Gaia Data Release 2

Gaia Data Release 2

Processing the spectroscopic data

P. Sartoretti[1], D. Katz[1], M. Cropper[2], P. Panuzzo[1], G. M. Seabroke[2], Y. Viala[1], K. Benson[2], R. Blomme[3],
G. Jasniewicz[4], A. Jean-Antoine[5], H. Huckle[2], M. Smith[2], S. Baker[2], F. Crifo[1], Y. Damerdji[6,7], M. David[8],
C. Dolding[2], Y. Frémat[3], E. Gosset[7,9], A. Guerrier[10], L. P. Guy[11], R. Haigron[1], K. Janßen[12], O. Marchal[1], G. Plum[1],
C. Soubiran[13], F. Thévenin[14], M. Ajaj[1], C. Allende Prieto[2,15,16], C. Babusiaux[1,17], S. Boudreault[2,18], L. Chemin[13,19],
C. Delle Luche[1,10], C. Fabre[20], A. Gueguen[1,21], N. C. Hambly[22], Y. Lasne[10], F. Meynadier[1,23],
F. Pailler[5], C. Panem[5], F. Riclet[5], F. Royer[1], G. Tauran[5], C. Zurbach[4], T. Zwitter[24], F. Arenou[1], A. Gomez[1],
V. Lemaitre[10], N. Leclerc[1], T. Morel[7], U. Munari[25], C. Turon[1], and M. Žerjal[24,26]

(Affiliations can be found after the references)

ABSTRACT

Context. The *Gaia* Data Release 2 (DR2) contains the first release of radial velocities complementing the kinematic data of a sample of about 7 million relatively bright, late-type stars.
Aims. This paper provides a detailed description of the *Gaia* spectroscopic data processing pipeline, and of the approach adopted to derive the radial velocities presented in DR2 .
Methods. The pipeline must perform four main tasks: (i) clean and reduce the spectra observed with the Radial Velocity Spectrometer (RVS); (ii) calibrate the RVS instrument, including wavelength, straylight, line-spread function, bias non-uniformity, and photometric zeropoint; (iii) extract the radial velocities; and (iv) verify the accuracy and precision of the results. The radial velocity of a star is obtained through a fit of the RVS spectrum relative to an appropriate synthetic template spectrum. An additional task of the spectroscopic pipeline was to provide first-order estimates of the stellar atmospheric parameters required to select such template spectra. We describe the pipeline features and present the detailed calibration algorithms and software solutions we used to produce the radial velocities published in DR2 .
Results. The spectroscopic processing pipeline produced median radial velocities for *Gaia* stars with narrow-band near-IR magnitude $G_{RVS} \leq 12$ (i.e. brighter than $V \sim 13$). Stars identified as double-lined spectroscopic binaries were removed from the pipeline, while variable stars, single-lined, and non-detected double-lined spectroscopic binaries were treated as single stars. The scatter in radial velocity among different observations of a same star, also published in *Gaia* DR2, provides information about radial velocity variability. For the hottest ($T_{eff} \geq 7000$ K) and coolest ($T_{eff} \leq 3500$ K) stars, the accuracy and precision of the stellar parameter estimates are not sufficient to allow selection of appropriate templates. The radial velocities obtained for these stars were removed from DR2 . The pipeline also provides a first-order estimate of the performance obtained. The overall accuracy of radial velocity measurements is around ~200–300 m s^{-1}, and the overall precision is ~1 km s^{-1}; it reaches ~200 m s^{-1} for the brightest stars.

Key words. techniques: spectroscopic – catalogs – techniques: radial velocities – surveys – methods: data analysis

1. Introduction

The *Gaia* mission (Gaia Collaboration 2016) will provide detailed phase space data for distant stars in the Milky Way galaxy, in addition to astrometry, including radial velocity for many stars (~150 million). For logistical reasons, the *Gaia* Radial Velocity Spectrometer (RVS; Cropper et al. 2018) adopts the wavelength range 845–872 nm, which includes the strong Ca II triplet lines and is useful for cross-correlation spectroscopy of Galactic stars of most spectral types, except for the earliest spectral types. This paper documents the reduction process and the method used to convert raw observed stellar spectra into the radial velocities presented in the *Gaia* Data Release 2 (hereafter DR2; Gaia Collaboration 2018).

