S-stars and HVSs and compared both populations to the latest observational results, taking observational selection effects into account. We find that, in particular, the ratio of main-sequence to evolved S-stars is highly informative on the star formation history of the progenitor binary population and investigated different progenitor binary populations.

We show that an old progenitor binary population, where the majority of the stellar content formed billions of years ago such as the NSC, cannot explain the ratio of main-sequence to evolved S-stars, nor the relative number of HVSs. Furthermore, our results show that a young progenitor binary population, such as the CWD, can simultaneously explain the observed pop-

ulation of main-sequence S-stars and HVS observations, but cannot explain the evolved S-stars that are observed.

We argue that the main-sequence and evolved S-star populations, if formed through the Hills mechanism, have different progenitor populations. Only the model in which progenitor binaries come from both the CWD and NSC can successfully the numbers of HVSs, main-sequence S-stars, and evolved S-stars. Furthermore, we show that the Ks-band and semi-major axis distributions predicted by this model are consistent with observations. The current Hills mechanism rate from the young binary population would need to be about an order of magnitude higher than that of the old population at roughly 10^{-4} and 10^{-5} yr^{-1} respectively. About 90% of ejected stars over the past $\lesssim 10$ Myr should thus originate from the CWD. In general, we expect most young HVSs to be ejected in bursts coinciding with star formation near Sgr A*.

We additionally argue that, provided S5-HVS1 was formed through the Hills mechanism, most (if not all) S-stars were formed through Hills mechanism disruptions of binaries.

Upcoming large spectroscopic surveys are expected to improve our understanding of the rate of the Hills mechanism, and thereby provide strong constraints on the Hills mechanism's contribution to the formation of the S-star cluster. In addition, continued observations of the GC will allow more precise measurements of the properties of the S-stars, facilitating increasingly constraining tests of different formation scenarios.

Acknowledgements

Software: `NumPy` (Harris et al. 2020), `SciPy` (Virtanen et al. 2020), `Matplotlib` (Hunter 2007), `Astropy` (Astropy Collaboration et al. 2013, 2018, 2022), `isochrones` (Morton 2015), `Speedystar` (Contigiani et al. 2019; Evans et al. 2022a).

5 | Searching for stars ejected from the Galactic Centre in DESI

Abstract

Dynamical interactions between stars and the supermassive black hole Sgr A* at the Galactic Centre (GC) may result in stars being ejected into the Galactic halo. While recent fast ejections by Sgr A* have been identified in the form of hypervelocity stars (hundreds to thousands of km/s), it is also believed that the stellar halo contains slower stars, ejected over the last few billion years. In this study we used the first data release of DESI to search for these slower GC ejecta, which are expected to stand out from the stellar halo population thanks to their combined high metallicity ($[\text{Fe/H}] \gtrsim 0$) and low vertical angular momentum (L_Z), whose distribution should peak at zero. Our search did not yield a detection but allowed us to place an upper limit on the ejection rate of stars from the GC of $\sim 2.8 \times 10^{-3}$ yr^{-1} over the past ~ 5 Gyr, which is ejection model independent. This implies that our result can be used to put constraints on different ejection models, including those that invoke mergers of Sgr A* with other massive black holes in the last last few billion years.

5.1   Introduction

The stellar halo is a highly complex environment made up of substructures containing information on the assembly history of the Galaxy. In particular, with the advent of *Gaia* (Gaia Collaboration et al. 2016), a wealth of knowledge has been gathered on the individual components of the halo (for a recent review, see Deason & Belokurov 2024). We now understand that the halo is perhaps entirely made up of the debris of past merger events. This means that the halo provides a means of studying the remains of high-redshift dwarf galaxies in great detail given the relative proximity of halo stars.

In our understanding of the halo, we rarely consider the effect of the Galactic Centre (GC) and in particular the supermassive black hole Sagittarius A* (Sgr A*). However, there is reason to think that this might not be entirely justified. In recent decades, it has become clear that Sgr A* ejects stars on unbound trajectories travelling through the halo via the Hills mechanism (Hills 1988; Yu & Tremaine 2003; Brown et al. 2005; Brown 2015; Koposov et al. 2020). Moreover, a significant fraction of these ejected stars are expected to remain bound to the Galaxy (e.g. Bromley et al. 2006; Rossi et al. 2014). This means that there would be a population of stars with the chemical properties of the GC that populate the halo (Brown et al. 2007). Stars ejected from the GC are often referred to as hypervelocity stars, but here we refer to them as GC ejecta since we focus on the stars still bound to the Galaxy, rather than the unbound population.

Galactic Centre ejecta provide interesting science cases since they are windows into the complex GC environment in parts of the sky that are easier to study, while unlocking wavelength ranges for observations that are inaccessible for sources at the GC. In addition, their trajectories can be used as tracers of the Galactic potential (e.g. Contigiani et al. 2019; Gallo et al. 2022; Armstrong et al. 2025). However, their application has been limited so far because only a single unambiguous star ejected from the GC is known (Koposov et al. 2020), alongside a handful of promising candidates (Brown et al. 2005, 2014, 2018). A big factor in the difficulty to accurately identify GC ejecta is that many of these candidates tend to be distant, which makes their past trajectories uncertain. The rate and properties of GC ejecta are therefore also still uncertain (Zhang et al. 2013; Brown et al. 2014; Marchetti et al. 2022; Evans et al. 2022a,c). The best constraints on the rate so far have been published in Verberne et al. (2024a, 2025), who find an upper limit of 10^{-5} yr^{-1} for stars more massive than 1 M$_\odot$ and a higher ejection rate over the past 10 Myr of about 10^{-4} yr^{-1}.

Additionally, past mergers of Sgr A* with intermediate-mass or supermassive black holes would have boosted the rate of stars ejected from the GC by orders of magnitude (e.g. Yu & Tremaine 2003; Gualandris et al.

2005; Baumgardt et al. 2006; Levin 2006; Evans et al. 2023). Furthermore, kicks from stellar-mass black holes and the disruption of infalling dwarf galaxies might have also produced GC ejecta (O'Leary & Loeb 2008; Abadi et al. 2009). For these reasons, it is unknown how large the halo population of GC ejecta is.

In this work we aim to constrain the population of GC ejecta that has built up in the stellar halo over the lifetime of the Galaxy. We expect that two characteristic features can be used to identify these stars: they should be metal-rich since they originated in the GC (e.g. Schultheis et al. 2019; Schödel et al. 2020), and their initial orbital angular momentum should be zero since they are ejected radially from the GC.

No other stellar halo populations with metallicities similar to that of the GC are known, making metallicity a crucial factor in the identification of this population of stars. We used the recently released Dark Energy Survey Instrument (DESI; DESI Collaboration et al. 2016a) Data Release 1 (DR1; DESI Collaboration et al. 2025b) in combination with *Gaia* DR3 (Gaia Collaboration et al. 2023a) to calculate the angular momenta of individual stars. We used the distributions of angular momenta and [Fe/H] to search for a statistical overdensity of low-angular-momentum, high-metallicity stars.

The structure of this work is as follows: We start by discussing the origin of GC ejecta and our simulations in Sect. 5.2, which is followed by a description of the observations we used in Sect. 5.3. In Sect. 5.4 we discuss our method for identifying a possible population of GC ejecta before presenting our results in Sect. 5.5. In Sect. 5.6 we discuss our assumptions and put our results into perspective. Finally, in Sect. 5.7 we summarise our work.

5.2   Galactic Centre ejecta

In this section we start by discussing a mechanism through which stars can be ejected from the GC and end up in the stellar halo, before we describe how we implemented this in our simulations. We explain the specific ejection mechanism used in our simulations below, but we argue that our results are in fact ejection model independent (see Sect. 5.6.1).

5.2.1   Hills mechanism

The Hills mechanism (Hills 1988) acts on a stellar binary that approaches a massive black hole within the tidal radius, where the tidal force of the black hole separates the stellar binary, ejecting one star and capturing its companion (for a review, see Brown 2015). The ejected stars can gain enough energy to become unbound to the Galactic potential. For a contact binary progenitor, for instance, the ejected star can be ejected at up to

about 3500 km s^{-1} (Rossi et al. 2014). However, a significant fraction of ejected stars will remain bound to the Galaxy, travelling on highly eccentric orbits. Although the fraction of stars that remain bound to the Galaxy is a function of the binary progenitor population properties, in particular the mass ratio and semi-major axis distributions, a population of stars ejected from the GC will accumulate in the stellar halo. For a star ejected from the GC on a bound orbit in a perfect axisymmetric potential, we expect L_Z to be conserved and equal to 0, while L_X and L_Y will oscillate due to the gravity of the disc.

The number of stars in this population is highly uncertain mainly because the ejection rate is only constrained to an order of magnitude and the progenitor binary properties are poorly constrained. Most literature estimates of the ejection rate are between 10^{-5} and 10^{-3} yr^{-1} (Hills 1988; Yu & Tremaine 2003; Zhang et al. 2013; Brown et al. 2014; Marchetti et al. 2022; Evans et al. 2022c; Verberne et al. 2024a) and are based on recent (past ~ 100 Myr) ejections. Verberne et al. (2025) extend this with a lower limit of the ejection rate over about a billion years of $\sim 10^{-5}$ yr^{-1}.

5.2.2 Simulations

To predict the physical and observational footprint of the population of stars ejected from the GC, we performed a suite of simulations. Our starting point was the simulations presented in Verberne et al. (2025, their Sects. 3 and 4), which we refer to for a detailed description. To summarise, we sampled from a progenitor binary population defined by the star formation history, initial mass function, log-period distribution, and mass ratio distribution. In Verberne et al. (2025) two progenitor populations are described: a young population with an average ejection rate over the past ~ 10 Myr of approximately 10^{-4} yr^{-1} and an old population with an ejection rate of at least $\sim 10^{-5}$ yr^{-1} over potentially billion-year timescales. We focussed here on the old population but note that past bursts of star formation near Sgr A* might have boosted the ejection rate, similar to what seems to have happened recently with the formation of the clockwise disc.

We randomly selected one of the stars in the progenitor binary to be ejected regardless of its mass, as appropriate for incoming centre of mass parabolic trajectories, (which is realistic; Sari et al. 2009b; Kobayashi et al. 2012). We assumed that stars are isotropically ejected from the GC and evolved their orbits using `Agama` (Vasiliev 2019a) in the McMillan (2017) Galactic potential until the 'present time'. We considered stars ejected over the past 5 Gyr to allow significant accumulation of ejected stars in the halo. We evaluated our choice of potential by comparing it with a non-axisymmetric potential in Sect. 5.6.2.

