Characterizing the Gaia radial velocity sample selection function in its native photometry

Jan Rybizki, Hans-Walter Rix, Markus Demleitner, Coryn A. L. Bailer-Jones and William J. Cooper

1 Max Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany
2 Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Mönchhofstraße 12-14, D-69120 Heidelberg, Germany
3 School of Physics, Astronomy and Mathematics, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK
4 Istituto Nazionale di Astrofisica, Osservatorio Astrofisico di Torino, Strada Osservatorio 20, I-10025 Pino Torinese, Italy

Accepted 2020 October 2. Received 2020 September 30; in original form 2020 June 26

ABSTRACT

The Gaia Data Release 2 (DR2) radial velocity sample (GDR2 RVS), which provides six-dimensional phase-space information on 7.2 million stars, is of great value for inferring properties of the Milky Way. Yet a quantitative and accurate modelling of this sample is hindered without knowledge and inclusion of a well-characterized selection function. Here we derive the selection function for objects entering the Gaia DR2 time span, through estimates of the internal completeness, i.e. the ratio of GDR2 RVS sources compared to all Gaia DR2 sources (GDR2 all). We show that this selection function or ‘completeness’ depends on basic observables, in particular the apparent magnitude $G_{\text{RVS}}$ and colour $G - G_{\text{RP}}$, but also on the surrounding source density and on sky position, where the completeness exhibits distinct small-scale structure. We identify a region of magnitude and colour that has high completeness, providing an approximate but simple way of implementing the selection function. For a more rigorous and detailed description we provide PYTHON code to query our selection function, as well as tools and ADQL queries that produce custom selection functions with additional quality cuts.

Key words: space vehicles: instruments – virtual observatory tools – software: public release – Galaxy: kinematics and dynamics – Galaxy: stellar content

1 INTRODUCTION

Gaia Data Release 2 (DR2) contains median radial velocities and their uncertainties for 7224 631 stars (Katz et al. 2019). To fully exploit this data set, the astronomical community requires the Gaia DR2 radial velocity sample (GDR2 RVS) selection function. A selection function quantifies the probability of an object entering a sample (here, the RVS sample) as a function of its observables, such as magnitudes, colours, and position on sky. This is required for essentially all ensemble modelling of such data, as in such cases any model that predicts observables must first be multiplied with the selection function before a meaningful comparison to data is possible. The characterization of the GDR2 RVS selection function has been hampered as no $G_{\text{RVS}}$ photometry, covering the far-red optical region of the Radial Velocity Spectrometer (RVS) spectra, has been published. This is because of the Gaia spectroscopic pipeline not yet being fully calibrated (Sartoretti et al. 2018). Approximating $G_{\text{RVS}}$ from $G$ and $G_{\text{RP}}$ magnitudes allows us to build the selection function: $S(G_{\text{RVS}}, G - G_{\text{RP}}, (\alpha, \delta))$. $G_{\text{RVS}}$ provides the brightness range over which the RVS instrument (Cropper et al. 2018) could collect enough signal over the Gaia DR2 time span and where the detector would not saturate. The $G - G_{\text{RP}}$ colour is a proxy for the effective temperature of the star, for which reliable radial velocity determinations can be obtained from the measured spectral window. The sky position can enter the selection function via the source density, which varies dramatically across the sky, through the Gaia scanning law (Boubert & Everall 2020; Boubert, Everall & Holl 2020). Sky position also enters the selection function through a mix of both the RVS spectral window assignment on neighbouring RVS sources and (in the case of Gaia DR2) through a pre-selection of RVS sources based on other input catalogues (Smart & Nicastro 2014).

This paper is structured as follows. In Section 2, we show how $G_{\text{RVS}}$ can be derived. Since our empirical approach derives the internal completeness of the GDR2 RVS sample with respect to the GDR2 all sample, we show how this can be generalized to the external completeness, i.e. selection function in Section 6. Section 3 looks at the completeness with colour and magnitude. In Section 4, we examine the completeness over the sky. Section 5 highlights correlations of the GDR2 RVS selection function with other parameters. We then explain the generation and usage of our GDR2 RVS selection function in Section 7, and conclude with a summary in Section 8.

2 USING THE NATIVE PHOTOMETRY OF THE RVS INSTRUMENT

The Radial Velocity Spectrometer (RVS) is an integral field spectrograph (Cropper et al. 2018) that observes in the near-infrared at $\lambda = [845, 872]$ nm, which is redder than the mean wavelength of the $G$ band ($\lambda = [330, 1050]$ nm) and also slightly redder than the mean wavelength of the $G_{\text{RP}}$ band ($\lambda = [630, 1050]$ nm; Evans et al. 2018).
For Gaia DR2, the processing of radial velocities was limited to sources with $G_{\text{RVS}} < 12$ mag, but since no $G_{\text{RVS}}$ magnitudes have been published this statement is ambiguous. To add to this confusion there are three different ways to determine $G_{\text{RVS}}$ (Sartoretti et al. 2018; Katz et al. 2019), which were used in different parts of the processing and affect the completeness of the GDR2RVS sample.

(i) The ‘on-board’ $G_{\text{RVS}}$ is derived on-board the satellite, is used for various on-board automated decisions, and is also transmitted to the ground.\(^1\)

(ii) The ‘external $G_{\text{RVS}}$’ was calculated by Smart & Nicastro (2014) from a collection of ground-based photometric catalogues. These $G_{\text{RVS}}$ are part of the so-called Initial Gaia Source List (IGSL).

(iii) The ‘internal $G_{\text{RVS}}$’ is the magnitude derived by the Gaia spectroscopic pipeline using the flux recorded in the RVS spectra.\(^2\)

Ideally we would like to use the ‘internal $G_{\text{RVS}}$’ in this work but since it was not published we are instead forced to use yet another quantity that we call the approximated $G_{\text{RVS}}$ and was fit to the ‘internal $G_{\text{RVS}}$’ after processing was finished (Gaia Collaboration et al. 2018). The approximated $G_{\text{RVS}}$ is a function of $G$ and $G_{\text{RP}}$ as specified in equations (2) and (3) from Gaia Collaboration et al. (2018), which gives $G_{\text{RVS}} = G_{\text{RP}}$ as a fourth-order polynomial in $G - G_{\text{RP}}$. These equations\(^3\) are valid within $0.1 < G - G_{\text{RP}} < 1.7$ and approximate the true $G_{\text{RVS}}$ to $\approx0.1$ mag precision.\(^4\)

The advantage of switching to this approximated $G_{\text{RVS}}$ value, which is closer to what RVS observes, is that it reduces the impact of colour variations on the selection function. This results in a sharper cut-off in the magnitude distribution that can be inspected in Fig. 1. For $G_{\text{RVS}}$, 5.3 million sources are brighter than the mode, whereas for $G$ this only holds for 4.5 million sources.

