Threading the needle: constraining the stellar content and dynamics of the Galactic Centre with hypervelocity stars

Verberne, S.

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

Work published in **Sill Verberne**, Sergey E. Koposov, Elena Maria Rossi, Tommaso Marchetti, Konrad Kuijken, Zephyr Penoyre, 2024, *Astronomy & Astrophysics*, 684, A29. Reprinted here in its entirety.

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], $\log g$, $T_{\rm eff}$, 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 ~ 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/`

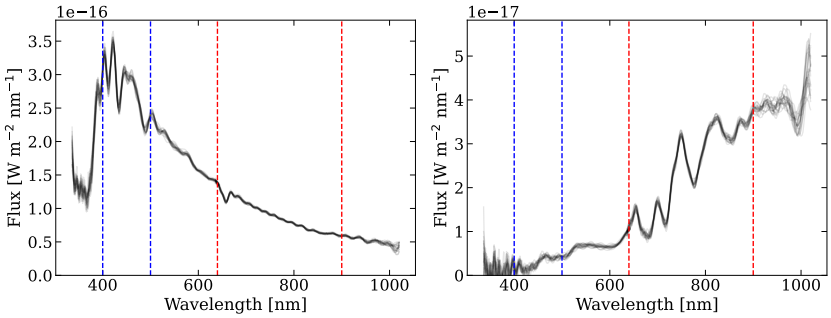

Figure 2.1: Example of two sampled *Gaia* XP spectra, in black. The uncertainties in the spectral coefficients are sampled over to indicate the uncertainties in the sampled spectra. The dashed blue and red lines indicate the BP and RP spectral ranges used in the fitting procedure, respectively. On the 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

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

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 $\mathbf{M}(T_{\rm eff},\ \log g,\ [\mathrm{Fe/H}],\ v_{\rm r},\ \mathrm{E(B-V)})$, where $T_{\rm eff}$, $\log g$, and [Fe/H] correspond to the effective temperature, surface gravity, and metallicity from the PHOENIX spectral library, respectively (Husser et al. 2013), $v_{\rm r}$ is the radial velocity, and E(B−V) the extinction along the line of sight. For the PHOENIX models, we only considered atmospheres with $[\alpha/\mathrm{H}] = 0$ 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 $\sigma = \mathrm{FWHM}/\left(2\sqrt{2\ln 2}\right)$. 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 $\mathrm{T_{eff}}$, $\log g$, 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 $\mathrm{BP}-\mathrm{RP}$ colour of a source and the extinction.

The initial guess for $T_{\rm eff}$ 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 $\pm$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

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

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

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

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 $\sim$ 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 $T_{\rm eff}$ outside grid range	15 041 788
E(B−V) ≥ 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 $T_{\rm eff}$ 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

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

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 BP − 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 ΔE(B−V) = 0.1. Additionally, we used two bins in $\log g$ for sources with BP − RP ≥ 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

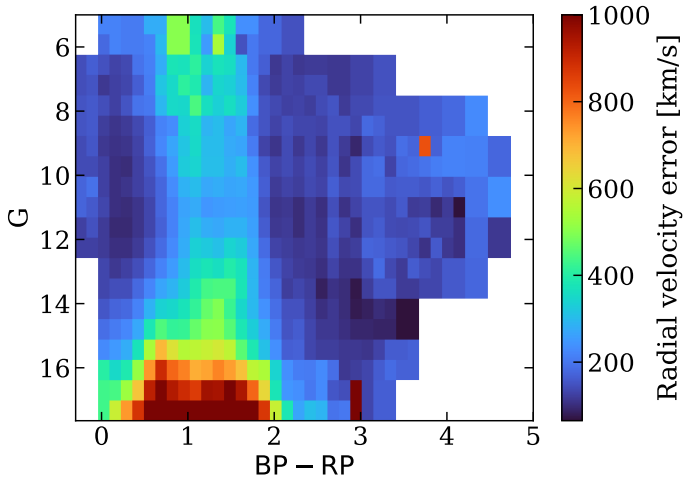

Figure 2.2: Median calibrated uncertainties as a function of colour and apparent magnitude for sources with E(B−V) < 0.1.

et al. 2013). We used flat priors throughout, except for $\sigma_{\rm out}$, 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 $\sim$ 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 $\geq$ 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.