To obtain observational parameters for our simulated stars, in partic-
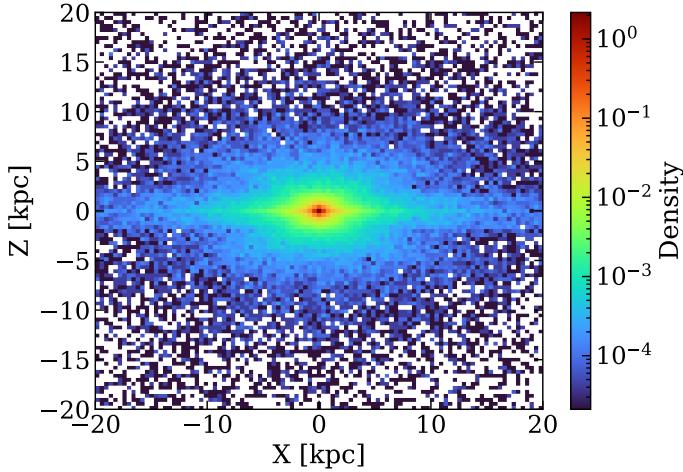
Figure 5.1: Galactic Cartesian X–Z distribution of the GC ejecta propagated in the McMillan (2017) potential over 5 Gyr.

ular the star brightness in different photometric bands, we utilised `MIST` isochrones (Dotter 2016b; Choi et al. 2016b). This allowed us to forward-model the population of ejected stars observable by DESI (DESI Collaboration et al. 2025b). Furthermore, we used `PyGaia`[1] to obtain estimated uncertainties on the measured parallax of stars in our simulations.

5.2.3   Results from simulations

Now that we have presented our simulations, we will discuss some of the key results they provide, which will help guide our search for GC ejecta in DESI DR1. We only analysed ejected stars with an apocentre above 0.1 kpc for computational reasons and because we are mainly interested in the ejected stars that reach the stellar halo. Firstly, we examined the distribution of GC ejecta in Galactocentric Cartesian coordinates (see Fig. 5.1). Although stars are isotropically ejected from the GC, we can see that gravitational focussing by the disc causes many ejected stars to end up in or near the stellar disc, which is in the $Z = 0$ plane (see also Kenyon et al. 2018). Of particular interest compared to stellar halo populations is the radial density profile, which we show in Fig. 5.2. In our definition, the fraction of stars n is calculated from the density $\rho(r)$ as $n = \int 4\pi r^2 \rho(r) \, \mathrm{d}r$. We performed a least-squares fit using a power law of the form $\rho(r) = \rho_0 r^{-\alpha}$ to the range [1, 100] kpc. We find a power-law slope of $\alpha = 3.41 \pm 0.01$, which is signifi-

---

[1]`https://github.com/agabrown/PyGaia`
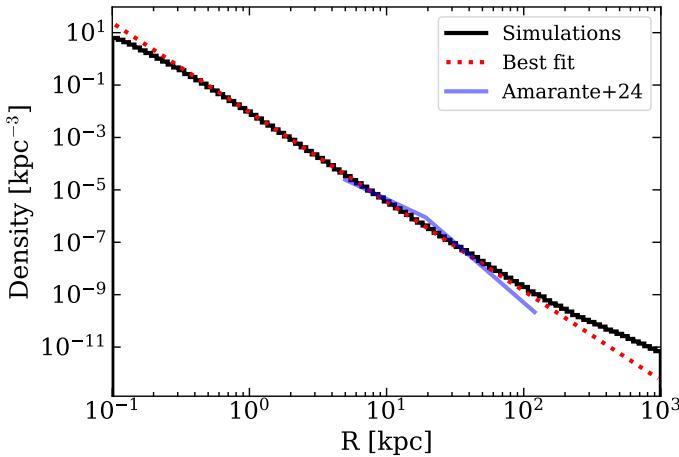
Figure 5.2: Radial density profile of the population of stars ejected from the GC in the McMillan (2017) potential. The red line shows the best fit power law in the range [1, 100] kpc, $\rho \propto r^{-3.41\pm0.01}$. For reference, we show the halo profile found in Amarante et al. (2024), which we scaled to our data at the solar circle.

cantly steeper compared to the stellar halo within ~ 20 kpc but tends to be less steep compared to measurements of the stellar halo outside ~ 20 kpc (e.g. Bell et al. 2008; Xue et al. 2015; Amarante et al. 2024; Yu et al. 2024). Moreover, at distances above about 100 kpc there is an overdensity compared to the simple power-law slope caused by unbound stars. The number and velocity distribution of unbound GC ejecta depends on the ejection model.

Another key characteristic of GC ejecta is their angular-momentum distribution. To calculate the angular momentum of a star, we needed the proper motion, radial velocity, and distance. We combined the proper motion from *Gaia* DR3 (Gaia Collaboration et al. 2023a), the RVSpecFit (RVS) radial velocity from the DESI Milky Way Survey (MWS) value-added catalogue (Koposov et al. 2024), and the distance from the DESI spectrophotometric distance value-added catalogue (Li et al. 2025). However, before we can present the expected distribution of L_Z in DESI, we first need to discuss the observational selection function. The one of the MWS bright survey in DESI is described in Cooper et al. (2023) and Koposov et al. (2025). We approximated it as follows:

- $16 < r < 19$,

- Dec > -15 deg,
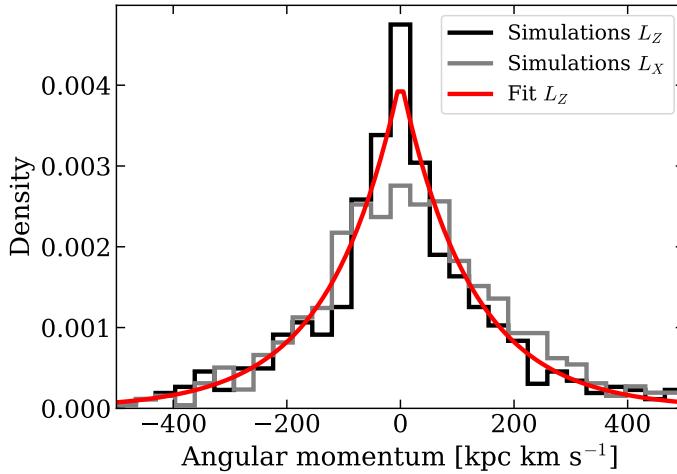
Figure 5.3: Measured distribution of L_Z and L_X for GC ejecta in the simulations. Since the true L_Z of ejected stars is 0, the observed spread is due to distance uncertainties. We provide a fit to the data in red. In grey we also plot the measured L_X distribution.

- $|b| > 40$ deg,

- completeness $= 20\%$,

- $(g - r < 0.7) \mid (g - r > 0.7 \;\wedge\; \varpi < 3\sigma_\varpi + 0.3)$,

where ϖ and σ_ϖ are the parallax and corresponding uncertainty, respectively. We know that the observed spread in the distribution of L_Z for GC ejecta in an axisymmetric potential will be dominated by observational uncertainty in the distance to individual stars, since $L_Z = 0$ intrinsically. To account for this, we randomly sampled from the distribution of distance/uncertainty for sources in the Li et al. (2025) value-added distance catalogue, for each simulated star. We convolved the true distances to the simulated stars with these uncertainties, only considering sources with distance/uncertainty > 5, since these will be the sources we analyse in Sect. 5.3. From our procedure, we obtained the expected observed angular-momentum distribution for GC ejecta in DESI, shown in Fig. 5.3. While the distribution in L_Z is exclusively set by the distance uncertainty, the spread in L_X is a combination of the intrinsic L_X distribution and the distance uncertainty. We performed an unbinned fit with a Laplace distribution of the form

$$f(x \mid \mu, b) = \frac{1}{2b} \exp\left(-\frac{|x - \mu|}{b}\right) \qquad (5.1)$$

to the L_Z distribution shown in Fig. 5.3, where we fix the offset $\mu = 0$. This provides us with the scale parameter for which we find the 50th percentile value $b = 1.2 \times 10^2$ kpc km s^{-1} (narrower than other populations; see Sect. 5.4). We used this distribution and shape as the characteristic shape of GC ejecta in the DESI data, and evaluate the (in)dependence of the assumed ejection model on our results in Sect. 5.6.1.

Given our simulations and the selection function of DESI, we can also compute the number of GC ejecta in the DESI MWS bright catalogue given an ejection rate. For GC ejecta from the old population we referenced above, we expect $\sim 0.18 * \frac{\eta}{10^{-5} \mathrm{yr}^{-1}}$ GC ejecta stars in DESI, with η the average ejection rate over the past 5 Gyr. Furthermore, we expect that recent ejections from the young population will contribute $\sim 0.27 * \frac{\eta}{10^{-4} \mathrm{yr}^{-1}}$ GC ejecta for an ejection rate of 10^{-4} yr^{-1} during the past 10 Myr to DESI. These estimates are likely somewhat pessimistic, since the bright programme includes some sources outside the selection function estimate that we used. Moreover, past mergers between Sgr A* and massive black holes could have additionally ejected large numbers of stars from the GC through the 'slingshot mechanism' that are not accounted for in our simulations, as mentioned in the introduction.

5.3   Observations

We have discussed the observational selection function and forward modelling of DESI. Here we describe the specific observational data from DESI used in this study.

We relied on data from DESI DR1 and in particular on the MWS catalogue (Cooper et al. 2023). The MWS catalogue contains spectra and properties derived from ~ 5M stars. We used the radial velocities and [Fe/H] abundance measurements from the RVS pipeline (Koposov 2019a; Cooper et al. 2023) in combination with the distances from the spectrophotometric MWS SpecDis catalogue (Li et al. 2025). For the [Fe/H] measurements, we used the calibration described in Koposov et al. (2025), which makes the measurements accurate to ~ 0.1 dex. Note that this recalibration is very important for our results, since high metallicity giants in DESI show a systematic offset towards too high [Fe/H] in RVS[2].

To ensure that we only used reliable measurements, we applied the following data quality cuts in DESI:

- `SUCCESS` = True (RVS flag),

- `RVS_WARN` = 0,

---

[2] Without the metallicity recalibration from Koposov et al. (2025), we find an overdensity of high-metallicity giants centred at $L_Z = 0$. The recalibration shifts these stars to lower metallicities, consistent with the GSE halo population.

- FEH_ERR < 0.2,

- RR_SPECTYPE $=$ STAR,

- DIST/DISTERR > 5.

Finally, we used the crossmatch of DESI and *Gaia* from the MWS value-added catalogue, to obtain proper motion measurements.

5.4   Different populations in the L_Z–[Fe/H] plane

Now that we have discussed our simulations and the observational dataset, we will describe how we analysed the data from DESI to find this population of stars ejected from the GC. We also discuss stellar populations that might be misclassified as GC ejecta because of possible overlap in parameter space.