---

\(^1\)For sources with on-board $G_{\text{RVS}} > 16.2$ mag RVS windows are assigned, sources with on-board $<7$ get 2D windows assigned (Sartoretti et al. 2018).

\(^2\)Sources with internal $G_{\text{RVS}} > 14$ mag were not published in Gaia DR2.

\(^3\)Can be inspected in the queries of Appendix A.

\(^4\)We add 0.05 mag at each limit to the colour range, i.e. $0.05 < G - G_{\text{RP}} < 1.75$ in order for all bins to have the same width, i.e. 0.1 mag.

---

Table 1. GDR2RVS star counts within different cuts.

| Source number in millions |
|----------------------------|
| GDR2RVS (all)              7.225 |
| $G_{\text{RVS}}$ approximation 7.213 |
| High-completeness colour range (equation 2) 6.872 |
| High-completeness mag range (equation 1) 5.699 |
| High-completeness CMD area (equations 1 and 2) 5.443 |

---

To select the stars to be processed for Gaia DR2, the consortium used the external $G_{\text{RVS}}$, when available. For the 8 per cent of sources where it was not, the on-board $G_{\text{RVS}}$ was used instead (Sartoretti et al. 2018).

The external $G_{\text{RVS}}$ has been derived from multiple catalogues, with varying zero-points, and a multitude of photometric bands, resulting in multiple transformation formulas. The sharp selection in external $G_{\text{RVS}}$ translates into a shallower one in the approximated $G_{\text{RVS}}$, though, as noted above, this is still sharper than when $G$ is used. Throughout the paper we refer to the approximated $G_{\text{RVS}}$ as $G_{\text{RVS}}$ unless we add the prefixes from the three bullet points above.\(^5\)

The tail of the $G_{\text{RVS}}$ magnitude distribution comprises mostly stars with an external $G_{\text{RVS}}$ brighter than 12th mag, but that are fainter in the approximated $G_{\text{RVS}}$.

For the $G_{\text{RVS}}$ distribution in Fig. 1 we note that for the magnitude range, $2.95 < G_{\text{RVS}} < 12.05$,\(^6\) the GDR2RVS sample follows the GDR2all sample distribution quite well. Stars $< 2.95$ $G_{\text{RVS}}$ mag do not usually enter the GDR2RVS sample owing to saturation of the core of RVS spectra at approximately $G > 4$ mag (Katz et al. 2019).

We limit our investigation to the colour range for which $G_{\text{RVS}}$ can be approximated. From Table 1 we deduce that we are only losing 0.2 per cent of GDR2RVS sources because of the colour limit and most of them actually owing to a missing $G_{\text{RP}}$ measurement.

In Fig. 2, we compare the GDR2all sources with GDR2RVS sources where $G < 12$ mag in order to assess the GDR2RVS sample completeness with colour. As we can see, only a minute fraction of GDR2RVS sources reside outside the $G_{\text{RVS}}$ approximation range.

---

\(^5\)Only in the context of GeDR3mock (Rybizki et al. 2020) is the $G_{\text{RVS}}$ not approximated but calculated using an RVS passband.
were excluded in redder sources. Colour dependence, which is blurred more at the red end owing to 2019). Therefore sources that needed radial velocity templates6 with the pseudo-continuum at the shortest wavelengths (Katz et al. are dominated by TiO molecular bands that inhibit the determination of the effective temperature range for radial velocity detection. The grey dashed lines show the $T_{\text{eff}}$ limits for radial velocity determination. (depicted with grey dashed lines). For a central colour range with $0.35 < G - G_{\text{RP}} < 1.25$ (depicted with red dashed lines), we see that the GDR2$_{\text{RVS}}$ sample is almost complete with respect to the GDR2$_{\text{all}}$ sample. The reason for the decrease in completeness for very blue and red sources is that hot stars have strong Paschen lines affecting the determination of radial velocities from the calcium triplet. Similarly, spectra of cool stars are dominated by TiO molecular bands that inhibit the determination of the pseudo-continuum at the shortest wavelengths (Katz et al. 2019). Therefore sources that needed radial velocity templates6 with an effective temperature outside the range $3550 < T_{\text{eff}} [\text{K}] < 6900$ were excluded in Gaia DR2. This translates into a relatively sharp colour dependence, which is blurred more at the red end owing to dust reddening as this extends the radial velocity determination to redder sources.

This can be seen in Fig. 3 where the $G - G_{\text{RP}}$ versus effective temperatures are shown for the GDR2$_{\text{RVS}}$ sample, colour coded by the extinction estimate from GDR2 (Andrae et al. 2018). Interestingly, the extinction estimates also increase for sources bluer than 0.35 mag and therefore hotter than 6900 K. Fig. 3 also illustrates that our defined high-completeness colour range from equation (2) comes from the effective temperature range for radial velocity detection.

In the following, we will inspect in more detail the joint colour–magnitude dependence of the internal completeness of the GDR2$_{\text{RVS}}$ sample. Ranges indicated by equations (1) and (2) will guide us where to expect nearly full completeness.

3 DEPENDENCE ON COLOUR AND MAGNITUDE

In order to obtain a better differentiated view of the GDR2$_{\text{RVS}}$ sample completeness, we show the internal completeness, i.e. GDR2$_{\text{RVS}}$/GDR2$_{\text{all}}$ source count, for colour–magnitude bins in the left-hand panel of Fig. 4. The red dashed lines show the colour–magnitude ranges of high completeness as specified in equations (1) and (2). For the GDR2$_{\text{RVS}}$ sample we have 7.21 million sources displayed (all sources within the $G_{\text{RVS}}$ approximation), of which 75 per cent are within the red dashed lines. For the GDR2$_{\text{all}}$ sample, 6.57 million sources are within the red dashed lines meaning that 83 per cent of the sources inside the red dashed lines have an RVS measurement. The colour bins that are adjacent but outside of the red dashed lines still have some higher fractions of completeness, but beyond these the completeness decreases fast. The only notable exception is the faint red corner at $G_{\text{RVS}} \sim 11$ mag and $G - G_{\text{RP}} \sim 1.4$ mag, which we attribute to dust-reddened sources with a well-defined calcium triplet despite their red colour. The drop off for objects with $G_{\text{RVS}} > 12$ mag is quite sharp for all colour bins. There seems to be a decrease in completeness at $G_{\text{RVS}} \sim 7$ mag. This was the on-board $G_{\text{RVS}}$ magnitude limit, where the Gaia pipeline switched from 2D to 1D window assignment (Sartoretti et al. 2018). Those sources with $G_{\text{RVS}} > 7$ are still relatively bright and produce spurious sources that also get 1D windows allocated by the SkyMapper. This then potentially produces window conflicts, which (because of the non-processing of truncated windows) leads to a lowered completeness in the range 7–9 mag $G_{\text{RVS}}$, cf. section 3.2 Sartoretti et al. (2018), contrary to 2D windows that are processed even when blended.