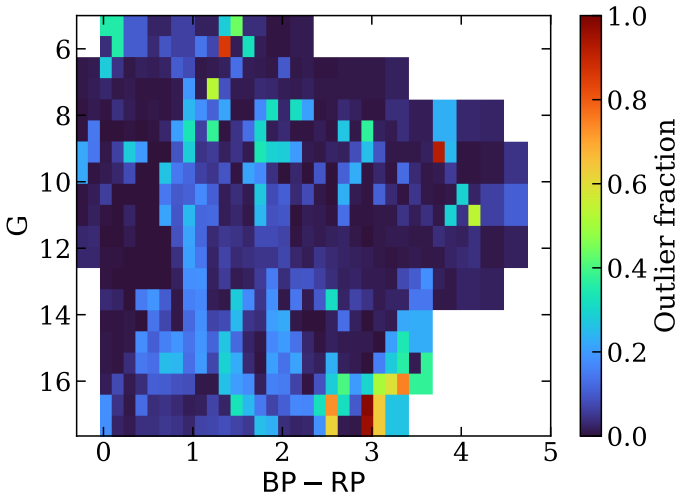

Figure 2.3: 2D histogram of the outlier fraction as a function of colour and apparent magnitude for sources with E(B−V) < 0.1.

We observe higher uncertainties for $1 \lesssim \mathrm{BP} - \mathrm{RP} \lesssim 2$ 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 $1 \lesssim \mathrm{BP}-\mathrm{RP} \lesssim 2$ 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 ($\mathrm{BP} - \mathrm{RP} \lesssim 0.7$) or red ($\mathrm{BP} - \mathrm{RP} \gtrsim 2$).

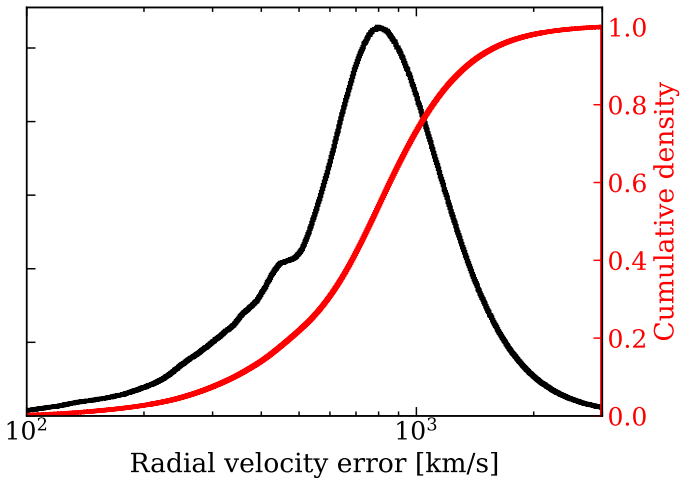

Figure 2.4: Histogram of the calibrated radial velocity uncertainties (in black) and the cumulative distribution (in red).

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 $\Delta \mathrm{v_r}/\sigma_{\mathrm{vr}} > 3$, with $\Delta \mathrm{v_r}$ 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