In the introduction we identified two characteristic properties that we expect for stars ejected from the GC over the past several billion years: we expect them to be metal rich and travel on initially radial trajectories. We discuss in Sect. 5.2.2 how these stars oscillate in L_X and L_Y, while L_Z remains constant at 0. We therefore attempted to find a statistical overdensity of high-metallicity stars with an angular-momentum distribution centred at $L_Z = 0$. Furthermore, we know the shape of the angular-momentum distribution in L_Z, since we expect this shape to be dominated by the observational uncertainties in the distances to individual stars, as discussed in Sect. 5.2.2.

In order to identify an overdensity in [Fe/H]–L_Z space, it is important to understand what other populations of stars overlap in this parameter space with GC ejecta. To gain intuition for this parameter space, we show metallicity against L_Z for all stars in our observational dataset (see Sect. 5.3) in Fig. 5.4. At low metallicity ($-2 <$ [Fe/H] < -1) and centred at $L_Z = 0$ we can see the population known as Gaia-Sausage-Enceladus (GSE; Belokurov et al. 2018; Helmi et al. 2018; Haywood et al. 2018). GSE dominates the stellar halo in the solar neighbourhood and consists of metal-poor stars (mean [Fe/H] ~ -1.2; Feuillet et al. 2021) on highly eccentric orbits (for a review, see Deason & Belokurov 2024). This population of stars is believed to be the remains of the most recent major merger of the Milky Way, which occurred ~ 10 Gyr ago. The L_Z distribution of GSE peaks roughly at $L_Z = 0$, the same as the GC ejecta. However, unlike our expectation for GC ejecta, GSE does have a measurable width in L_Z space by DESI.

Metallicity is the other important differentiator: the metal-rich boundary of the GSE is typically found at around [Fe/H] ~ -0.6 (e.g. Myeong et al. 2019; Hasselquist et al. 2021; Horta et al. 2023), much lower in metallicity compared to typical stars expected from the GC (Do et al. 2015; Feldmeier-Krause et al. 2017a; Schödel et al. 2020; Feldmeier-Krause et al.
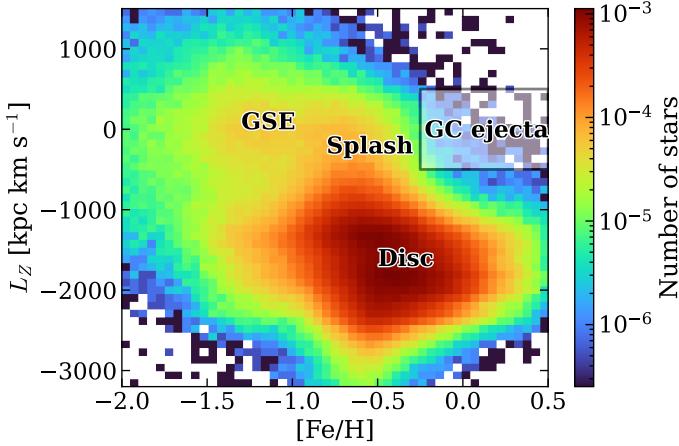
Figure 5.4: Histogram of [Fe/H] against L_Z for stars in the DESI DR1 selection described in Sect. 5.3. We highlight areas occupied by different galactic components and GC ejecta.

2025). However, it is not excluded that there is a more metal-rich compact GSE remnant (e.g. Hasselquist et al. 2021). Finally, uncertainties in the measured [Fe/H] might lead to the misclassification of a GSE star.

The other population of stars that can end up in the high-metallicity, low-angular-momentum part of parameter space comes from the (thick) disc. In Fig. 5.4 we can see that most stars reside at large negative L_Z, but a tail towards low angular momenta extends from this population. This tail is possibly related to the recently discovered population of relatively metal-rich ([Fe/H] > −0.7) stars on highly eccentric orbits known as the 'Splash' (Belokurov et al. 2020). These stars have little to no angular momentum and their relatively high metallicity means that they might overlap significantly with GC ejecta. A suggested interpretation of the Splash is that these stars were part of the proto-disc of the Galaxy and were kicked out during the merger event that created GSE. For detailed information on these and other halo populations, we refer to recent reviews (e.g. Rix & Bovy 2013; Bland-Hawthorn & Gerhard 2016; Deason & Belokurov 2024; Hunt & Vasiliev 2025).

Now that we have a better understanding of the background stellar populations at high metallicity and low angular momentum, we will describe our model. We performed all our fits in L_Z space in bins of metallicity. Furthermore, we separately treated dwarfs and giants, which we separated at $\log g = 4$ as determined in the DESI MWS catalogue from RVS. We focussed on the region $|L_Z| < 500$ kpc km s^{-1}, since the population ex-

tending from large negative L_Z can be described by a linear slope in this region. We refer to this population as the Splash. We described the GSE using a Gaussian centred at $L_Z = 0$ and determined its width in L_Z by performing an unbinned fit of this Gaussian and a linear slope to the distribution of L_Z. We limited this fit to the range $|L_Z| < 500$ kpc km s^{-1} and $-1.7 < $ [Fe/H] $ < -1.3$ since this parameter range is dominated by GSE stars in the DESI data. We find a standard deviation of ~ 170 kpc km s^{-1} for giants and ~ 410 kpc km s^{-1} for dwarfs and use these values to fix the width in L_Z of GSE for our fits in metallicity bins. We left the slope of the Splash population as a free parameter.

Our probability density function to describe the distribution of stars in L_Z within the range $[-L_{Z,\ \max}, L_{Z,\ \max}]$ per bin of metallicity is then

$$
\begin{aligned}
f(L_Z)\ =\ & p_{\mathrm{GSE}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{L_Z^2}{2\sigma^2}\right) \frac{1}{Z_{\mathrm{norm,\ GSE}}} \\
& + p_{\mathrm{ej}} \frac{1}{2b} \exp\left(\frac{-|L_Z|}{b}\right) \frac{1}{Z_{\mathrm{norm,\ ej}}} \\
& + (1 - p_{\mathrm{GSE}} - p_{\mathrm{ej}})\left(\frac{1}{2L_{Z,\ \max}} + aL_Z\right),
\end{aligned}
\tag{5.2}
$$

where p_{GSE} and p_{ej} are the fractions of stars belonging to in the GSE and the GC ejecta, respectively, σ is the width of the GSE in L_Z, $Z_{\mathrm{norm,\ GSE}}$ and $Z_{\mathrm{norm,\ ej}}$ are the integrals of the normal and Laplace distributions, respectively, over the range $[-L_{Z,\ \max}, L_{Z,\ \max}]$, b is the scale parameter of the GC ejecta, and a is the slope of the linear Splash population. For bins in metallicity, we performed unbinned fits to the probability density function for $L_{Z,\ \max} = 500$ kpc km s^{-1} using a Markov chain Monte Carlo approach. We used 100 walkers, with 500 burn-in steps and 1000 additional steps to explore the posterior and calculate percentile confidence intervals. The free parameters in this fit are p_{GSE}, p_{ej}, and a. We used uniform priors on p_{GSE} and p_{ej} between 0 and 1, and for a we used a uniform prior for $|a| \leq 1/(2L_{Z,\ \max})$, so that the posterior is positive for $|L_Z| < L_{Z,\ \max}$. Finally, we required that $1 - p_{\mathrm{GSE}} - p_{\mathrm{ej}} \geq 0$ so that our probability density function integrates to 1.

5.5   Results

Now that we have discussed our model and the data, we will look at our results. In Fig. 5.5 we show the fraction and number of stars that belong to the Splash, GSE, and GC ejecta populations, as a function of metallicity. The fraction of GC ejecta is always consistent with 0 if we consider that $p_{\mathrm{ej}} \geq 0$ and the fact that the posteriors are highly non-Gaussian given this truncation. In fact, the posterior on p_{ej} usually peaks at $p_{\mathrm{ej}} = 0$. For
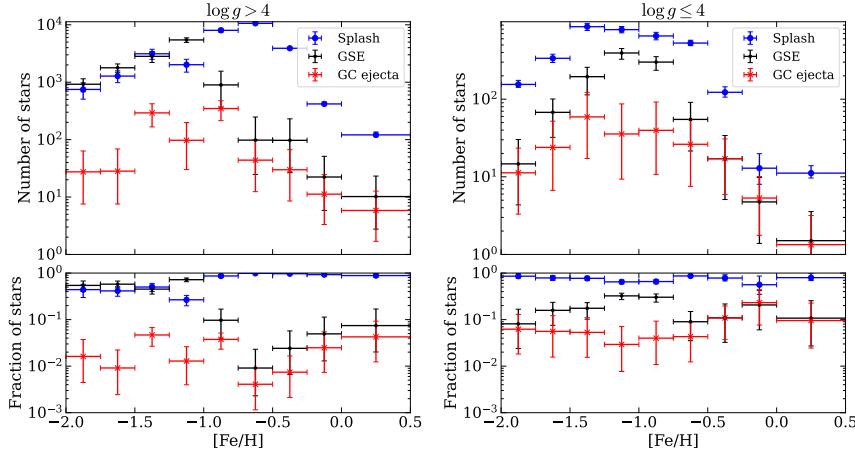
**Figure 5.5:** Relative contributions of Splash, GSE, and GC ejecta for stars with $|L_Z| < 500$ kpc km s^{-1} as a function of metallicity. The rows show the number of stars and fraction in each population, while the columns show dwarfs and giants. The horizontal error bars correspond to the bin size in [Fe/H], and the vertical error bars are from the 16th and 84th percentiles of the posterior.

dwarfs we can see that at low metallicity the Splash and GSE contribute about equally to the stellar population, while for giants the Splash always dominates. Note that these numbers are not necessarily representative of the 'true' fractions in the halo, since we did not correct for observational selection effects nor is our model optimised for detecting populations other than GC ejecta.