Curiously, the decrease in completeness at around seventh $G_{\text{RVS}}$ happens at a somewhat brighter $G_{\text{RVS}}$ in the red than in the blue. This strongly indicates a colour dependence of the on-board $G_{\text{RVS}}$ estimate. The lowered completeness vanishes for fainter sources as these produce less spurious sources. In crowded regions of the sky (especially where the number of visits is still low, i.e. towards the Galactic Centre and the anticentre)7 the effect of lowered completeness at the beginning of the 1D window assignment is worst, cf. left-hand panel of Fig. 6.

In the right-hand panel of Fig. 4 the source density of the GDR2$_{\text{RVS}}$ sample per colour–magnitude bin is shown. Here we see that in absolute numbers there are still a significant number of GDR2$_{\text{RVS}}$ sources that have $G_{\text{RVS}} > 12$ mag or $G - G_{\text{RP}} > 1.25$ mag.

4 DEPENDENCE ON SOURCE DENSITY AND SKY POSITION

Here we investigate the dependence of the selection function with respect to the position on the sky $S(\alpha, \delta)$. As we will see in Section 4.1, this has imprints of the global source density, as well as the close RVS pairs that compete in the spectral window allocation. Furthermore, the dependence on input catalogues is still visible in Gaia DR2 as we will see in Section 4.2. In the generation of our selection function, the sky position enters through HEALPix bins, while the source density is only accounted for indirectly by lowering the HEALPix level until enough sources are in a respective selection function bin. See Section 7 for details.

4.1 Source density

4.1.1 Global

To the first order, the completeness is driven by projected stellar density (we neglect the sources from the second telescope in our work, but discuss its contribution in Appendix B), which results

6For 18 per cent of the GDR2$_{\text{RVS}}$ sources, ground-based effective temperatures were used and the remaining were derived from the spectra themselves using only 28 templates (Sartoretti et al. 2018; Katz et al. 2019). We did not find strong signatures in the selection function resulting from these two different ways of determining the template spectra.

7This can be well seen when inspecting the HEALpix CMDs of https://www2.mpia-hd.mpg.de/homes/rybizki/Jan-per-cent20Rybizki-per-cent120-per-cent20Homepage_files/rvs_selection_visualisation_internmap.html switching between regions of high and low median transits as in fig. 9 of Katz et al. (2019).

Figure 3. $G - G_{\text{RP}}$ versus $T_{\text{eff}}$ from GDR2, colour coded by the mean extinction value for sources in the GDR2$_{\text{RVS}}$ sample. Of the 7.2 million sources, 4.8 million have those estimates from GDR2 (Andrae et al. 2018). Red dashed lines indicate the high-completeness colour from equation (2). The grey dashed lines show the $T_{\text{eff}}$ limits for radial velocity determination.

\[ T_{\text{eff}} [\text{K}] < 6900 \]

\[ G - G_{\text{RP}} < 1.25 \]
Figure 4. GDR2\textsubscript{RVS} internal completeness in colour and magnitude in \(0.1 \times 0.1\) mag binning. In the left-hand panel, the colour coding represents the ratio of GDR2\textsubscript{RVS} versus GDR2\textsubscript{all}; in the right-hand panel the number of stars in GDR2\textsubscript{RVS}. The white bars in the left-hand panel indicate the modes of the magnitude distribution in each colour bin. The red dashed lines indicate the recommended ‘high-completeness’ ranges in colour and magnitude from equations (1) and (2).

Figure 5. Aitoff projection of the sky density in Galactic coordinates. The Galactic Centre is in the middle, with longitude increasing to the left at HEALPIX level 5. This figure shows how we split the GDR2\textsubscript{RVS} sample into three source density regimes: low density (GDR2\textsubscript{RVS} 1.4 million; GDR2\textsubscript{all} 100.3 million), intermediate density (GDR2\textsubscript{RVS} 3.5 million; GDR2\textsubscript{all} 540.0 million), and high density (GDR2\textsubscript{RVS} 2.4 million; GDR2\textsubscript{all} 1052.6 million). The intermediate-density area is shown with pale transparency; the high-density part is towards the Galactic plane and the Magellanic Clouds. The low-density parts are at higher Galactic latitudes.

4.1.2 Close pairs

There are fundamental limitations of the Gaia satellite for the detection of nearby sources (sky separation), the so-called ‘contrast sensitivity’, which is a function of projected distance and magnitude difference (de Bruijne et al. 2015). For GDR2\textsubscript{all} sources with a \(G\) magnitude and a sky position measurement this has been characterized in Brandeker & Cataldi (2019), which is at approximately 0.4 arcsec for equal brightness sources. When requiring a colour measurement this distance increases to approximately 2.0 arcsec (fig. 9 in Arenou et al. 2018),\(^8\) because of truncated windows, which have not been included in Gaia DR2 photometry (Riello et al. 2018).

For the GDR2\textsubscript{RVS} sample there is an additional factor at the processing stage owing to its extremely elongated windows\(^9\) and relatively small source densities. The latter will change with future data releases, when going from a GRVS limit of 12 mag in GDR2, to 14 mag in GDR3, and perhaps as faint as 16 mag in GDR4. The maximum source densities per degree\(^2\) will increase from 2700 in GDR2 to (a theoretical maximum of) 50 000 in GDR3 or 300 000 in GDR4 (though the instrument limit is reached at 36 000). Sources in denser areas are lost because deblending was not yet activated and truncated windows that were not rectangular were not processed (Sartoretti et al. 2018). This does not apply to 2D windows, which were assigned to sources with \(G_{\text{RVS}} < 7\) mag.

In Fig. 7,\(^10\) we show log density plots of the sky separation versus \(G_{\text{RVS}}\) difference for GDR2\textsubscript{RVS} sources. We show this for high- (left-hand panel) and low-density sky areas (right-hand panel) as defined by Fig. 5. Owing to the low number statistics and relatively low \(G_{\text{RVS}}\) range of the sample the contrast sensitivity is not well sampled. In high-density areas chance alignments, which increase with distance squared, dominate the close pairs. For low-density areas these play

\(^8\)Except for closer equal brightness binaries that still entered the catalogue.

\(^9\)RVS spectral windows are 10 across-scan pixels times \(\sim 1300\) along-scan pixels wide, which corresponds to 1.77 \(\times \sim 75\) arcsec\(^2\) (Prusti 2012). This also means that a single RVS spectrum effectively takes up 135.4 arcsec\(^2\) on the sky. If tightly packed this would allow for approximately 100 000 sources degree\(^{-2}\) that is a factor of 3 higher than the instrument limit. In GDR2\textsubscript{RVS} sample the highest density per degree\(^2\) is 2681. If isotropically distributed, the mean distance of neighbouring sources in the highest density HEALPix would be approximately 55 arcsec.