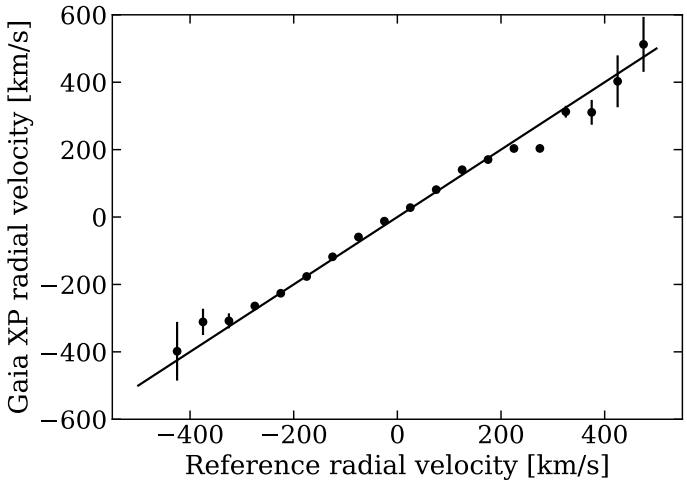

Figure 2.5: Median of the calibrated binned XP radial velocities as a function of the reference radial velocity measurements. These sources have an outlier fraction below 0.2, `bad_measurement` < 0.1, and calibrated uncertainty below 300 km/s. The solid line is the bisection.

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

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

where the overline indicates that the mean is taken, v_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-`

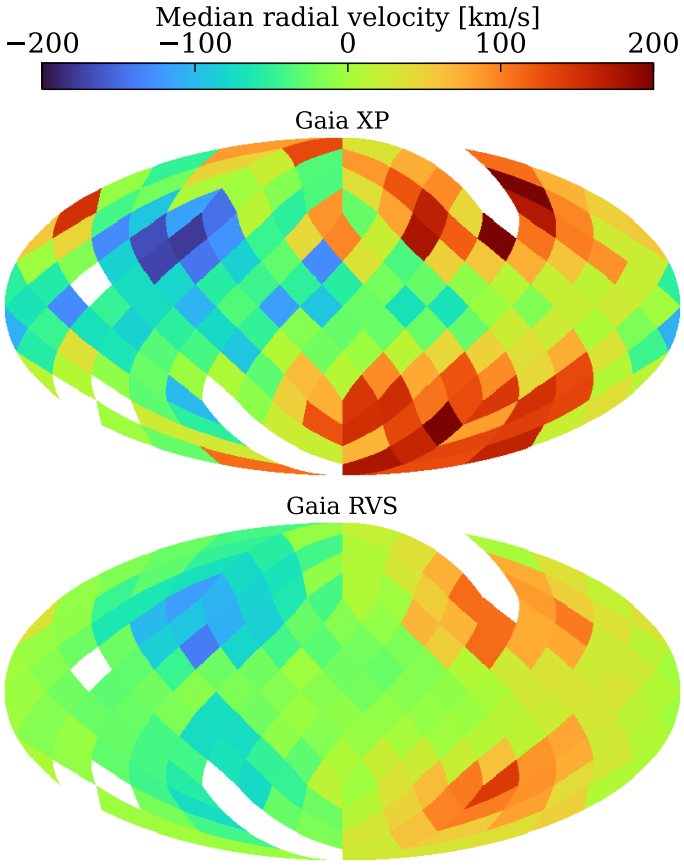

Figure 2.6: Median XP radial velocity as a function of sky position in Galactic coordinates (left) and the same but using *Gaia* RVS reference measurements (right). To highlight halo stars, we only show low metallicity stars ([Fe/H] ≤ -1). This increases the radial velocity amplitude as a function of 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.