So far, we have not made any selections on angular momentum other than in L_Z. However, for GC ejecta, we expect that L_X is also centred and concentrated around $L_X = 0$. To improve our statistics, we performed an additional fit selecting only $L_X < 300$ kpc km s^{-1} to constrain the contribution from the GC ejecta population to the metallicity bin $-0.25 <$ [Fe/H] < 0.5. We chose this metallicity bin to encompass all GC ejecta. For both dwarfs and giants we again find a fraction of GC ejecta consistent with 0. The 95% upper limit on the number of GC ejecta within this metallicity bin is about 38 dwarfs and 13 giants, for a total of about 51 stars. However, not all GC ejecta will be in our observational sample. The 95% upper limit on the number of GC ejecta with $16 < r < 19$ and $|b| > 40$ deg can be written as

$$N_{\mathrm{GC,ej}}^{95}(16 < r < 19 \ \wedge \ |b| > 40 \ \mathrm{deg}) = \frac{p_{\mathrm{ej}}^{95} N_{\mathrm{[Fe/H]}}}{C_{\mathrm{spec}} C_{\mathrm{colour}} C_{\mathrm{Dec}} C_{\mathrm{L}}}, \qquad (5.3)$$

with p_{ej}^{95} the 95% upper limit on the fraction of GC ejecta, $N_{\mathrm{[Fe/H]}}$ the

number of stars in the metallicity bin, and C_{spec}, C_{colour}, C_{Dec}, and C_{L} the completenesses on the spectroscopy, colour selection, Dec range, and angular momentum, respectively. The spectroscopic completeness of DESI DR1 is 20%, as mentioned before, and from our simulations we find that the product of the other completeness factors is about 60%. The total number of GC ejected stars with $16 < r < 19$ at $|b| > 40$ can therefore be no higher than $51/(0.2 \cdot 0.6) \simeq 4.3 \times 10^2$ at the 95% confidence limit, assuming that the ejection mechanism is isotropic.

The upper limit for the number of GC ejecta in DESI implies constraints on the ejection rate of stars from the GC over a timescale of billions of years, since GC ejecta will have accumulated in the halo for billions of years. Since we expect $\sim 0.18 * \frac{\eta}{10^{-5}\mathrm{yr}^{-1}}$ GC ejecta in DESI for an ejection rate η (see Sect. 5.2.3), the average ejection rate over this timescale has to be at most $\sim 2.8 \times 10^{-3}$ yr^{-1}.

5.6   Discussion

In this section we provide context for our results, review some of our assumptions, and discuss future prospects. Since our simulations (excluding orbit integrations) are directly taken from the work of Verberne et al. (2025), we refer to that paper for a discussion of the ejection model and the progenitor binary population assumptions.

In this work we investigated whether we could identify an overdensity of high-metallicity stars at low angular momenta in the Galactic halo using data from DESI DR1. Such a population could point to Hills mechanism disruptions of binary stars near Sgr A* or past mergers of Sgr A* with intermediate-mass or massive black holes. Our analysis yields a non-detection that we used to constrain the GC ejection rate over a timescale of ~ 5 Gyr, a novelty with respect to previous constraints obtained with unbound stars that were therefore limited to the shorter timescales of their fly times (~ 100 Myr). Specifically, we conclude that the average ejection rate over the past ~ 5 Gyr cannot exceed $\sim 2.8 \times 10^{-3}$ yr^{-1}. This rate is in line with previous estimates and constraints for the Hills mechanism, which tend to span the range $10^{-5} - 10^{-3}$ yr^{-1} (Hills 1988; Yu & Tremaine 2003; Bromley et al. 2012; Zhang et al. 2013; Brown et al. 2014; Evans et al. 2022a,c; Marchetti et al. 2022; Verberne et al. 2024a, 2025).

5.6.1   Testing the ejection model dependence

Here we evaluate the dependence of our results on the specific ejection model assumed. A previous study showed that the velocity distribution of stars ejected from the GC is determined by the potential rather than the ejection velocity spectrum for stars travelling on bound orbits (Rossi et al. 2014).

We verified that this holds for our simulations by using different ejection velocity distributions. We find that the width of the zero-angular-momentum population and the number of expected GC ejecta in DESI are not significantly affected, which means that our results hold for any population of stars ejected isotropically from the GC and are thus ejection model independent. Anisotropic ejections could potentially bias our results if, for instance, stars are preferentially ejected in the Galactic plane, since DESI targets high latitude sources. Additionally, anisotropic ejections could change the angular-momentum distribution of GC ejecta, which would influence our results presented in Fig. 5.5.

5.6.2   Testing the (non-)axisymmetric potential dependence

So far, we have used the axisymmetric McMillan (2017) potential for our orbit integration. The effect is that the intrinsic L_Z distribution is a delta function at $L_Z = 0$ and the spread in the observed L_Z distribution is exclusively caused by observational uncertainties. However, the potential of the Galaxy contains non-axisymmetric components that will impart L_Z onto GC ejecta. To investigate this effect, we made use of the potential defined in Hunter et al. (2024, excluding Sgr A*), which contains the nuclear star cluster, nuclear stellar disc, bar, thin and thick stellar discs, gas discs, dark halo, and spiral arms. Both the bar and spiral arms are non-axisymmetric and are rotating in this potential. We followed the same procedure we used for the McMillan (2017) potential and integrate the orbits using `Agama`. The result of the non-axisymmetric components is that the L_Z distribution becomes wider, with a scale parameter of $\sim 1.5 \times 10^2$ kpc km s^{-1}, compared to $\sim 1.2 \times 10^2$ kpc km s^{-1} for the McMillan (2017) potential. We also evaluated if the increased width in L_Z affects our conclusions, by reanalysing the DESI data using our model (Eq. 5.2) with this increased width of the GC ejecta population. We find that the increased width of the L_Z distribution does not significantly impact the ratios of stars we assigned to the Splash, GSE, or GC ejecta populations. Our rate and number constraints for GC ejecta increase by about 15% if we assume the Hunter et al. (2024) potential over the McMillan (2017) one. We therefore conclude that our conclusions are not significantly affected by the assumption that the potential is axisymmetric.

Separate from the adopted potential, a factor that might introduce additional angular momentum is scattering from individual stars. We implicitly assumed in this work that this is negligible, but especially for stars that experience multiple pericentre passages through the dense GC region, scattering might become important. Since we only considered stars with apocentres outside the GC, we effectively limited the number of orbits for

these stars. Indeed, we find that for stars with an apocentre greater than 1 kpc, the number of orbits is at most a few tens in our simulations. We therefore assumed that scattering by individual stars is not important.

5.7   Summary and conclusion

In this work we searched for the population of GC ejecta that we expect has accumulated in the stellar halo over billions of years via the Hills mechanism and massive black hole binary ejections. To this purpose, we used the newly released DESI DR1 in combination with simulations to predict the angular-momentum distribution of stars ejected from the GC. We find that the radial distribution of GC ejecta between 1 and 100 kpc from the GC can be described by a single power law with $\rho \propto r^{-3.41 \pm 0.01}$. Gravitational focussing by the disc means that many stars ejected from the GC will have ended up in the stellar disc, making identification more challenging due to the large number of disc stars. We built a mixture model to describe the L_Z distribution around $L_Z = 0$ as a function of metallicity and fitted this to the DESI data. We find that on a statistical level there is no overdensity at high metallicities of stars on low L_Z orbits, as would be expected for GC ejecta. Based on this null-detection, we place an ejection-model-independent upper limit on the ejection rate of stars from the GC of $\sim 2.8 \times 10^{-3}$ yr^{-1}. We also place a 95% upper limit on the number of GC ejecta at $|b| > 40$ deg and $16 < r < 19$ of 4.3×10^2 stars.

The analysis presented here can be repeated for other large spectroscopic surveys, for instance using the *Gaia* DR3 radial velocity subsample (Gaia Collaboration et al. 2023a), LAMOST (Cui et al. 2012), or APOGEE (Majewski et al. 2017b). Furthermore, the upcoming *Gaia* DR4 and DESI DR2 will significantly increase the number of sources to which this analysis can be applied. DESI DR2, for instance, is expected to contain about three times more stars than DR1 (Koposov et al. 2025). Our method would be enhanced by considering additional tracers, such as elemental abundances, which would add weight to any claimed GC ejecta discoveries (see e.g. Hattori et al. 2025).

Our results can be used to constrain ejection mechanisms; a particularly interesting application is constraining the merger history of Sgr A*. Gravitational slingshots of single stars by a massive binary black hole eject stars with properties similar to the Hills mechanism (Yu & Tremaine 2003). Evans et al. (2023) already used this in combination with the lack of uncontroversial GC ejecta in *Gaia* DR3 to determine that mergers of Sgr A* with a possible companion of > 500 M$_\odot$ cannot have happened within the past 10 Myr. Expanding on this, we believe that our non-detection of a population of bound ejecta constrains the merger history of Sgr A* to a timescale of billions of years, something we know very little about.

Acknowledgements

Software: `NumPy` (Harris et al. 2020), `SciPy` (Virtanen et al. 2020), `Matplotlib` (Hunter 2007), `Astropy` (Astropy Collaboration et al. 2013,

2018, 2022), `isochrones` (Morton 2015), `Speedystar` (Contigiani et al. 2019; Evans et al. 2022a), `emcee` (Foreman-Mackey et al. 2013), `Agama` (Vasiliev 2019b).

Bibliography

Abadi M. G., Navarro J. F., Steinmetz M., 2009, ApJ, 691, L63

Abdurro'uf et al., 2022, ApJS, 259, 35

Abolfathi B., et al., 2018, ApJS, 235, 42

Aihara H., et al., 2011, ApJS, 193, 29

Akiba T., Naoz S., Madigan A.-M., 2024, arXiv e-prints, p. arXiv:2410.19881

Alexander T., 2017, ARA&A, 55, 17

Amarante J. A. S., Koposov S. E., Laporte C. F. P., 2024, A&A, 690, A166

Andrae R., Rix H.-W., Chandra V., 2023a, The Astrophysical Journal Supplement Series, 267, 8

Andrae R., et al., 2023b, A&A, 674, A27

Antonini F., Merritt D., 2013, ApJ, 763, L10

Armstrong I., Evans F. A., Bovy J., 2025, ApJ, 984, 56

Ashkenazy Y., Balberg S., 2025, A&A, 695, A98

Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, ARA&A, 47, 481

Astropy Collaboration et al., 2013, A&A, 558, A33

Astropy Collaboration et al., 2018, AJ, 156, 123

Astropy Collaboration et al., 2022, apj, 935, 167

Baganoff F. K., et al., 2001, Nature, 413, 45

Bahcall J. N., Wolf R. A., 1976, ApJ, 209, 214

Bailer-Jones C. A. L., 2015, PASP, 127, 994

Bailer-Jones C. A. L., et al., 2013, A&A, 559, A74

Bailer-Jones C. A. L., Rybizki J., Fouesneau M., Mantelet G., Andrae R., 2018, AJ, 156, 58

Bailer-Jones C. A. L., Rybizki J., Fouesneau M., Demleitner M., Andrae R., 2021, AJ, 161, 147

Balick B., Brown R. L., 1974, ApJ, 194, 265

Bar-Or B., Alexander T., 2016, ApJ, 820, 129

Barbuy B., Chiappini C., Gerhard O., 2018, ARA&A, 56, 223

Bartko H., et al., 2009, ApJ, 697, 1741

Bartko H., et al., 2010, ApJ, 708, 834

Baumgardt H., Gualandris A., Portegies Zwart S., 2006, MNRAS, 372, 174

Bell E. F., et al., 2008, ApJ, 680, 295

Belokurov V., Erkal D., Evans N. W., Koposov S. E., Deason A. J., 2018, MNRAS, 478, 611