\(^10\)The query for the close pair data can be inspected in Appendix A where we also provide a link to the data that includes all pairs between GDR2\textsubscript{RVS} sources and GDR2\textsubscript{all} sources.
Gaia RVS selection function

Figure 6. Same as left-hand panel of Fig. 4 (using the same colour scale) but in density subsets and only within the recommended high-completeness colour and magnitude ranges, see equations (1) and (2). From left to right are the high-, intermediate-, and low-density sample with sources degree$^{-2} > 100,000$, $10,000 < \text{and} < 100,000$, and $< 100,000$, respectively. The white bars show the mode of the magnitude distribution per colour bin. $f_{\text{comp}}$ represents the fractional completeness marginalized over colour or magnitude. The number of GDR2RVS sources (GDR2all sources) for these figures is 1.55 million (2.12 million), 2.72 million (3.14 million), and 1.18 million (1.30 million), from left to right (the HEALPIX fractions are 9 per cent, 39 per cent, and 52 per cent). The partition over the sky is shown in Fig. 5.

Figure 7. Density plots (colour scale is in log and not the same for both panels) of the distance versus $G_{\text{RVS}}$ difference for GDR2RVS sources. The left-hand panel shows high-density areas with more than 100,000 sources degree$^{-2}$. For the right-hand panel these close pairs are shown for low-density areas on the sky (less than 10,000 sources degree$^{-2}$). The number of pairs for each of those subsets (not all are necessarily depicted) is 86,000 and 10,000, respectively.

less of a role and true binaries in very close vicinity (overdensity up to 5 arcsec) are more abundant. The minimum separation of two GDR2RVS sources$^{11}$ is 5 across-scan pixels (Sartoretti et al. 2018) corresponding to 0.85 arcsec (Prusti 2012). The very close true binaries (1–2 arcsec) are less well sampled in the high-density regions indicative of source loss because of window truncation.

We do not attempt to include the effect of close pairs into our selection function, still from Fig. 7, one can approximate one if needed for binary treatment, e.g.

$$ S_{\text{close pairs}} = \begin{cases} \text{0 if } \Delta G_{\text{RVS}} > 0.5 \text{ mag and } \text{dist} < 2 \text{ arcsec}, \\ 1 \text{ else}. \end{cases} \quad (3) $$

4.2 Sky position

It is important to recall for the following subsections that the external $G_{\text{RVS}}$, on which the magnitude limit of 12 for most of the sources was applied comes from the IGSL3 (Smart & Nicastro 2014). This used several catalogues and transformation formula in generating a $G_{\text{RVS}}$ estimate, which we defined as external $G_{\text{RVS}}$ in this paper. IGSL3 has 15 million sources that have $G_{\text{RVS}} < 12$ mag. The $G_{\text{RVS}}$ determination was prioritized in order of the following input catalogue list (percentage of IGSL sources with external $G_{\text{RVS}} < 12$ mag in brackets): Sloan Digital Sky Survey (SDSS; Strauss et al. 2002) (4), Tycho-2 (Høg et al. 2000) (16), GSC23 (Lasker et al. 2008) (80), and negligible fractions from other input catalogues.

In Fig. 8, we can see that the priorities in choosing from different catalogues result in structure that we will recognize in the GDR2RVS completeness function. The Tycho set on the left has the highest priority and appears well-behaved. The SDSS footprint in the middle panel does not cover the whole sky. Areas and magnitude ranges that are left out are filled with the GSC23 data from the right-hand panel.

This results in the overall IGSL footprint seen in the left-hand panel of Fig. 9. For comparison we show the GDR2RVS sample density on the right. Differences are mainly a result of overlapping windows in dense areas (and therefore lost sources in the GDR2RVS sample on the right) and for 8 per cent of GDR2RVS sources also the on-board $G_{\text{RVS}}$ has been used instead of the external $G_{\text{RVS}}$. But still patterns from

$^{11}$This does not apply for sources with 2D windows.
the GSC23 and SDSS footprint can be recognized in the GDR2RVS sample.

4.2.1 Magnitude dependence of spatial completeness

In Fig. 10, we show the completeness over sky with $G_{\text{RVS}}$ magnitude when only using sources within the high-completeness colour ranges of equation (2). A video version can be downloaded from here.\(^{12}\)

At $G_{\text{RVS}} = 11$ mag (left-hand panel) the completeness shows little spatial structure, with only slightly lower completeness towards the Galactic plane. What can be already recognized is the SDSS footprint (lower completeness stripes perpendicular to the Galactic plane) that can also be seen in the left-hand panel of Fig. 9, which shows the source density of the IGSL catalogue with external $G_{\text{RVS}} < 12$ mag.

Since the SDSS transformation was preferentially used when calculating the sources in these stripes (cf. Fig. 8, middle panel), it likely means that the SDSS transformation results in a fainter external $G_{\text{RVS}}$ magnitude compared to the GSC23 transformation. This is consistent with fig. 16 of Evans et al. (2018) where the $G$ magnitude estimate from SDSS data varies over the sky and seems to be especially different in parts of the stripes.

At 12th $G_{\text{RVS}}$ mag (middle panel of Fig. 10), small patches of lower completeness start to emerge. These areas are point like and distributed over the entire sky. It could be that these spots are correlated with brighter zero-points in the respective GSC23 plates.

At $G_{\text{RVS}} = 12.5$ mag there are still a few areas of high completeness visible. The circular glass-canopy pattern that can be seen is a result of the derivation of most of the external $G_{\text{RVS}}$ magnitudes from the GSC23 BJ and RF magnitudes (Lasker et al. 2008). The magnitude zero-points of the plates seem to be a little fainter at the edges (possibly where they overlap each other; Katz, private communication). The canopy pattern delineates the borders of the

---

\(^{12}\)https://keeper.mpdl.mpg.de/f/db4ca9ee9bc34513bed1/
plates. It is also clearly visible in both panels of Fig. 9. There also seem to be plates for which the total area has a fainter zero-point.

The star counts for the GDR2\textsubscript{RVS} (GDR2\textsubscript{all}) in the 0.1 mag bins in Fig. 10 are 202 000 (234 000), 391 000 (569 000), and 114 000 (863 000) from left to right.

4.2.2 Colour dependence of spatial completeness

In Fig. 11, we show the completeness over sky with $G - G_{RVS}$ when only using sources within the high-completeness magnitude range of equation (1). A video version can be downloaded from here.\textsuperscript{13}

In the left-hand panel of Fig. 11 where the completeness for the colour bin at $G - G_{RVS} = 0.4$ mag is shown, we can see notably lower completeness wherever dust reddening is in place. This is owing to hot stars that have no treatment from the RVS pipeline getting reddened to colours that would usually (in the absence of dust reddening) result in a radial velocity measurement; this can be seen by the good completeness out of plane. A second-order effect can be witnessed in lower completeness areas that can also be seen but are even less pronounced in the left-hand panel of Fig. 10, for instance, the stripe from the South Galactic Pole left up to approximately $l = 90$ and $b = 0$. This feature and similarly lower completeness areas in the top right, Galactic anticentre, etc. are because of the scanning law and the necessity of at least two transits for a radial velocity measurement,\textsuperscript{14} cf. fig. 9 of Katz et al. (2019) or figs 5(c) and (f) of Boubert & Everall (2020).

