169. A billion radial velocities

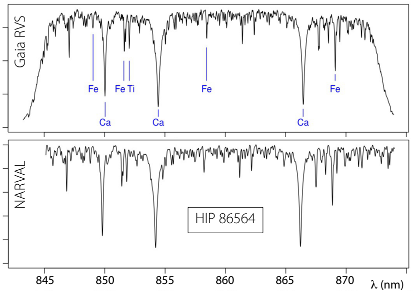
I have described the acquisition and use of Gaia's radial velocities in several previous essays: why they are important (essay 8), the resolution and wavelength range chosen for the radial velocity spectrometer (85), their acquisition on board (86), and a summary of their content in Data Release 3 (87).

Let me summarise a few key points. The star's radial velocity (i.e. *along* the line-of-sight) is required, together with the projected motion on the sky given by its proper motion, to determine the star's full 3-d space motion. This is important, and often crucial, for many kinematic and dynamical studies (e.g. Wilkinson et al., 2005). They also contribute, deeply, to orbital studies of binaries.

However, astrometry *per se* cannot determine this radial component of motion. Accordingly, a dedicated radial velocity spectrometer (RVS) was included in Gaia's instrument focal plane. RVS is an integral-field spectrograph, covering the wavelength range 845–872 nm at a resolution $R = \lambda/\Delta\lambda \sim 11\,700$. This combination was carefully chosen to cover the important Ca II infrared triplet, the Paschen series of hydrogen in early-type stars, and certain diffuse interstellar bands (essay 85).

Due to practical constraints, RVS obtains spectra only for sources $G_{\rm RVS} \lesssim 16$ mag, and with 43% fewer focal plane transits than the astrometric and photometric fields, typically ~80 over the 10-year mission.

As an example, the figure shows the $V = 6.7$ mag K5 star HIP 86564 from a single 4.4 sec Gaia exposure (top), and with the NARVAL spectrograph at the Observatoire du Pic du Midi (at the same spectral resolution, bottom).



Cropper et al. (2018)

The processing of the RVS data is carried out, on the ground, within an extensive and dedicated processing 'pipeline', described for Gaia DR2 by Sartoretti et al. (2018), and for Gaia DR3 by Katz et al. (2023).

While Gaia is today into its tenth year of observations, only the first 34 months (July 2014–May 2017) of data have been processed and released by the Gaia Data Processing and Analysis Consortium (DPAC), as Gaia Data Release 3 on 13 June 2022.

Along with 1.8 billion sources with astrometry, DR3 provides radial velocities for 33 million sources down to $G_{\rm RVS} \sim 14$ mag. When DR4 is released in 2025, based on 66 months of data, RVS results for the 100 million sources down to $G_{\rm RVS} \sim 16$ mag should be available.

Let me put these numbers in context. When the Hipparcos catalogue of 120 000 stars was released in 1997, radial velocities were known for just 20 000. Surveys by Coravel, amongst others (often focused on exoplanets) later measured several thousand more.

Subsequently, and on a much larger scale, and in the northern hemisphere between 2005–2010, the Sloan Digital Sky Survey's SEGUE extension obtained spectra for 240 000 stars, with typical radial velocity accuracies of $10\,\mathrm{km\,s^{-1}}$ (Yanny et al., 2009). SEGUE–2 (2008–2009) observed a further 120 000 (Rockosi et al., 2022).

Complementing SEGUE in the south, RAVE (RAdial Velocity Experiment) was a multi-fiber spectroscopic survey using the 1.2-metre AAO–UK Schmidt Telescope. Conducted between 2004–2013, and partly motivated by the prospects of Gaia, RAVE acquired some 574 000 spectra for around 483 000 stars (Kunder et al., 2017).

The LAMOST–II medium-resolution ($R \sim 7500$) spectral survey measured 1 597 675 spectra for 281 515 stars (Wang et al., 2019), achieving radial velocity accuracies of around $1\,\mathrm{km\,s^{-1}}$. The (AAO–HERMES) GALAH+ survey includes 584 015 dwarfs and giants in the magnitude range $G = 11 - 14$ (Zwitter et al., 2021).

And the merged and homogeneous 'Survey of Surveys' contains almost 11 million stars with radial velocity precision in the range 0.05–$1.50\,\mathrm{km\,s^{-1}}$, of which half are exclusively from Gaia (Tsantaki et al., 2022).

THIS BACKGROUND sets the scene for some work recently reported by Verberne et al. (2023). They turned their attention to the possibility of deriving radial velocities from the low-resolution BP/RP spectra that Gaia measures for *all* sources. Unlike the RVS which (aside from data rate limitations) is photon-starved at $G \gtrsim 16$ mag, the low-resolution BP/RP spectra are acquired for all sources crossing the main focal plane.

Currently, DR3 (June 2022) provides the full astrometric solution for 1.5 billion sources, while the low-resolution spectra have been published for 'only' some 220 million. The BP/RP spectra for Gaia's full 2 billion or more sources will be made available with DR4 in 2025.