Belokurov V., Sanders J. L., Fattahi A., Smith M. C., Deason A. J., Evans N. W., Grand R. J. J., 2020, MNRAS, 494, 3880

Bennett M., Bovy J., 2019, MNRAS, 482, 1417

Benz W., Hills J. G., 1987, ApJ, 323, 614

Binney J., Tremaine S., 2008, Galactic Dynamics: Second Edition. Princeton University Press

Binney J., Gerhard O., Spergel D., 1997, MNRAS, 288, 365

Blaauw A., 1961, BAIN, 15, 265

Bland-Hawthorn J., Gerhard O., 2016, ARA&A, 54, 529

Boehle A., et al., 2016, ApJ, 830, 17

Bonaca A., Price-Whelan A. M., 2025, New Astronomy Reviews, 100, 101713

Boubert D., Guillochon J., Hawkins K., Ginsburg I., Evans N. W., Strader J., 2018, MNRAS, 479, 2789

Boubert D., Erkal D., Gualandris A., 2020, MNRAS, 497, 2930

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., Geller M. J., Brown W. R., 2012, ApJ, 749, L42

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., Kurtz M. J., 2006a, ApJ, 640, L35

Brown W. R., Geller M. J., Kenyon S. J., Kurtz M. J., 2006b, ApJ, 647, 303

Brown W. R., Geller M. J., Kenyon S. J., Kurtz M. J., Bromley B. C., 2007, ApJ, 660, 311

Brown W. R., Geller M. J., Kenyon S. J., 2009, ApJ, 690, 1639

Brown W. R., Geller M. J., Kenyon S. J., 2012, ApJ, 751, 55

Brown W. R., Geller M. J., Kenyon S. J., 2014, ApJ, 787, 89

Brown W. R., Lattanzi M. G., Kenyon S. J., Geller M. J., 2018, ApJ, 866, 39

Burkert A., Gillessen S., Lin D. N. C., Zheng X., Schoeller P., Eisenhauer F., Genzel R., 2024, ApJ, 962, 81

Cantat-Gaudin T., et al., 2023, A&A, 669, A55

Carnall A. C., 2017, arXiv e-prints, p. arXiv:1705.05165

Castro-Ginard A., et al., 2021, A&A, 652, A162

Castro-Ginard A., et al., 2023, A&A, 677, A37

Cavieres M., et al., 2025, ApJ, 983, 83

Chakrabarti S., et al., 2023, AJ, 166, 6

Chandra V., et al., 2024, ApJ, 972, 112

Chen Z., et al., 2023, ApJ, 944, 79

Choi J., Dotter A., Conroy C., Cantiello M., Paxton B., Johnson B. D., 2016a, ApJ, 823, 102

Choi J., Dotter A., Conroy C., Cantiello M., Paxton B., Johnson B. D., 2016b, ApJ, 823, 102

Conroy C., Naidu R. P., Garavito-Camargo N., Besla G., Zaritsky D., Bonaca A., Johnson B. D., 2021, Nature, 592, 534

Contigiani O., Rossi E. M., Marchetti T., 2019, MNRAS, 487, 4025

Cooper A. P., et al., 2023, ApJ, 947, 37

Craig M., et al., 2022, astropy/ccdproc: 2.3.1 – fixes astropy 5.1 compatibility, doi:10.5281/zenodo.6533213, `https://doi.org/10.5281/zenodo.6533213`

Cropper M., et al., 2018, A&A, 616, A5

Cui X.-Q., et al., 2012, Research in Astronomy and Astrophysics, 12, 1197

DESI Collaboration et al., 2016a, arXiv e-prints, p. arXiv:1611.00036

DESI Collaboration et al., 2016b, arXiv e-prints, p. arXiv:1611.00037

DESI Collaboration et al., 2025a, arXiv e-prints, p. arXiv:2503.14745

DESI Collaboration et al., 2025b, arXiv e-prints, p. arXiv:2503.14745

Dalton G., et al., 2014, in Ramsay S. K., McLean I. S., Takami H., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 9147, Ground-based and Airborne Instrumentation for Astronomy V. p. 91470L (`arXiv:1412.0843`), doi:10.1117/12.2055132

Dark Energy Survey Collaboration et al., 2016, MNRAS, 460, 1270

De Angeli F., et al., 2023, A&A, 674, A2

De Silva G. M., et al., 2015, MNRAS, 449, 2604

Deason A. J., Belokurov V., 2024, New Astronomy Reviews, 99, 101706

Degenaar N., Wijnands R., 2009, A&A, 495, 547

Do T., et al., 2013, ApJ, 779, L6

Do T., Kerzendorf W., Winsor N., Støstad M., Morris M. R., Lu J. R., Ghez A. M., 2015, ApJ, 809, 143

Do T., et al., 2019, Science, 365, 664

Dotter A., 2016a, ApJS, 222, 8

Dotter A., 2016b, ApJS, 222, 8

Drimmel R., Poggio E., 2018, Research Notes of the AAS, 2, 210

Dunstall P. R., et al., 2015, A&A, 580, A93

Dwek E., et al., 1995, ApJ, 445, 716

Eckart A., Genzel R., 1996, Nature, 383, 415

Edelmann H., Napiwotzki R., Heber U., Christlieb N., Reimers D., 2005, ApJ, 634, L181

Eisenhauer F., et al., 2005, The Astrophysical Journal, 628, 246

Eisenhauer F., et al., 2011, The Messenger, 143, 16

El-Badry K., et al., 2023, MNRAS, 518, 1057

Erkal D., et al., 2019, MNRAS, 487, 2685

Erkal D., et al., 2021, MNRAS, 506, 2677

Evans J., Friedlander M. W., 2025, Encyclopedia Britannica

Evans K. A., Massey P., 2015, AJ, 150, 149

Evans F. A., Renzo M., Rossi E. M., 2020, MNRAS, 497, 5344

Evans F. A., Marchetti T., Rossi E. M., 2022a, MNRAS, 512, 2350

Evans F. A., Marchetti T., Rossi E. M., 2022b, MNRAS, 517, 3469

Evans F. A., Marchetti T., Rossi E. M., 2022c, MNRAS, 517, 3469

Evans F. A., Rasskazov A., Remmelzwaal A., Marchetti T., Castro-Ginard A., Rossi E. M., Bovy J., 2023, MNRAS, 525, 561

Evans F. A., Grondin S. M., Ye C. S., Webb J., Bovy J., 2025, arXiv e-prints, p. arXiv:2506.14273

Event Horizon Telescope Collaboration et al., 2022a, ApJ, 930, L12

Event Horizon Telescope Collaboration et al., 2022b, ApJ, 930, L13

Event Horizon Telescope Collaboration et al., 2022c, ApJ, 930, L14

Event Horizon Telescope Collaboration et al., 2022d, ApJ, 930, L15

Event Horizon Telescope Collaboration et al., 2022e, ApJ, 930, L16

Everall A., Boubert D., Koposov S. E., Smith L., Holl B., 2021, Monthly Notices of the Royal Astronomical Society, 502, 1908

Feldmeier-Krause A., Kerzendorf W., Neumayer N., Schödel R., Nogueras-Lara F., Do T., de Zeeuw P. T., Kuntschner H., 2017a, MNRAS, 464, 194

Feldmeier-Krause A., Kerzendorf W., Neumayer N., Schödel R., Nogueras-Lara F., Do T., de Zeeuw P. T., Kuntschner H., 2017b, MNRAS, 464, 194

Feldmeier-Krause A., Veršič T., van de Ven G., Gallego-Cano E., Neumayer N., 2025, arXiv e-prints, p. arXiv:2506.06014

Feuillet D. K., Sahlholdt C. L., Feltzing S., Casagrande L., 2021, MNRAS, 508, 1489

Fitzpatrick E. L., 1999, PASP, 111, 63

Foreman-Mackey D., 2016, The Journal of Open Source Software, 1, 24

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306

Fragione G., Loeb A., 2017, New Astronomy, 55, 32

Fritz T. K., et al., 2011, ApJ, 737, 73

GRAVITY Collaboration et al., 2018, A&A, 615, L15

GRAVITY Collaboration et al., 2019, A&A, 625, L10

GRAVITY Collaboration et al., 2020, A&A, 636, L5

GRAVITY+ Collaboration et al., 2022a, The Messenger, 189, 17

GRAVITY Collaboration et al., 2022b, A&A, 657, L12

Gaia Collaboration et al., 2016, A&A, 595, A1

Gaia Collaboration et al., 2021, A&A, 649, A1

Gaia Collaboration Vallenari, A. Brown, A.G.A. Prusti, T. et al. 2022, A&A

Gaia Collaboration et al., 2023a, A&A, 674, A1

Gaia Collaboration et al., 2023b, A&A, 674, A37

Galilei G., 1610, Sidereus Nuncius

Gallego-Cano E., Fritz T., Schödel R., Feldmeier-Krause A., Do T., Nishiyama S., 2024, A&A, 689, A190

Gallo A., Ostorero L., Chakrabarty S. S., Ebagezio S., Diaferio A., 2022, A&A, 663, A72

Garavito-Camargo N., Besla G., Laporte C. F. P., Johnston K. V., Gómez F. A., Watkins L. L., 2019, ApJ, 884, 51

Generozov A., 2021, MNRAS, 501, 3088

Generozov A., Madigan A.-M., 2020, ApJ, 896, 137

Generozov A., Perets H. B., 2022, MNRAS, 513, 4257

Generozov A., et al., 2025, A&A, 696, A68

Genzel R., et al., 2003, ApJ, 594, 812

Genzel R., Eisenhauer F., Gillessen S., 2010, Reviews of Modern Physics, 82, 3121

Gezari S., 2021, ARA&A, 59, 21

Ghez A. M., Klein B. L., Morris M., Becklin E. E., 1998, ApJ, 509, 678

Ghez A. M., et al., 2003, The Astrophysical Journal, 586, L127

Ghez A. M., et al., 2008, The Astrophysical Journal, 689, 1044

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., Reid N., 1983, MNRAS, 202, 1025