The middle panel of Fig. 11 is similar to left-hand panel of Fig. 10 with the SDSS stripes of lower completeness and a lower completeness in the bulge because of blends; interestingly enough however, the Galactic plane overall suffers less from incompleteness. When going redder in the right-hand panel the completeness towards the plane improves the most. The reason for this counterintuitive effect is because sources in the good temperature regime ($3550 < T_{\text{eff}}$ [K] $< 6900$) for Ca triplet radial velocity determination fall into this colour bin (i.e. $G - G_{RVS} = 1.3$ mag) only when dust reddened and therefore in the Galactic plane.

The star counts for the GDR2\textsubscript{RVS} (GDR2\textsubscript{all}) in the 0.1 mag bins are 533 000 (746 000), 626 000 (731 000), and 115 000 (205 000) from left to right.

5 CORRELATION WITH OTHER QUANTITIES

We want to assess how well the GDR2\textsubscript{RVS} sample is representative of the GDR2\textsubscript{all} sample in terms of Gaia-derived quantities as a function of both magnitude and colour. If those Gaia-derived quantities were similar, one could apply a simple completeness correction by up-sampling the GDR2\textsubscript{RVS} sources to the numbers of GDR2\textsubscript{all} sources in their respective CMD bin.

In Fig. 12, we colour code the fractional difference of the mean PARALLAX, RUWE,\textsuperscript{15} and PHOT\textsubscript{BP-RP} EXCESS\_FACTOR\textsuperscript{16} (from left to right) between the GDR2\textsubscript{RVS} sample and the GDR2\textsubscript{all} sample on the CMD. Comparing to the left-hand panel of Fig. 4 we see that lower values of internal completeness generally coincide with differences of the GDR2\textsubscript{RVS} sample to the GDR2\textsubscript{all} sample for the shown parameters, and the patterns generally differ between the parameters. For the parallaxes it seems that sources with $G - G_{RP} < 0.35$ that are in the GDR2\textsubscript{RVS} sample have generally higher parallaxes compared to the GDR2\textsubscript{all} sample, whereas for sources with $G - G_{RP} > 1.05$ the opposite applies. The reason for this is that blue sources that are dust reddened are usually further away whilst simultaneously being too hot for a radial velocity determination (cf. the left-hand panel of Fig. 11 where for blue sources a low completeness in the Galactic plane areas can be seen). Therefore, blue sources in GDR2\textsubscript{RVS} are generally closer to us than the blue sources in GDR2\textsubscript{all}. Vice versa, the red sources ($1.05 < G - G_{RP} < 1.35$) in the GDR2\textsubscript{RVS} sample are dust-reddened bluer stars for which a RVS measurement is possible (cf. right-hand panel of Fig. 11). Those are usually further away than unreddened sources of the same colour for which molecular lines in the spectrum prohibit the RVS measurement. This of course also means that the proper motions are different between those two samples, potentially biasing kinematic selections.

For the RUWE and the PHOT\textsubscript{BP-RP} EXCESS\_FACTOR it seems that generally the GDR2\textsubscript{RVS} sample has better quality sources than the overall GDR2\textsubscript{all} sample. There is one exception to this rule at the red end (outside of the high-completeness zone) for $G_{RVS} \approx 11$ mag

\textsuperscript{13}https://keeper.mpdl.mpg.de/f/fedbe4cc51df544b738ab5/

\textsuperscript{14}The GDR2\textsubscript{RVS} sample has a hard lower limit of two transits, a median of 7 and a maximum of 201.

\textsuperscript{15}The renormalized unit weight error (RUWE) is expected to be around 1.0 for sources where the single-star model provides a good fit to the astrometric observations. A value significantly greater than 1.0 (e.g. >1.4) could indicate that the source is non-single or otherwise problematic for the astrometric solution.

\textsuperscript{16}The BP/RP excess factor is the sum of the integrated BP and RP fluxes divided by the flux in the G band (BP and RP are dispersed and therefore more prone to light from nearby sources than G). This excess is believed to be caused by background and contamination issues affecting the BP and RP data. Therefore a large value of this factor for a given source indicates systematic errors in the BP and RP photometry.
where quality indicators are better than for the GDR2_{all} sample. However, this only applies to a tiny fraction of the GDR2_{RVS} sources.

Therefore, we recommend to not use bins of the completeness function with completeness below some threshold, e.g. 90 per cent. If a larger sample is needed because of a weak signal, we caution that the GDR2_{RVS} sample properties might not be representative of the GDR2_{all} sample.

5.1 RVS selection in GeDR3mock with parallax bias

In order to assess the impact of the parallax bias on the spatial distribution of the GDR2_{RVS} sample, we applied the completeness function (HEALPix level 5, 0.1 G_{RVS} bins, 0.1 G - G_{RP} bins) to the GeDR3mock catalogue (Rybizki et al. 2020). We randomly chose 1000 subsets in each bin and took the subset that had the mean parallax closest to the mean of the respective:

(a) GDR2_{RVS} bin;
(b) GDR2_{all} bin.

And also (c) a truly random subset, which we do not consider here, but can, together with (a) and (b), be retrieved from here.\footnote{http://dc.g-vo.org/browse/gedr3mock/q}

In Fig. 13, we display the spatial distribution of sample (a) black lines versus sample (b) in blue lines. The left-hand panel shows the Galactic XZ projection for the blue sources (0.15 < G - G_{RP} < 0.35 mag) where we see that in the Galactic plane the GDR2_{RVS} sample (a) does not probe as deep as the GDR2_{all} sample (b) with about 0.3 kpc difference at the outer (3\sigma) density contour. Similarly, the right-hand panel shows the Galactic XY projection for the red sources (1.05 < G - G_{RP} < 1.35 mag) where we see that the GDR2_{RVS} sample probes approximately 1 kpc further than the GDR2_{all} sample on the far side of the Galaxy at the outer (3\sigma) density contour.

The difference is less pronounced than it probably is with the real Gaia data because we only chose from random subsets of sources from GeDR3mock; this choice does not perfectly represent the GDR2_{all} parallax distribution. Nevertheless, it should be beneficial to investigate those subsets and see how other selection effects can bias a sample, e.g. cuts on fractional parallax uncertainty.

6 FROM INTERNAL COMPLETENESS TO SELECTION FUNCTION

While in this paper we mainly investigate the internal completeness of GDR2_{RVS} sample with respect to GDR2_{all} sample and specifically the GDR2_{all} sample with G_{RP} measurement (because we require G - G_{RP} colour), evidence shows (Rybizki & Drimmel 2018) that we can use this internal completeness and approximate the external completeness (i.e. selection function) from it.