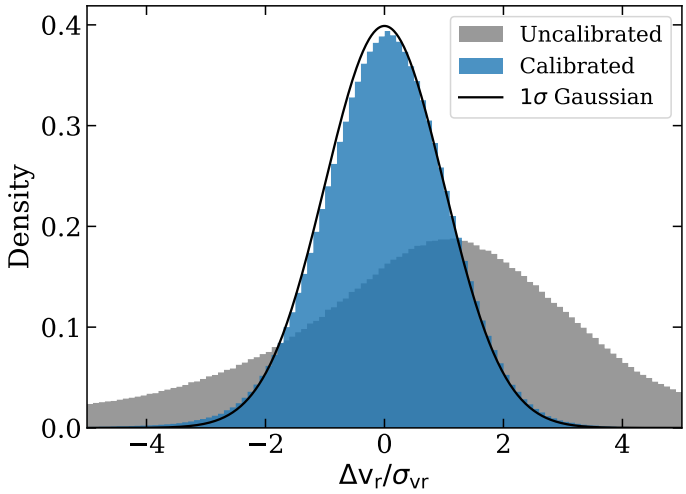

Figure 2.7: Difference in radial velocity over the uncertainty of both our calibrated and uncalibrated results compared to reference measurements. The solid line is a Gaussian distribution with a standard deviation of 1. The sample used to make this figure has calibrated radial velocity uncertainties of < 300 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`≤ 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

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

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 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`	$T_{\rm eff}$ 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

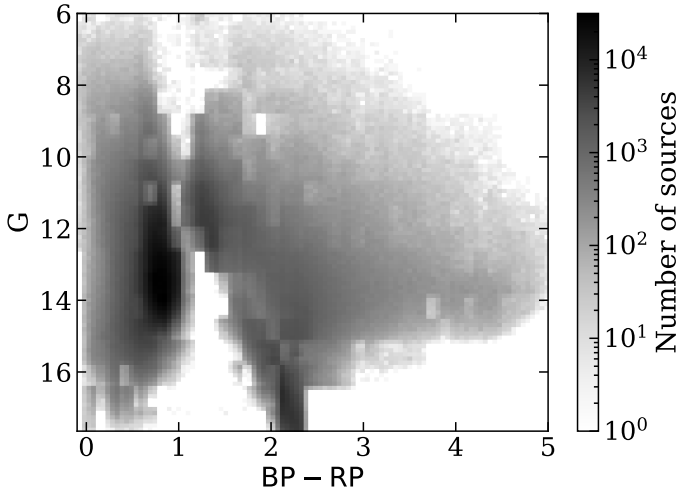

Figure 2.8: Colour-magnitude diagram for the sources in our Main Catalogue.

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

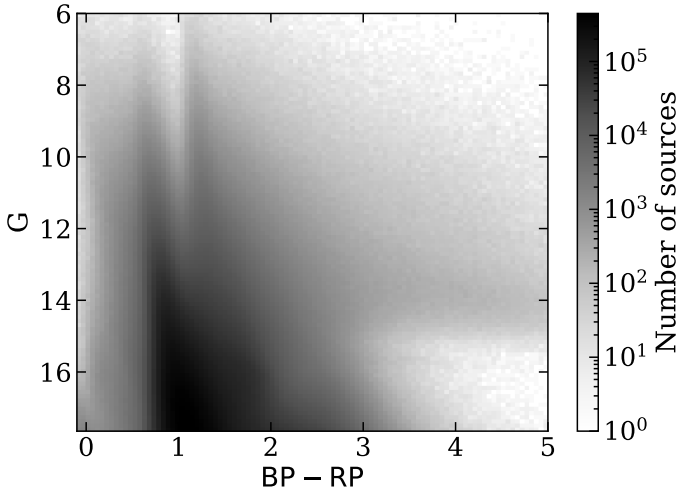

Figure 2.9: Colour-magnitude diagram for the sources in the Extended Catalogue.

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 $\mathrm{km\,s^{-1}}$ and a heliocentric radial velocity of 1017 ± 2.7 $\mathrm{km\,s^{-1}}$ (Koposov et al. 2020). The calibrated radial velocity we measure is 799 ± 273 $\mathrm{km\,s^{-1}}$, 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 $\mathrm{km\,s^{-1}}$ (i.e. $v_\mathrm{r} > 300 + 3 \cdot \sigma_\mathrm{vr}$ $\mathrm{km\,s^{-1}}$) and plot those for which reference radial velocity measurements were available. Although the distribution of our selection still peaks at 0 $\mathrm{km\,s^{-1}}$, 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-