Gilmore G., Wyse R. F. G., 1991, ApJ, 367, L55

Ginsburg I., Loeb A., 2006, MNRAS, 368, 221

Gnedin O. Y., Gould A., Miralda-Escudé J., Zentner A. R., 2005, ApJ, 634, 344

Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759

Gould A., Quillen A. C., 2003, The Astrophysical Journal, 592, 935

Graczyk D., et al., 2020, ApJ, 904, 13

Gravity Collaboration et al., 2024, A&A, 692, A242

Green G., 2018a, The Journal of Open Source Software, 3, 695

Green G., 2018b, The Journal of Open Source Software, 3, 695

Gualandris A., Merritt D., 2009, ApJ, 705, 361

Gualandris A., Portegies Zwart S., 2007, MNRAS, 376, L29

Gualandris A., Portegies Zwart S., Sipior M. S., 2005, MNRAS, 363, 223

Guiglion G., et al., 2023, arXiv e-prints, p. arXiv:2306.05086

Guy J., et al., 2023, AJ, 165, 144

Häberle M., et al., 2024, Nature, 631, 285

Habibi M., et al., 2017a, ApJ, 847, 120

Habibi M., et al., 2017b, ApJ, 847, 120

Hailey C. J., Mori K., Bauer F. E., Berkowitz M. E., Hong J., Hord B. J., 2018, Nature, 556, 70

Han J. J., El-Badry K., Lucchini S., Hernquist L., Brown W., Garavito-Camargo N., Conroy C., Sari R., 2025, ApJ, 982, 188

Harris C. R., et al., 2020, Nature, 585, 357

Hasselquist S., et al., 2021, The Astrophysical Journal, 923, 172

Hattori K., Valluri M., Castro N., 2018, ApJ, 869, 33

Hattori K., Taniguchi D., Tsujimoto T., Matsunaga N., Sameshima H., Elgueta S. S., Otsubo S., 2025, arXiv e-prints, p. arXiv:2502.20266

Haywood M., Di Matteo P., Lehnert M. D., Snaith O., Khoperskov S., Gómez A., 2018, ApJ, 863, 113

Helmi A., 2020, ARA&A, 58, 205

Helmi A., Babusiaux C., Koppelman H. H., Massari D., Veljanoski J., Brown A. G. A., 2018, Nature, 563, 85

Hills J. G., 1975, Nature, 254, 295

Hills J. G., 1988, Nature, 331, 687

Hills J. G., 1991, AJ, 102, 704

Hirsch H. A., Heber U., O'Toole S. J., Bresolin F., 2005, A&A, 444, L61

Hogg D. W., Bovy J., Lang D., 2010, arXiv e-prints, p. arXiv:1008.4686

Hopman C., Alexander T., 2006, ApJ, 645, 1152

Horta D., et al., 2023, MNRAS, 520, 5671

Hou L. G., Han J. L., 2014, A&A, 569, A125

Hou L. G., Han J. L., Shi W. B., 2009, A&A, 499, 473

Hunt J. A. S., Vasiliev E., 2025, New Astronomy Reviews, 100, 101721

Hunter J. D., 2007, Computing in Science & Engineering, 9, 90

Hunter G. H., et al., 2024, A&A, 692, A216

Hurley J. R., Pols O. R., Tout C. A., 2000, Monthly Notices of the Royal Astronomical Society, 315, 543

Husser T. O., Wende-von Berg S., Dreizler S., Homeier D., Reiners A., Barman T., Hauschildt P. H., 2013, A&A, 553, A6

Irrgang A., Kreuzer S., Heber U., 2018, A&A, 620, A48

Irrgang A., Dimpel M., Heber U., Raddi R., 2021, A&A, 646, L4

Jin S., et al., 2024, MNRAS, 530, 2688

Jurić M., et al., 2008, ApJ, 673, 864

Katz D., et al., 2023, A&A, 674, A5

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., 2018, ApJ, 864, 130

Klessen R. S., Spaans M., Jappsen A.-K., 2007, MNRAS, 374, L29

Kobayashi S., Hainick Y., Sari R., Rossi E. M., 2012, The Astrophysical Journal, 748, 105

Koposov S. E., 2019a, RVSpecFit: Radial velocity and stellar atmospheric parameter fitting, Astrophysics Source Code Library, record ascl:1907.013

Koposov S. E., 2019b, RVSpecFit: Radial velocity and stellar atmospheric parameter fitting, Astrophysics Source Code Library, record ascl:1907.013 (ascl:1907.013)

Koposov S. E., et al., 2020, MNRAS, 491, 2465

Koposov S. E., et al., 2023, MNRAS, 521, 4936

Koposov S. E., et al., 2024, MNRAS, 533, 1012

Koposov S. E., et al., 2025, arXiv e-prints, p. arXiv:2505.14787

Kreuzer S., Irrgang A., Heber U., 2020, A&A, 637, A53

Kroupa P., 2001, MNRAS, 322, 231

Lam S. K., Pitrou A., Seibert S., 2015, in Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC. LLVM '15. Association for Computing Machinery, New York,

NY, USA, doi:10.1145/2833157.2833162, https://doi.org/10.1145/2833157.2833162

Laureijs R., et al., 2011, arXiv e-prints, p. arXiv:1110.3193

Leonard P. J. T., 1991, AJ, 101, 562

Leonard P. J. T., Duncan M. J., 1988, AJ, 96, 222

Leonard P. J. T., Duncan M. J., 1990, AJ, 99, 608

Levin Y., 2006, ApJ, 653, 1203

Levin Y., 2007, MNRAS, 374, 515

Levin Y., Beloborodov A. M., 2003, ApJ, 590, L33

Li T. S., et al., 2019a, MNRAS, 490, 3508

Li T. S., et al., 2019b, MNRAS, 490, 3508

Li J., Wong K. W. K., Hogg D. W., Rix H.-W., Chandra V., 2023, arXiv e-prints, p. arXiv:2309.14294

Li S., et al., 2025, arXiv e-prints, p. arXiv:2503.02291

Lin C. C., Shu F. H., 1964, ApJ, 140, 646

Lindegren L., et al., 2018, A&A, 616, A2

Lindegren L., et al., 2021, A&A, 649, A4

Löckmann U., Baumgardt H., Kroupa P., 2010, MNRAS, 402, 519

Lu J. R., Ghez A. M., Hornstein S. D., Morris M. R., Becklin E. E., Matthews K., 2009, ApJ, 690, 1463

Lu J. R., Do T., Ghez A. M., Morris M. R., Yelda S., Matthews K., 2013, ApJ, 764, 155

Luri X., et al., 2018, A&A, 616, A9

Madigan A.-M., Levin Y., Hopman C., 2009, ApJ, 697, L44

Madigan A.-M., Pfuhl O., Levin Y., Gillessen S., Genzel R., Perets H. B., 2014, ApJ, 784, 23

Majewski S. R., et al., 2017a, AJ, 154, 94

Majewski S. R., et al., 2017b, AJ, 154, 94

Maness H., et al., 2007, ApJ, 669, 1024

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

Marchetti T., Evans F. A., Rossi E. M., 2022, MNRAS, 515, 767

Mastrobuono-Battisti A., Church R. P., Davies M. B., 2021, MNRAS, 505, 3314

McMillan P. J., 2017, MNRAS, 465, 76

McWilliam A., Zoccali M., 2010, ApJ, 724, 1491

Merritt D., 2013, Classical and Quantum Gravity, 30, 244005

Moe M., Di Stefano R., 2017, ApJS, 230, 15

Moe M., Stefano R. D., 2013, The Astrophysical Journal, 778, 95

Moe M., Stefano R. D., 2015, The Astrophysical Journal, 810, 61

Montegriffo P., et al., 2023, A&A, 674, A3

Morton T. D., 2015, isochrones: Stellar model grid package, Astrophysics Source Code Library, record ascl:1503.010

Myeong G. C., Vasiliev E., Iorio G., Evans N. W., Belokurov V., 2019, MNRAS, 488, 1235

Naidu R. P., et al., 2021, ApJ, 923, 92

Nandakumar G., Ryde N., Schultheis M., Thorsbro B., Jönsson H., Barklem P. S., Rich R. M., Fragkoudi F., 2018, MNRAS, 478, 4374

Naoz S., Will C. M., Ramirez-Ruiz E., Hees A., Ghez A. M., Do T., 2020, ApJ, 888, L8

Ness M., Lang D., 2016, AJ, 152, 14

Neumayer N., Seth A., Böker T., 2020, A&A Rev., 28, 4

Nogueras-Lara F., et al., 2020, Nature Astronomy, 4, 377

O'Leary R. M., Loeb A., 2008, MNRAS, 383, 86

Oh S., Kroupa P., 2016, A&A, 590, A107

Paumard T., et al., 2006, ApJ, 643, 1011

Pedregosa F., et al., 2011, Journal of Machine Learning Research, 12, 2825

Pelupessy F. I., van Elteren, A. de Vries, N. McMillan, S. L. W. Drost, N. Portegies Zwart, S. F. 2013, A&A, 557, A84

Penoyre Z., Belokurov V., Evans N. W., 2022, MNRAS, 513, 2437

Penoyre Z., Rossi E. M., Stone N. C., 2025, arXiv e-prints, p. arXiv:2505.06344

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., 2009a, ApJ, 697, 2096

Perets H. B., Gualandris A., Kupi G., Merritt D., Alexander T., 2009b, ApJ, 702, 884

Perryman M. A. C., et al., 1997, A&A, 323, L49

Petersen M. S., Peñarrubia J., 2021, Nature Astronomy, 5, 251

Pfuhl O., et al., 2011, ApJ, 741, 108

Pietrzyński G., et al., 2019, Nature, 567, 200

Poggio E., et al., 2021, A&A, 651, A104

Portegies Zwart S. F., 2000, ApJ, 544, 437

Portegies Zwart S., McMillan S., 2018, Astrophysical Recipes. 2514-3433, IOP Publishing, doi:10.1088/978-0-7503-1320-9, https://dx.doi.org/10.1088/978-0-7503-1320-9

Portegies Zwart S., et al., 2009, New Astronomy, 14, 369

Portegies Zwart S. F., McMillan S. L., van Elteren A., Pelupessy F. I., de Vries N., 2013, Computer Physics Communications, 184, 456

Poveda A., Ruiz J., Allen C., 1967, Boletin de los Observatorios Tonantzintla y Tacubaya, 4, 86

Press W. H., Schechter P., 1974, ApJ, 187, 425

Przybilla N., Nieva M. F., Heber U., Firnstein M., Butler K., Napiwotzki R., Edelmann H., 2008a, A&A, 480, L37

Przybilla N., Fernanda Nieva M., Heber U., Butler K., 2008b, ApJ, 684, L103

Rantala A., Naab T., 2024, MNRAS, 527, 11458

Rauch K. P., Tremaine S., 1996, New Astronomy, 1, 149

Rees M. J., 1988, Nature, 333, 523

Reino S., Rossi E. M., Sanderson R. E., Sellentin E., Helmi A., Koppelman H. H., Sharma S., 2021, MNRAS, 502, 4170