First, we need to look at the internal completeness of GDR2_{all} sources that have a G_{RP} measurement with respect to GDR2_{all} sources that at least have a G measurement for G < 15 mag. As we see in Fig. 14, the completeness over the sky is virtually complete except for small patches with lower completeness; these are mainly towards the North Galactic Pole plus a tiny patch with missing colour information.

Further on we need to verify that GDR2_{all} sample is externally complete in the magnitude range brighter than 15th G magnitude. Rybizki & Drimmel (2018) did this via a cross-match to the Two Micron All-Sky Survey (2MASS; Skrutskie et al. 2006) sources. Figures for this can be inspected in tutorial [2].\footnote{https://github.com/jan-rybizki/gdr2_completeness/blob/master/tutorials/\~\per\-cent5B2\-per\-cent5DCompleteness\-per\-cent20tutorial\_gdr2\_light.ipynb} While they find external completeness of GDR2_{all} to be virtually complete in the range 8 < G < 15 mag, Gaia seems to be losing sources in the Galactic disc and bulge for sources with G > 15 mag. Their external completeness assessment relies on the assumption that Gaia and 2MASS do independent measurements of the true sources within their respective G magnitude bins per HEALPix (they marginalize over colour). Despite their crude assumptions and large magnitude bins, their findings are indicative of the GDR2_{all} sample being close to the true completeness for sources with G < 15 mag. Similarly we expect no bias from spurious sources in GDR2_{all} because for G > 15 mag these have negligible contribution, cf. appendix C of Lindegren et al. (2018).

For the bright end, Boubert & Everall (2020) have shown that for sources with G > 3 mag the detection probability (and therefore completeness) is high.

We provide the GDR2_{all} sources with G_{RP} measurement internal completeness as a function of G and HEALPix of level 6 and integrate it into our RVS internal completeness function so that it should reflect the overall selection function.

7 ACCESSING GDR2_{RVS} COMPLETENESS FUNCTION

An example query illustrating how to download the data necessary to produce the GDR2_{RVS} selection function is given in Appendix A. User-specific quality cuts should be included in the query.

The internal completeness function created for easy usage as a PYTHON function has been generated in the following way.

(i) Fixed bin size of 0.2 mag in G_{RVS} covering a range from 2.9 to 14.1 mag.
The completeness function and the upper and lower $1\sigma$ percentile can be queried (for the right-hand panel the 743 000 stars with a colour range of $1.05 < G - G_{\text{RP}} < 1.35$ mag are shown. The contour lines represent the $1\sigma$, $2\sigma$, and $3\sigma$ levels encompassing 39.3 per cent, 86.5 per cent, and 98.9 per cent of the stellar density distribution projected on to the respective Galactic plane, i.e. $XZ$ on the left and $XY$ on the right.

(ii) Fixed bin size of 0.1 in $G - G_{\text{RP}}$, covering 0.05–1.75 mag.
(iii) HEALPIX level 6 baseline but with degradation down to level 0 until at least five GDR2$_{\text{all}}$ sources are available in that bin. If level 0 HEALPIX has less than five sources, then we use the whole sky also including bins with less than five sources.
(iv) In each level 6 HEALPIX the number of GDR2$_{\text{RVS}}$ and GDR2$_{\text{all}}$ sources are saved according to the above scheme and the respective HEALPIX level from which the numbers were taken.

The completeness function and the upper and lower $1\sigma$ percentile were generated according to the following scheme.

(i) Whenever a bin had no GDR2$_{\text{all}}$ or no GDR2$_{\text{RVS}}$ entries the completeness function was set to zero, as well as the upper and lower percentile.
(ii) In the rest of the cases the number of sources in the GDR2$_{\text{RVS}}$ bin was divided by the number of sources of the GDR2$_{\text{all}}$ bin that yields the completeness function.
(iii) For the upper and lower percentile we assumed a Poisson distribution with expected value of the number of sources in the GDR2$_{\text{RVS}}$ bin. We took the value at the 16th and 84th percentile of this distribution and divided by the number of sources in the GDR2$_{\text{all}}$ bin. We forced the value for upper to be less or equal the number of GDR2$_{\text{all}}$.

As a result, the completeness function is smoothed on the sky wherever source densities were too low. An illustration of this can be seen in the left-hand panel of Fig. 15, where the whole sky selection function is depicted for $G_{\text{RVS}} = 10.84$ mag and $G - G_{\text{RP}} = 1.2$ mag. In the middle panel, we show the corresponding fractional uncertainty calculated as $((\text{upper} - \text{lower})/\text{completeness})$. The right-hand panel shows the respective HEALPIX level, over which the star counts in the CMD bin have been averaged. The GDR2$_{\text{RVS}}$ completeness function can be queried (i.e. $G$, $b$, mag, col) and returns the above quantities and also the number of GDR2$_{\text{all}}$ and GDR2$_{\text{RVS}}$ sources in that respective bin (adding all of those yields total star counts), but also the sky area smoothed value (which is actually used to calculate the selection function). This is returned to a precision of 49 152 HEALPIXes, 56 magnitude bins, and 17 colour bins and can be accessed via the GDR2$_{\text{RVS}}$ completeness package (Rybizki & Drimmel 2018). Tutorial 5 illustrates its usage and shows some visualizations. We also provide the colour transformations ($G$, $G_{\text{RP}}$) → $G_{\text{RVS}}$ and the internal completeness of the GDR2$_{\text{all}}$ sample with $G_{\text{RP}}$ measurement versus the GDR2$_{\text{all}}$ sample. Interactive web visualizations of the completeness maps over the CMD and the sky are available here.19

8 SUMMARY
The GDR2$_{\text{RVS}}$ sample is an important subset of the Gaia DR2 catalogue. We have characterized the internal completeness of the GDR2$_{\text{RVS}}$ sample with respect to the GDR2$_{\text{all}}$ sample in the three dimensions sky position (represented by HEALPIX), $G_{\text{RVS}}$ magnitude, and $G - G_{\text{RP}}$ colour. In the magnitude range $3 < G_{\text{RVS}}$ [mag] < 14 evidence shows (Rybizki & Drimmel 2018) that this should be close to the external completeness (i.e. selection function). We show that the internal completeness is well characterized in the native $G_{\text{RVS}}$ band that can be approximated from $G_{\text{RP}}$ and $G - G_{\text{RP}}$. Imprints in the GDR2$_{\text{RVS}}$ Sample internal completeness arise from the IGSL3 input catalogues (mainly SDSS and GSC23) and its estimated external $G_{\text{RVS}}$ on which the magnitude limit of 12th mag has been applied. Completeness is lowered in high-density areas mainly owing to blended RVS spectral windows resulting in non-rectangular windows that have not been processed in Gaia DR2. Another important effect is the spectral template range between 3550 and 6900 K that could be used to determine radial velocity measurements. Together with dust reddening in the Galactic plane this leads to the counterintuitive effect that for red sources the completeness is best in the Galactic plane. A second-order effect

19https://www2.mpia-hd.mpg.de/homes/rybizki/index.html#publ
Figure 15. Example selection function (all-sky output without RP to G completeness correction) for $G_{\text{RVS}} = 10.84$ mag and $G - G_{\text{RP}} = 1.2$ mag. From left to right, the completeness, fractional uncertainty ((upper – lower)/(2 x completeness)), and the HEALPix level over which the star counts were averaged to calculate the completeness. The grey areas in the middle panel come from a zero division where the completeness function is zero.