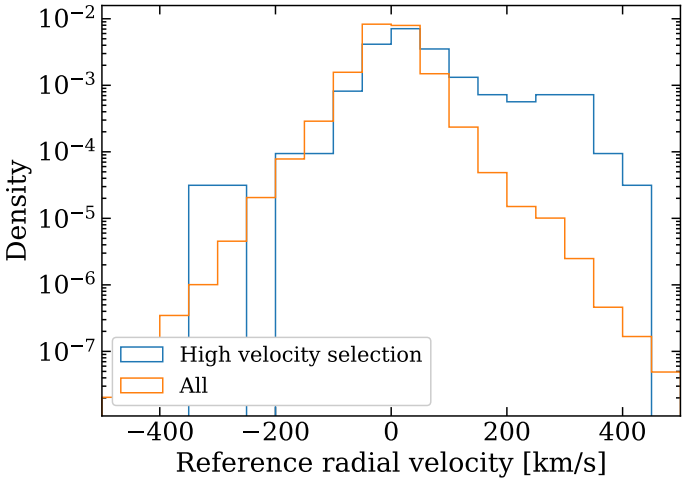

Figure 2.10: Density of the reference radial velocity distribution of all sources against those selected by $v_r > 300 + 3 \cdot \sigma_{vr}$ km s^{-1} in our Main Catalogue and for which reference radial velocity measurements are available.

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 $v_r > 300 + 3 \cdot \sigma_{vr}$ 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 $T_{\rm eff}$. 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 $T_{\rm eff}$ 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 $\sim$ 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 $\sim$ 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 $\sim$ 153M and $\sim$ 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 $\sim$ 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 $G_{RVS} \sim 16$ 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

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

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 ($\mathrm{BP}-\mathrm{RP} \gtrsim 2$) and blue ($\mathrm{BP}-\mathrm{RP} \lesssim 0.7$) 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 $\sim$ 125M sources for which we have obtained a radial velocity measurement. This constitutes $\sim$ 84% of sources with *Gaia* XP spectra and $\mathrm{E(B-V)} < 0.5$. The extended catalogue, however, contains a significant number of unreliable measurements and should therefore only be used with caution.

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

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

The authors thank the anonymous referee for their insightful comments and suggestions on this work. In addition, the authors would like to thank the attendees at the *Gaia* XPloration workshop for their input and enthusiasm. Special thanks goes to Francesca De Angeli, Anthony Brown, and Vasily Belokurov for their support, helpful insight, and discussions. We would also like to thank Anthony Brown for his feedback on a first draft of this manuscript.

EMR acknowledges support from European Research Council (ERC) grant number: 101002511/project acronym: VEGA_P. TM acknowledges a European Southern Observatory (ESO) fellowship. This work has made use of data from the European Space Agency (ESA) mission *Gaia* (`https://www.cosmos.esa.int/gaia`), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, `https://www.cosmos.esa.int/web/gaia/dpac/consortium`). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. This project was developed in part at the 2023 *Gaia* XPloration, hosted by the Institute of Astronomy, Cambridge University. This paper made use of the Whole Sky Database (wsdb) created and maintained by Sergey Koposov at the Institute of Astronomy, Cambridge with financial support from the Science & Technology Facilities Council (STFC) and the European Research Council (ERC). This work was performed using the ALICE compute resources provided by Leiden University. This research or product makes use of public auxiliary data provided by ESA/Gaia/DPAC/CU5 and prepared by Carine Babusiaux.

Software: `NumPy` (Harris et al. 2020), `SciPy` (Virtanen et al. 2020), `Matplotlib` (Hunter 2007), `Astropy` (Astropy Collaboration et al. 2013, 2018, 2022), `emcee` (Foreman-Mackey et al. 2013), `Numba` (Lam et al. 2015), `dustmaps` (Green 2018a), `GaiaXPy`, `SpectRes` (Carnall 2017), `extinction`[6], `healpy` (Górski et al. 2005; Zonca et al. 2019), `corner` (Foreman-Mackey 2016), `scikit-learn` (Pedregosa et al. 2011).