Renzo M., et al., 2019, A&A, 624, A66

Rix H.-W., Bovy J., 2013, A&A Rev., 21, 61

Rix H.-W., et al., 2022, The Astrophysical Journal, 941, 45

Rose S. C., Mockler B., 2024, arXiv e-prints, p. arXiv:2412.00975

Rose S. C., Naoz S., Sari R., Linial I., 2023, ApJ, 955, 30

Rossi E. M., Kobayashi S., Sari R., 2014, The Astrophysical Journal, 795, 125

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

Rossi E. M., Stone N. C., Law-Smith J. A. P., Macleod M., Lodato G., Dai J. L., Mandel I., 2021, Space Sci. Rev., 217, 40

Ryan S. G., Norton A. J., 2010, Stellar Evolution and Nucleosynthesis. Cambridge University Press, Cambridge, UK

Sana H., et al., 2012, Science, 337, 444

Sana H., et al., 2013, A&A, 550, A107

Santana F. A., et al., 2021, AJ, 162, 303

Sari R., Fragione G., 2019, ApJ, 885, 24

Sari R., Kobayashi S., Rossi E. M., 2009a, The Astrophysical Journal, 708, 605

Sari R., Kobayashi S., Rossi E. M., 2009b, The Astrophysical Journal, 708, 605

Schlafly E. F., Finkbeiner D. P., 2011, ApJ, 737, 103

Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525

Schödel R., et al., 2007, A&A, 469, 125

Schödel R., Merritt D., Eckart A., 2009, A&A, 502, 91

Schödel R., Najarro F., Muzic K., Eckart A., 2010, A&A, 511, A18

Schödel R., Feldmeier A., Neumayer N., Meyer L., Yelda S., 2014, Classical and Quantum Gravity, 31, 244007

Schödel R., Nogueras-Lara F., Gallego-Cano E., Shahzamanian B., Gallego-Calvente A. T., Gardini A., 2020, A&A, 641, A102

Schödel R., et al., 2023, arXiv e-prints, p. arXiv:2310.11912

Schultheis M., Rich R. M., Origlia L., Ryde N., Nandakumar G., Thorsbro B., Neumayer N., 2019, A&A, 627, A152

Sersante B., Penoyre Z., Rossi E. M., 2025, arXiv e-prints, p. arXiv:2505.08499

Shen J., Zheng X.-W., 2020, Research in Astronomy and Astrophysics, 20, 159

Shipp N., et al., 2021, ApJ, 923, 149

Shu F. H., 2016, ARA&A, 54, 667

Silva M. D. V., Napiwotzki R., 2011, MNRAS, 411, 2596

Steele J., 2021, astronomy, Babylonia, doi:10.1093/acrefore/9780199381135.013.8631, `https://oxfordre.com/classics/view/10.1093/acrefore/9780199381135.001.0001/acrefore-9780199381135-e-8631`

Stone N. C., Metzger B. D., 2016, MNRAS, 455, 859

Stone N. C., Vasiliev E., Kesden M., Rossi E. M., Perets H. B., Amaro-Seoane P., 2020, Space Sci. Rev., 216, 35

Tauris T. M., 2015, MNRAS, 448, L6

Tauris T. M., Takens R. J., 1998, A&A, 330, 1047

Tody D., 1986, in Crawford D. L., ed., Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 627, Instrumentation in astronomy VI. p. 733, doi:10.1117/12.968154

Tody D., 1993, in Hanisch R. J., Brissenden R. J. V., Barnes J., eds, Astronomical Society of the Pacific Conference Series Vol. 52, Astronomical Data Analysis Software and Systems II. p. 173

Toomre A., 1964, ApJ, 139, 1217

Trac H., Sills A., Pen U.-L., 2007, MNRAS, 377, 997

Vasiliev E., 2019a, MNRAS, 482, 1525

Vasiliev E., 2019b, MNRAS, 482, 1525

Vasiliev E., Belokurov V., Erkal D., 2020, Monthly Notices of the Royal Astronomical Society, 501, 2279

Vasiliev E., Belokurov V., Erkal D., 2021, MNRAS, 501, 2279

Verberne S., et al., 2024a, MNRAS, 533, 2747

Verberne S., Koposov S. E., Rossi E. M., Marchetti T., Kuijken K., Penoyre Z., 2024b, A&A, 684, A29

Verberne S., Rossi E. M., Koposov S. E., Penoyre Z., Cavieres M., Kuijken K., 2025, A&A, 696, A218

Virtanen P., et al., 2020, Nature Methods, 17, 261

Wada K., Baba J., Saitoh T. R., 2011, ApJ, 735, 1

Wegg C., Gerhard O., 2013, MNRAS, 435, 1874

Wegg C., Gerhard O., Portail M., 2015, MNRAS, 450, 4050

Weiland J. L., et al., 1994, ApJ, 425, L81

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

Xiang M., Rix H.-W., 2022, Nature, 603, 599

Xue X.-X., Rix H.-W., Ma Z., Morrison H., Bovy J., Sesar B., Janesh W., 2015, ApJ, 809, 144

Yan H., et al., 2022, The Innovation, 3, 100224

Yelda S., Ghez A. M., Lu J. R., Do T., Meyer L., Morris M. R., Matthews K., 2014, ApJ, 783, 131

York D. G., et al., 2000, AJ, 120, 1579

Young A., et al., 2023, A&A, 670, A36

Yu Q., Madau P., 2007, MNRAS, 379, 1293

Yu Q., Tremaine S., 2003, ApJ, 599, 1129

Yu F., et al., 2024, ApJ, 975, 81

Zhang F., Lu Y., Yu Q., 2013, ApJ, 768, 153

Zhang X., Green G. M., Rix H.-W., 2023, MNRAS, 524, 1855

Zhao G., Zhao Y.-H., Chu Y.-Q., Jing Y.-P., Deng L.-C., 2012, Research in Astronomy and Astrophysics, 12, 723

Zonca A., Singer L., Lenz D., Reinecke M., Rosset C., Hivon E., Gorski K., 2019, Journal of Open Source Software, 4, 1298

de Jong R. S., et al., 2019, The Messenger, 175, 3

von Fellenberg S. D., et al., 2022, ApJ, 932, L6

Samenvatting

Door het oog van de naald

Iedereen heeft waarschijnlijk wel eens gehoord van zwarte gaten: objecten die zo massief zijn dat zelfs licht er niet aan kan ontsnappen. Grofweg zijn er twee soorten zwarte gaten: stellaire zwarte gaten en superzware zwarte gaten. Het verschil is dat stellaire zwarte gaten de relatief talrijke restanten zijn van zware sterren, terwijl superzware zwarte gaten zeldzaam zijn en zich meestal bevinden in de centra van sterrenstelsels, die miljarden sterren bevatten.

Van superzware zwarte gaten is bekend dat ze worden omringd door vele sterren in een dicht opeengepakte stercluster. Deze complexe dynamische omgeving roept de vraag op: wat als twee sterren die om elkaar heen draaien dicht langs een superzwaar zwart gat komen, maar niet zo dicht dat ze vernietigd worden? Dit is de vraag die de astronoom Jack G. Hills zichzelf stelde in een wetenschappelijk artikel dat in 1988 werd gepubliceerd. Hij voorspelde dat door de getijdenkracht van het zwarte gat één van de sterren gevangen zou worden in een korte baan rond het superzware zwarte gat, terwijl de andere ster met een extreem hoge snelheid zou worden weggeslingerd—sterren die hij supersnelle sterren noemde (Hills 1988). In Fig. 1 tonen we een artistieke impressie van deze interactie, die het Hills mechanisme wordt genoemd.

Figuur 1: Artistieke impressie van het wegslingeren van een supersnelle ster na de interactie tussen een massief zwart gat en een dubbelster. *Afbeelding: James Josephides, Swinburne Astronomy Productions.*

In de tijd dat Hills deze voorspelling maakte werd het bestaan van superzware zwarte gaten nog bediscussieerd. Hills verwachtte zelfs dat de

ontdekking van een supersnelle ster bijna onweerlegbaar bewijs zou leveren voor superzware zwarte gaten. Toch duurde het bijna 20 jaar voordat de eerste supersnelle ster werd ontdekt (Brown et al. 2005). Tegen die tijd waren superzware zwarte gaten al algemeen geaccepteerd. Desondanks zorgde de ontdekking voor een hernieuwde belangstelling voor het onderwerp. Supersnelle sterren vormen een uniek hulpmiddel om allerlei astrofysische vraagstukken te onderzoeken: zo kunnen we bijvoorbeeld leren over dynamische interacties in de drukke en chaotische omgeving nabij het superzware zwarte gat in het centrum van ons Melkwegstelsel en over de massa en vorm van de Melkweg.

Tot op heden kennen we slechts een handvol supersnelle sterren, en er is slechts één ster gevonden waarvan ondubbelzinnig kan worden vastgesteld dat deze uit het Galactisch Centrum is weggeslingerd. Gelukkig maken recente grootschalige waarnemingscampagnes het mogelijk om systematisch naar deze zeldzame sterren te zoeken. De ruimtetelescoop *Gaia* heeft in het bijzonder de posities en bewegingen van bijna twee miljard sterren gemeten. Deze schat aan gegevens wordt aangevuld met waarnemingen vanaf de grond die extra informatie leveren over miljoenen van deze sterren.

Dit onderzoek richt zich op het karakteriseren van de populatie supersnelle sterren in de Melkweg en het bepalen van de eigenschappen van de omgeving nabij het superzware zwarte gat in het Galactisch Centrum. We gebruiken een combinatie van de meest geavanceerde waarnemingen en speciaal ontwikkelde simulaties om het probleem gelijktijdig vanuit een observationeel en theoretisch perspectief te benaderen.

In Hoofdstuk 2 richtten we ons op het uitbreiden van het aantal sterren met snelheidsmetingen langs de gezichtslijn (radiële snelheid), die *Gaia* 'slechts' voor ongeveer 34 miljoen van de bijna twee miljard sterren in haar catalogus levert. Deze snelheden zijn namelijk nodig om een driedimensionaal beeld te vormen van de baan van een potentiële supersnelle ster en daarmee van zijn oorsprong. Hiervoor analyseerden we de lage resolutie spectroscopische metingen die *Gaia* in zijn laatste catalogus openbaar heeft gemaakt. Hoewel deze gegevens oorspronkelijk niet voor dit doel bedoeld waren, motiveerden de extreme snelheden van supersnelle sterren en de daardoor lagere vereiste meetnauwkeurigheid ons om het toch te proberen. We vergeleken de *Gaia* spectra met een raster van modelspectra om een eerste meting te verkrijgen. We corrigeerden systematische fouten in deze eerste metingen door onze resultaten te kalibreren met behulp van een groot aantal sterren met bekende radiële snelheden uit diverse catalogi. Onze volledige gepubliceerde catalogus bevat radiële snelheden voor ongeveer 125 miljoen objecten, naar ons weten de grootste in zijn soort op het moment van schrijven. Hoewel we met succes radiële snelheden konden meten met deze spectra, bleek het identificeren van potentiële supersnelle sterren lastig door een fractie van bronnen met onbetrouwbare metingen. Aangezien supersnelle sterren extreem zeldzaam zijn (ongeveer 1 op 100 miljoen sterren),

kan zelfs een zeer kleine fractie onbetrouwbare metingen problematisch zijn.