This research or product makes use of public auxiliary data provided by ESA/Gaia/DPAC as obtained from the publicly accessible ESA Gaia SFTP.

This work was funded by the DLR (German space agency) via grant 50 QG 1403.

WJC studentship provided by University of Hertfordshire and Istituto Nazionale di Astrofisica, Osservatorio Astrofisico di Torino with computing infrastructure provided by STFC grant ST/R000905/1.

Software used: TOPCAT (Taylor 2005), HEALPix (Górski et al. 2005), ASTROPY (Astropy Collaboration et al. 2018), and CORNER (Foreman-Mackey 2016).

DATA AVAILABILITY

The data underlying this paper are available in the paper and in its online supplementary material.

ACKNOWLEDGEMENTS

We thank David Katz for insightful discussions on the RVS selection function. We also thank the Gaia group at MPIA for valuable feedback and Berry Holl for a very helpful email exchange on the Gaia scanning law. We thank the anonymous referee (also over a second iteration) for their thorough inspection and helpful comments. JR will not travel anywhere by aeroplane for the purpose of promoting this paper.

This work has made use of data from the European Space Agency (ESA) mission Gaia, processed by the Gaia Data Processing and Analysis Consortium (DPAC). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement.

REFERENCES

Andrae R. et al., 2018, A&A, 616, A8
ARENou F. et al., 2018, A&A, 616, A17
Astropy Collaboration et al., 2018, AJ, 156, 123
Boubert D., Everall A., 2020, MNRAS, 497, 4246
Boubert D., Everall A., Holl B., 2020, MNRAS, 497, 1826
Brandeker A., Cataldi G., 2019, A&A, 621, A86
Crapper M. et al., 2018, A&A, 616, A5
de Bruijne J. H. J., Allen M., Azaz S., Krone-Martins A., Prod’homme T., Hestroffer D., 2015, A&A, 576, A74
Evans D. W. et al., 2018, A&A, 616, A4
Foreman-Mackey D., 2016, J. Open Source Software, 1, 24
Gaia Collaboration et al., 2016, A&A, 595, A2
Gaia Collaboration et al., 2018, A&A, 616, A1
Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Hse E. et al., 2000, A&A, 355, L27
Katz D. et al., 2019, A&A, 622, A205
Lasker B. M. et al., 2008, AJ, 136, 735
Lindegren L. et al., 2018, A&A, 616, A2
Prusti T., 2012, Astron. Nachr., 333, 453
Riello M. et al., 2018, A&A, 616, A3
Rybizki J., Drimmel R., 2018, Astrophysics Source Code Library, record ascl:1811.018
Rybizki J. et al., 2020, PASP, 132, 074501
Sartoretti P. et al., 2018, A&A, 616, A6
Skrutskie M. F. et al., 2006, AJ, 131, 1163
Smart R. L., Nicastro L., 2014, A&A, 570, A87
Strauss M. A. et al., 2002, AJ, 136, 735
Taylor M. B., 2005, in Shopbell P., Britton M., Ebert R., eds, ASP Conf. Ser. Vol. 347, Astronomical Data Analysis Software and Systems XIV. Astron. Soc. Pac., San Francisco, p. 29

arising from the scanning law and the requirement of at least two transits with radial velocity measurement is also visible.

We show that the GDR2$^{\text{RVS}}$ sample can have partially different properties over the CMD than the corresponding GDR2$^{\text{all}}$ sample, e.g. in parallax but also in proper motion (not depicted here, though), which might bias kinematic selections. We similarly show that GDR2$^{\text{RVS}}$ sources usually have better quality indicators such as RUWE and PHOT$_{\text{BP,RP}}$:EXCESS$_{\text{NOISE}}$ compared to the GDR2$^{\text{all}}$ sample that prohibits easy completeness corrections. We apply the completeness function to a mock stellar catalogue, GE3mock, and explore the impact of the parallax difference on the spatial extent of colour subsets of the GDR2$^{\text{RVS}}$ sample.

We provide a completeness function in PYTHON that delivers the internal completeness as a function of HEALPix, magnitude and colour together with quality and uncertainty indicators, together with the functionality to generate the external completeness, i.e. the selection function, by taking into account missing $G_{\text{RP}}$ measurements in Gaia DR2. The necessary data are included but can also be generated for individual use cases by adapting our example ADQL query from Appendix A.

In Gaia DR3 (which is anticipated to be published end of 2021 at time of writing), a much improved GDR2$^{\text{RVS}}$ sample will be provided. It will be much deeper with a magnitude limit of internal $G_{\text{RVS}} = 14$ mag and with the dependency on external catalogues removed. Also the treatment of blended spectra will be included and the effective temperature range for which radial velocities will be determined might increase. This will greatly simplify future selection function determinations and should allow for more sophisticated modelling, e.g. including the nearby source contamination (as projected on to the sky) and taking into account the different scanning angles.
APPENDIX A: QUERYING THE SELECTION FUNCTION

This query uses Common Table Expressions, which will be part of the upcoming ADQL 2.1 standard, which at the time of writing among the Gaia-carrying VO data centres are only available on GAVO’s TAP service.\(^{20}\)

\(G_{RVS}\) is approximated using equations (2) and (3) from Gaia Collaboration et al. (2018).