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

Appendix

2.A $T_{\rm eff}$ initial guess

For computational efficiency, we made an initial guess for the $T_{\rm eff}$ 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 $T_{\rm eff}$ 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 $T_{\rm eff}$ grid range, we used an empirical exponential fit to the relation between BP − RP and $T_{\rm eff}$. 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

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

where c denotes the extinction corrected BP − RP colour. If the initial guess for $T_{\rm eff}$ 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 $T_{\rm eff}$ 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 $\Delta T_{\rm eff}$. 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`

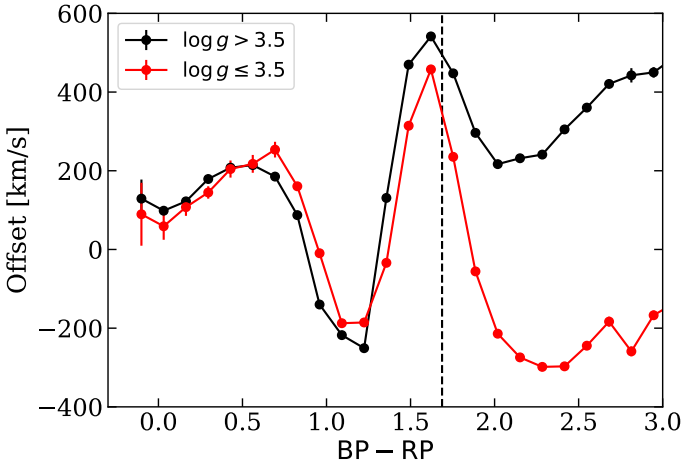

Figure 2.11: Offset between the uncalibrated *Gaia* XP radial velocities and reference ones as a function of BP − RP colour. We split the dwarfs and the giants using the $\log g$ measurements from Zhang et al. (2023). The dashed vertical line shows the limit above which we used two bins in $\log g$ for our calibration model. The figure only includes sources with E(B−V) < 0.1 and 12 < G < 13.

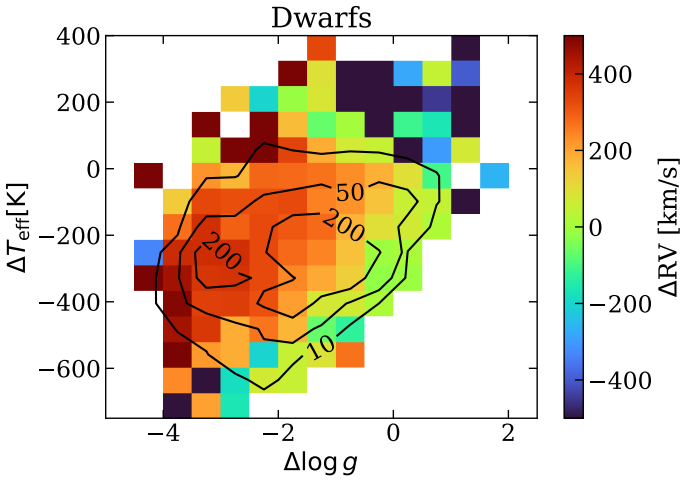

Figure 2.12: Radial velocity offset for dwarfs ($\log g > 3.5$ according to Zhang et al. (2023)) as a function of the difference in the best-fit $\log g$ and $T_{\rm eff}$ we obtain versus those from APOGEE. The contour lines give the underlying source density. The stars in the sample have 10 < G < 16, 2 < BP − RP < 2.5, and E(B−V) < 0.1.