In Hoofdstuk 3 gebruikten we een andere methode om potentiële supersnelle sterren te vinden. We gebruikten de gemeten eigenbeweging (hoeksnelheid) aan de hemel die *Gaia* voor ongeveer 1,5 miljard sterren heeft bepaald, om te zoeken naar sterren die lijken te bewegen langs een baan die rechtstreeks uit het Galactisch Centrum komt. We maakten een catalogus van 600 veelbelovende kandidaten en namen vervolgwaarnemingen vanaf de grond voor ongeveer 200 van deze sterren. We gebruikten het uitblijven van nieuwe ontdekkingen in deze steekproef om de zeldzaamheid van supersnelle sterren en hun massaverdeling te bepalen. Bovendien voorspelden we de eigenschappen van de nog onontdekte populatie supersnelle sterren die in de *Gaia*-gegevens verborgen zit.

In Hoofdstuk 4 gebruikten we de waarnemingen uit Hoofdstuk 3 om ons te informeren over de populatie en eigenschappen van sterren die gevangen worden in het Hills mechanisme. Het Galactisch Centrum herbergt een populatie sterren op sterk elliptische banen dicht bij het superzware zwarte gat, bekend als de S-ster cluster. We simuleerden gelijktijdig de uitgestoten en gevangen sterren, waarbij we de uitgestoten sterren vergeleken met onze supersnelle sterren zoektocht en de gevangen sterren met waarnemingen van het Galactisch Centrum. We vinden dat geen enkele enkele voorouderpopulatie van dubbelsterren zowel de waargenomen eigenschappen van supersnelle sterren als die van sterren nabij het superzware zwarte gat in het Galactisch Centrum kan verklaren. In plaats daarvan vinden we dat een recent gevormde schijf van zware sterren waarschijnlijk de frequentie van supersnelle sterren in de afgelopen ~ 10 miljoen jaar met ongeveer een orde van grootte heeft verhoogd ten opzichte van de achtergrondwaarde.

In Hoofdstuk 5 breidden we onze zoektocht naar supersnelle sterren uit naar de sterren die tot miljarden jaren geleden zijn weggeslingerd, maar binnen ons sterrenstelsel zijn gebleven. Hiervoor gebruikten we gegevens van de recent gepubliceerde, DESI Milky Way Survey. We simuleerden de verdeling van het impulsmoment van sterren die uit het Galactisch Centrum zijn weggeslingerd, en gebruiken dit in combinatie met de karakteristiek hoge abundantie metalen die voor deze sterren wordt verwacht, om statistisch te zoeken naar een populatie van weggeslingerde sterren uit het Galactisch Centrum in DESI. We gebruikten het uitblijven van een detectie van deze populatie om voor het eerst de frequentie (hoe vaak iets gebeurt) van supersnelle sterren uit het Galactisch Centrum te bepalen over een tijdschaal van ~ 5 miljard jaar, onafhankelijk van het onderliggende mechanisme. Dit betekent dat ons resultaat niet alleen de frequentie van het Hills mechanisme bepaald, maar ook de frequentie van alternatieve wegslingermechanismen die in de literatuur zijn voorgesteld.

Publications

First author

- *Searching for stars ejected from the Galactic Centre in DESI*
  Sill Verberne, Sergey E. Koposov, Elena Maria Rossi, Zephyr Penoyre
  Accepted in Astronomy & Astrophysics

- *Double progenitor origin of the S-star cluster*
  Sill Verberne, Elena Maria Rossi, Sergey E. Koposov, Zephyr Penoyre,
  Manuel Cavieres, Konrad Kuijken
  Published in Astronomy & Astrophysics
  Volume 696, Article Number A218, April 2025

- *Improved constraints on Galactic Centre ejection of hypervelocity stars
  based on novel search method*
  Sill Verberne, Elena Maria Rossi, Sergey E. Koposov, Tommaso
  Marchetti, Konrad Kuijken, Zephyr Penoyre, Fraser A. Evans, Dimitris Souropanis, Clár-Bríd Tohill
  Published in Monthly Notices of the Royal Astronomical Society
  Volume 533, Issue 3, September 2024, Pages 2747–2761

- *Radial velocities from Gaia BP/RP spectra*
  Sill Verberne, Sergey E. Koposov, Elena Maria Rossi, Tommaso
  Marchetti, Konrad Kuijken, Zephyr Penoyre
  Published in Astronomy & Astrophysics
  Volume 684, Article Number A29, April 2024

- *The radial supernova remnant distribution in the Galaxy*
  Sill Verberne, Jacco Vink
  Published in Monthly Notices of the Royal Astronomical Society
  Volume 504, Issue 1, June 2021, pages 1536–1544

- *First Ultraviolet Outburst Detected from ASASSN-18eh Strengthens
  Its Interpretation as a Cataclysimic Variable*
  Sill Verberne, David Modiano, Rudy Wijnands
  Published in Research Notes of the AAS
  Volume 4, Issue 12, id.233, December 2020

Contributed

- *TUVO-21acq: A new cataclysmic variable discovered through a UV outburst*
  David Modiano, Rudy Wijnands, David A.H. Buckley, Mariusz Gromadzki, **Sill Verberne**, Marieke van Etten
  Published in Astronomy & Astrophysics
  Volume 664, Article Number A165, August 2022

- *TUVOpipe: A pipeline to search for UV transients with Swift-UVOT*
  David Modiano, Rudy Wijnands, Aastha Parikh, Jari van Opijnen, **Sill Verberne**, Marieke van Etten
  Published in Astronomy & Astrophysics
  Volume 663, Article Number A5, July 2022

- *Discovery prospects of dwarf spheroidal galaxies for indirect dark matter searches*
  Shin'ichiro Ando, Bradley J. Kavanagh, Oscar Macias, Tiago Alves, Siebren Broersen, Stijn Delnoij, Thomas Goldman, Jim Groefsema, Jorinde Kleverlaan, Jaïr Lenssen, Toon Muskens, Liam X. Palma Visser, Ebo Peerbooms, Bram van der Linden, **Sill Verberne**
  Published in Journal of Cosmology and Astroparticle Physics
  Issue 10, article id. 040, October 2019

Curriculum Vitae

I was born on July 17th, 1998, in Amsterdam, Netherlands. Astronomy first peaked my interest in high school, which I attended at the Gerrit van der Veen College between 2010 and 2016. The school had a focus on art and culture, but I was more interested in the STEM subjects. I already enjoyed watching documentaries about space on TV, but about halfway through high school I got a new physics teacher that really ignited my interest in space and physics more generally. Gideon Koekoek was still an active researcher in gravitational waves, but had decided to also teach highschool physics during this period. I became more and more curiosity driven and decided to study physics and astronomy at the University of Amsterdam and the Vrije Universiteit Amsterdam.

During my bachelors degree I noticed that I found the astronomy courses consistently the most fun and interesting. I took every opportunity to register for the astronomy activities and courses on offer. For my bachelor thesis project I decided to work with dr. Jacco Vink on the distribution of the remnants of exploding stars, known as supernova remnants. The project was my first exposure to mostly independent research and a great learning experience for me. The results were interesting enough that I decided that I wanted to try to get it published during my masters, which resulted in my earliest first-author peer-reviewed publication in 2021. Already during my bachelors, but especially during my masters, I came to realise that especially observational astronomy appealed greatly to me. In parallel to publishing my bachelors research I started my 1 year master research project in the group of prof. dr. Rudy Wijnands. This research projected was very observationally oriented and the goal was to contribute to a pipeline that could automatically detect ultraviolet transient events using public data from the Neil Gehrels Swift Observatory. I learned a lot about research, critical thinking, and academic writing during this time, not in small part due to the mentoring of dr. David Modiano.

At the start of the second year of my master, I decided that I wanted to continue in astronomy after the master and pursue a PhD. My most important criterion on choosing a PhD position was that it needed to be observational. I found the perfect fit with a PhD position in the group of prof. dr. Elena Maria Rossi at Leiden Observatory, co-supervised by prof. dr. Sergey E. Koposov at the University of Edinburgh and dr. Tommaso Marchetti at the European Southern Observatory. I started my PhD at the tail-end of COVID, which fortunately allowed me to come into the office regularly. The aim of my PhD has been to search for and characterise rare, extremely fast stars ejected by a dynamical interaction with a massive black hole using large surveys while also obtaining my own observations. I managed to write four more first author peer-reviewed publications during my PhD which can be read in this thesis.

Even though I will not continue my career in academic astronomy research, the experiences and skills I acquired during my studies and PhD will certainly form who I am and how I work moving forward.

Acknowledgements

These acknowledgements would not be complete without a shout out to the Pedro Pancakes. Our many dinners and TV-show nights were a much needed relief from science. Special thanks to Chloe, Colin, Nicco, and Nicole for hosting so many of them.

Of course, a massive thanks also goes to everyone who was there for the coffee (or puzzle) breaks throughout these years! Every day, these moments were something to look forward to and I will cherish the memories I made during them. I'm sorry if I missed people: Alessia, Andrés, Ani, Beth, Celine, Christiaan, Darío, Dennis, Elena, Elia, Erik, Esther, Gijs, Jelle, Jinyi, Jurjen, Karin, Kevin, Louis, Lucie, Luna, Marta, Orestis, Osmar, Piyush, Pranjal, Rico, Rob, Roland, Sai, Sam, Sid, Silvia, Thomas, Timo, Vanesa, Victor, Will, Yuze, and Zeynab.

Muyao, thank you so much for your unconditional love and support over this past year. You have given me so much and it means the world to me!

Natuurlijk zou ik hier niet zijn zonder de steun en toeverlaat van mijn familie. Papa, mama, dank jullie voor het altijd ondersteunen en aanmoedigen van mijn ambities. Ik heb nooit druk gevoeld om een bepaalde richting op te gaan, maar ik kan altijd rekenen op jullie steun om mijn eigen weg te vinden. Finn, dank je wel voor je steun en jouw inspirerende en eigenwijze blik. Dide, jouw nieuwsgierige en aanmoedigende karakter maken het altijd leuk om te praten.