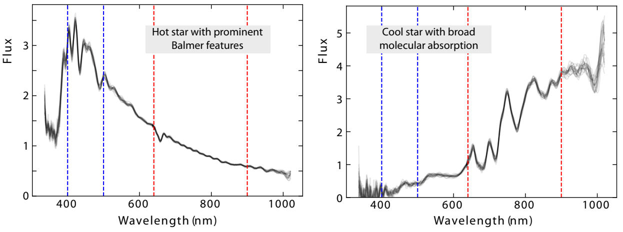
THE LOW-RESOLUTION BP/RP spectra provide astrophysical information on each astrometric source. The design, including wavelength range and resolution, was optimised to provide information needed to determine the astrophysical parameters of each source, such as $T_{\rm eff}$, log(g), [M/H], and extinction, across a wide range of spectral type and luminosity class. Within the DPAC Consortium (CU8), the 'astrophysical parameters inference system' pipeline (Apsis, Bailer-Jones et al., 2013; Recio-Blanco et al., 2023) runs 13 modules, using different combinations of data and models, to produce astrophysical parameters for stars, galaxies and quasars.

Amongst the modules identifying and classifying the extragalactic sources, the 'quasar classifier' module (QSOC) estimates the quasar redshifts using a χ^2 approach, in which the BP and RP spectra are compared to a composite quasar spectrum over trial redshifts in the range $z = 0-6$ (Bailer-Jones et al., 2023; §2.2). A similar approach estimates the redshifts of galaxies using the 'unresolved galaxy classifier' module (UGC). An animation illustrating this procedure is shown here.

As a result, 6.4 million quasars and 1.4 million galaxies have redshifts given in DR3 (Delchambre et al., 2023).

A SIMILAR APPROACH has been adopted for determining stellar radial velocities by Verberne et al. (2023). The principle is straightforward, although of course the details are not: the radial velocity (and its estimated error) is obtained by fitting the BP/RP spectra to a wide range of models based on a grid of synthetic spectra.

The figure shows two example Gaia source spectra, with the blue and red dashed lines indicating the BP and RP spectral ranges used in the fitting procedure.

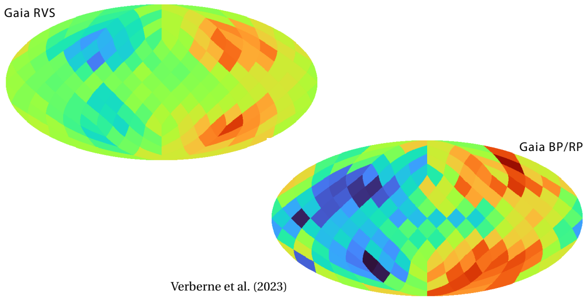


Verberne et al. (2023)

THEIR MOST reliable subset comprises 6.4 million sources with uncertainties <300 km s^{-1}, around one quarter of which have no radial velocity yet available from RVS. They also constructed an extended catalogue with all 125 million sources for which they could obtain a valid radial velocity estimate.

The figure below shows: (top) the median radial velocity for low-metallicity stars, [Fe/H] < –1, from the BP/RP spectra as a function of sky position in Galactic coordinates, and (bottom) the median radial velocity from RVS. Colour coding ranges from −200 km s^{-1} (dark blue), through 0 (green), to +200 km s^{-1} (dark red). The dipole component attributable to the solar motion is evident in both maps.

The typical uncertainties are, of course, far higher than those obtained through precision spectroscopy. Their median uncertainty on the calibrated BP/RP radial velocity measurements is ~770 km s^{-1}, but extends to below 100 km s^{-1} (their Fig. 4). Sources with the best accuracy tend to be either very blue (BP–RP $\lesssim$ 0.7) or very red (BP–RP $\gtrsim$ 2). The scientific value of such imprecise values remains to be seen, but might include searches for black hole binaries or hypervelocity stars.



Verberne et al. (2023)

LET ME summarise. Gaia DR3 has provided radial velocities for 33 million sources to $G_{\rm RVS} \sim 14$, with the 66-month based DR4 expected to deliver around 100 million to $G_{\rm RVS} \sim 16$ mag in 2025. Very coarse measurements from BP/RP are now also available for 125 million.

A VERY DIFFERENT approach to estimating the missing radial velocities is to infer them from the 5 other astrometric parameters, as first implemented using neural networks by Dropulic et al. (2021). Their 6 million source training set was later used to infer radial velocities for a further 92 million, which they applied to stars in the Enceladus merger using EDR3 (Dropulic et al., 2023).

Naik & Widmark (2022) used a Bayesian neural network approach to predict the values for 16 million stars in EDR3. These predictions were tested by Naik & Widmark (2024), and DR3 astrometry in turn used to predict the missing values for 185 million stars to $G < 17.5$. They estimate that the predictions are reliable, to 25–30 km s^{-1}, for stars within 7 kpc of the Sun.