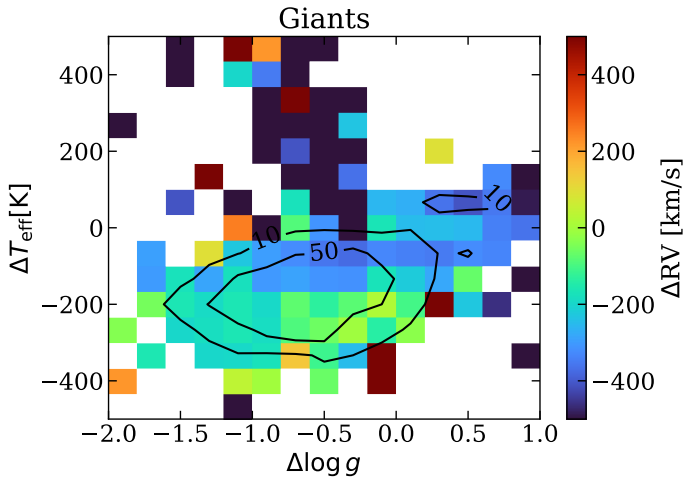

Figure 2.13: Same as Fig. 2.12, but now only for giants.

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 $\mathrm{BP}-\mathrm{RP} \sim 1.5$. 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.

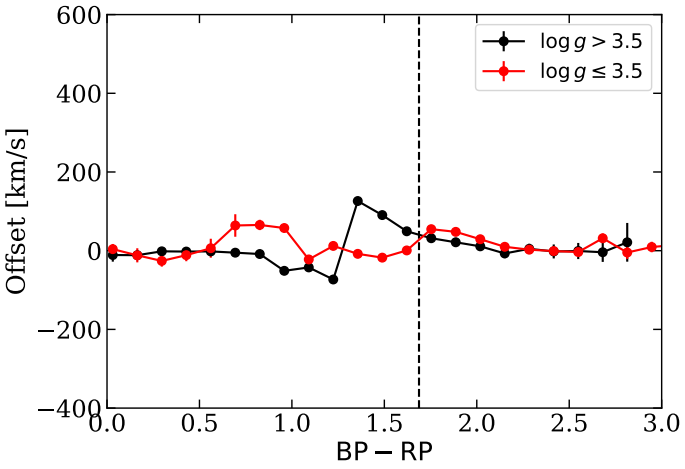

Figure 2.14: Same as Fig. 2.11, but after calibration is applied to the bin $11.9575 < \mathrm{G} < 12.59$.

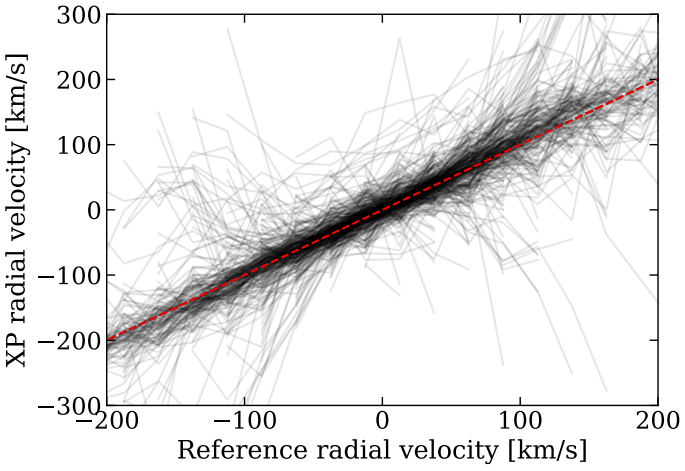

Figure 2.15: Median XP radial velocity as a function of bins in reference radial velocity. The dashed red line gives the bisection and the transparent black lines the data for individual bins in the calibration.