The autocorrelation query for the close pairs is as follows:

The result of this query (which was subdivided in different HEALPIX in order to be able to retrieve all sources) can be downloaded from here\(^{21}\) in a 5, 10, and 20 arcsec version. It has also been cleaned from double entries and a \(G_{RVS}\) magnitude has been calculated for the second source.

\[^{20}\]http://www.g-vo.org/
\[^{21}\]https://keeper.mpdl.mpg.de/d/f2b841c75d7a42f6aad9/
WITH with_rvs AS (  -- Overall a similar construct to
SELECT radial_velocity, source_id, ra, dec,
    phot_g_mean_mag, phot_rp_mean_mag,
    phot_rp_mean_mag=0.402319-0.65124*(phot_g_mean_mag - phot_rp_mean_mag) + 1.0215 * POWER(phot_g_mean_mag - phot_rp_mean_mag, 2) - 1.3947
    * POWER(phot_g_mean_mag - phot_rp_mean_mag, 3) + 0.53768
    * POWER(phot_g_mean_mag - phot_rp_mean_mag, 4) AS phot_rvs
FROM gaia_dr2light
WHERE phot_g_mean_mag-phot_rp_mean_mag>1.4 AND phot_g_mean_mag-phot_rp_mean_mag>0.95
UNION ALL
    SELECT radial_velocity, source_id, ra, dec,
        phot_g_mean_mag, phot_rp_mean_mag,
        phot_rp_mean_mag=132.32-377.28*(phot_g_mean_mag - phot_rp_mean_mag) + 462.32 * POWER(phot_g_mean_mag - phot_rp_mean_mag, 2) - 190.97
        * POWER(phot_g_mean_mag - phot_rp_mean_mag, 3) + 34.926
        * POWER(phot_g_mean_mag - phot_rp_mean_mag, 4) AS phot_rvs
FROM gaia_dr2light
WHERE phot_g_mean_mag-phot_rp_mean_mag>=1.4 AND phot_g_mean_mag-phot_rp_mean_mag<=1.75)
rvobject AS (  
SELECT source_id, ra, dec, phot_rvs, phot_g_mean_mag, phot_rp_mean_mag
FROM with_rvs
WHERE radial_velocity IS NOT NULL)
SELECT a.source_id AS sid1, b.source_id AS sid2,
    a.phot_rvs AS rvsmag1, a.ra AS ral, a.dec AS dec1,
    a.phot_g_mean_mag AS gmag1, a.phot_rp_mean_mag AS rpmag1,
    a.radial_velocity AS rv1, b.phot_g_mean_mag AS gmag2, b.phot_rp_mean_mag AS rpmag2,
    b.radial_velocity AS rv2, DISTANCE(a.ra, a.dec, b.ra, b.dec)*3600 AS dist
FROM rvobject AS a  -- Here we only request pairs with an
JOIN gaia_dr2light AS b
ON (a.source_id=b.source_id AND DISTANCE(b.ra, b.dec, a.ra, a.dec)<20/3600.)
-- If adding any WHERE statement here the source_id
index gets lost and the query takes long
WITH with_rvs AS (  -- Overall a similar construct to
SELECT radial_velocity, source_id, ra, dec,
    phot_g_mean_mag, phot_rp_mean_mag,
    phot_rp_mean_mag=0.402319-0.65124*(phot_g_mean_mag - phot_rp_mean_mag) + 1.0215 * POWER(phot_g_mean_mag - phot_rp_mean_mag, 2) - 1.3947
    * POWER(phot_g_mean_mag - phot_rp_mean_mag, 3) + 0.53768
    * POWER(phot_g_mean_mag - phot_rp_mean_mag, 4) AS phot_rvs
FROM gaia_dr2light
WHERE phot_g_mean_mag-phot_rp_mean_mag>1.4 AND phot_g_mean_mag-phot_rp_mean_mag>0.95
UNION ALL
    SELECT radial_velocity, source_id, ra, dec,
        phot_g_mean_mag, phot_rp_mean_mag,
        phot_rp_mean_mag=132.32-377.28*(phot_g_mean_mag - phot_rp_mean_mag) + 462.32 * POWER(phot_g_mean_mag - phot_rp_mean_mag, 2) - 190.97
        * POWER(phot_g_mean_mag - phot_rp_mean_mag, 3) + 34.926
        * POWER(phot_g_mean_mag - phot_rp_mean_mag, 4) AS phot_rvs
FROM gaia_dr2light
WHERE phot_g_mean_mag-phot_rp_mean_mag>=1.4 AND phot_g_mean_mag-phot_rp_mean_mag<=1.75)
rvobject AS (  
SELECT source_id, ra, dec, phot_rvs, phot_g_mean_mag, phot_rp_mean_mag
FROM with_rvs
WHERE radial_velocity IS NOT NULL)
SELECT a.source_id AS sid1, b.source_id AS sid2,
    a.phot_rvs AS rvsmag1, a.ra AS ral, a.dec AS dec1,
    a.phot_g_mean_mag AS gmag1, a.phot_rp_mean_mag AS rpmag1,
    a.radial_velocity AS rv1, b.phot_g_mean_mag AS gmag2, b.phot_rp_mean_mag AS rpmag2,
    b.radial_velocity AS rv2, DISTANCE(a.ra, a.dec, b.ra, b.dec)*3600 AS dist
FROM rvobject AS a  -- Here we only request pairs with an
JOIN gaia_dr2light AS b
ON (a.source_id=b.source_id AND DISTANCE(b.ra, b.dec, a.ra, a.dec)<20/3600.)
-- If adding any WHERE statement here the source_id
index gets lost and the query takes long
APPENDIX B: INCREASED DENSITIES FROM SECOND FIELD OF VIEW

Since the two telescopes share the focal plane, the sources competing for CCD window allocation are more than just the sources at a single position in the sky. We neglect this effect in our work, because the viewing angle of 106.5° is sufficiently large and the position of the second telescope always changes for a specific position of the first telescope because of the scanning law. Therefore, plus the fact that most fractions of the sky contain low-density areas, we assume that only a small fraction of transits will have the situation where both fields are in crowded regions.

Figure B1. Mollweide projection of the mean per transit fractional increase in source density for GDR2 RVS sample owing to the second telescope in Galactic coordinates. The Galactic Centre is in the middle, with longitude increasing to the left at HEALpix level 6.

To make an approximate but quantitative assessment of the increase in source density per transit coming from the second telescope, we use the public Gaia scanning law\(^{22}\) (Gaia Collaboration et al. 2016), without accounting for the breaks (e.g. due to lost telemetry or the telescope going into safe mode) and estimate: for each field of view for telescope 1 (FoV1) on the sky, how many more sources there are from FoV2 (on average). We take every fifth data point in the scanning law file (which has a time resolution of 10 s), such that the individual pointings are 50 arcmin apart (50 s), which corresponds approximately to the distance of two neighbouring HEALpix at level 6. Level 6 HEALpix also have roughly the size of the field of view (FoV) of one telescope. Instead of an exact solution, each pointing is moved to the nearest HEALpix of level 6. We assume that per transit in each FoV Gaia observes the same amount of sources, which is simply taken from the HEALpix level 6 source count of GDR2RVS sample (as displayed in Fig. 9).

As can be seen in Fig. B1 the resulting mean per transit fractional increase for the GDR2RVS sample (average counts in FoV2 divided by counts in FoV1) is very low in high-density regions, cf. right-hand panel of Fig. 9. Therefore neglecting the sources from the second telescope in high-density regions seems to be a valid assumption. For the GDR2all sample the situation is similar and will even improve in future data releases with longer observational baselines. The shown plot and others can be recreated using Tutorial 6 in Rybizki & Drimmel (2018).

\(^{22}\)https://www.cosmos.esa.int/web/gaia/scanning-law-pointings

This paper has been typeset from a \TeX/Bibtex file prepared by the author.