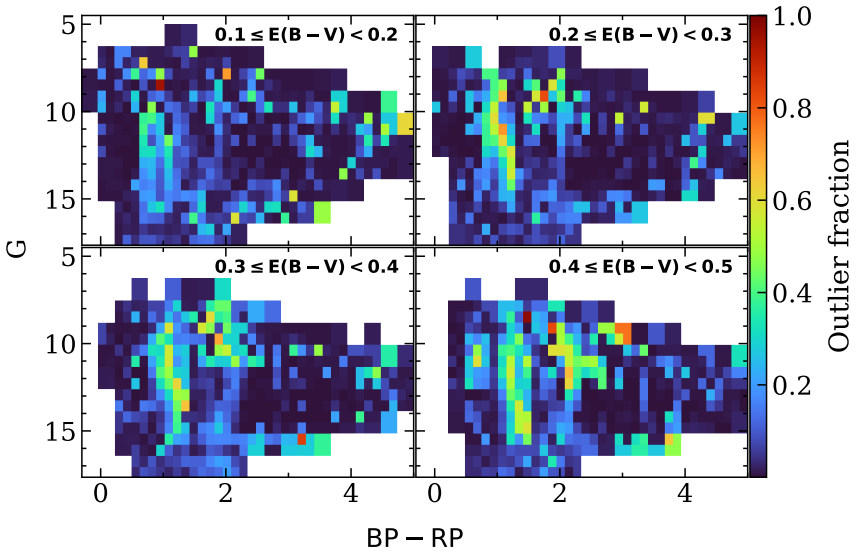

Figure 2.16: Outlier fractions as a function of colour and magnitude for the remaining E(B−V) bins not shown in the main text.

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.

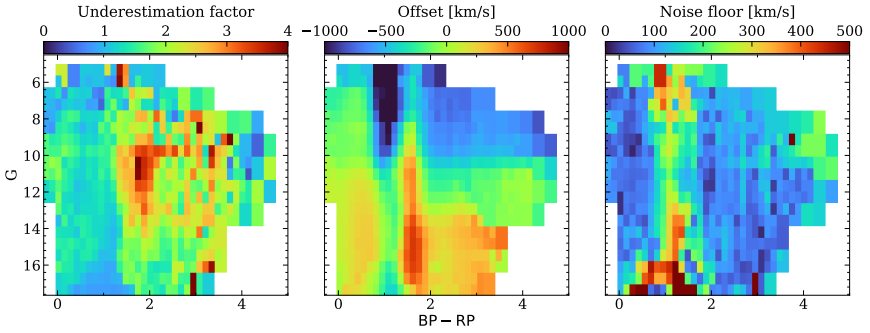

Figure 2.17: From left to right: Underestimation factor, offset, and noise floor for sources with E(B−V) < 0.1.

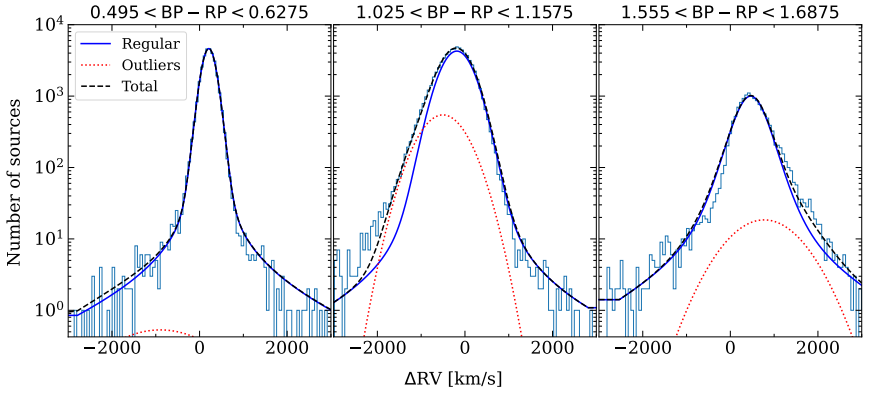

Figure 2.18: Histograms of the radial velocity difference between the raw XP measurement and the reference values. The three lines show the parameters found by the fit of the predicted distributions of the regular, outlier, and combined populations. The different panels show different bins in BP − RP colour, as indicated above the panels.

Instead a more elaborate strategy would have to be employed to correct for the remaining systematics.