With its two telescopes, *Gaia* continuously scans the sky and collects data of any source detected by the onboard system. Sources are first observed by the astrometric and photometric instruments, when their magnitude in the RVS instrument is estimated. If a source transiting through one of the four RVS rows (as depicted in Fig. 1) is bright enough ($G_{RVS} < 16.2$, see Sect. 2.1), and if the onboard limit on the number of data that can be obtained simultaneously (corresponding to a source density of ~35 000 sources deg^{-2}) is not already reached, a spectrum of the source is recorded by each of the three CCDs on the row.

Each source will be observed many times during the nominal 60 months of the mission, the expected number of transits per star in the RVS being on average around 40. In the RVS, starlight is dispersed over ~1000 pixels, while the exposure time on the *Gaia* CCDs is fixed at 4.4 s by the scanning requirements. The resulting low signal-to-noise ratio (S/N) per pixel for stars near the limiting magnitude of $G_{RVS} \sim 16$ implies that the combination of many transit spectra will be necessary over the entire mission lifetime for the radial velocities of these stars to be measured. In fact, intermediate data releases such as DR2 are only
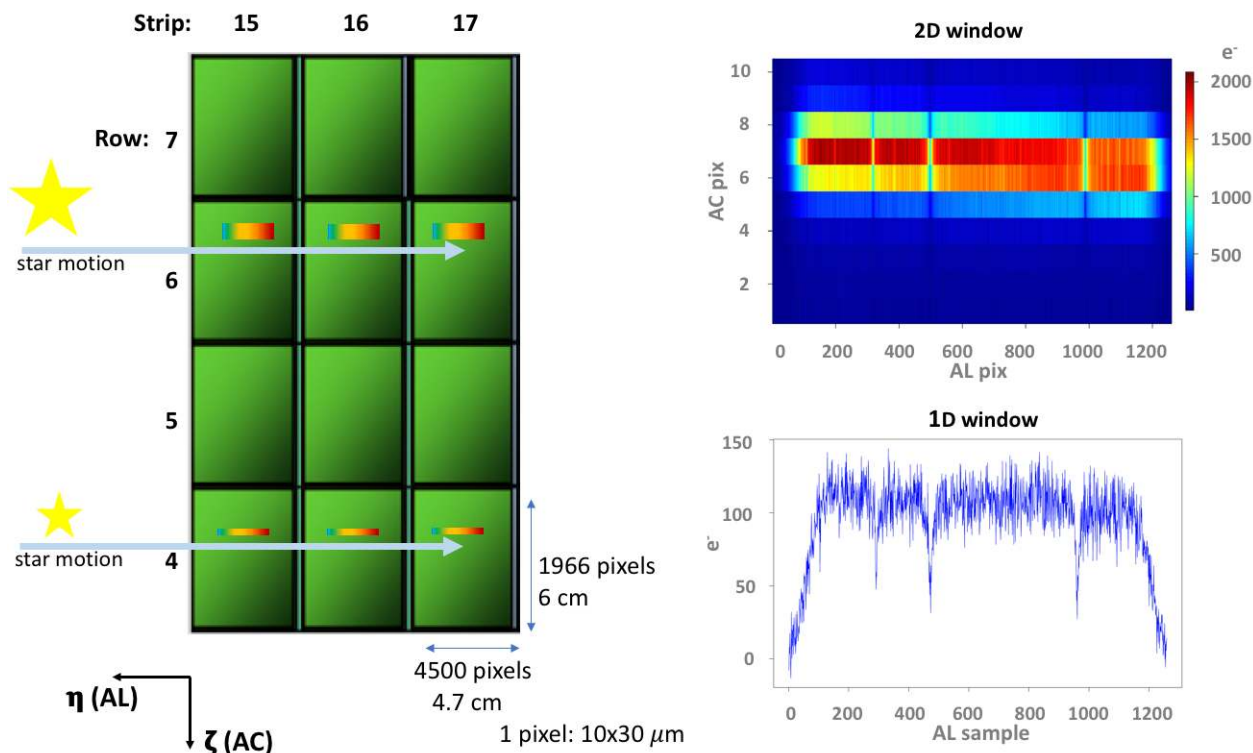
Fig. 1. RVS focal plane. For the complete *Gaia* focal plane, see Gaia Collaboration (2016, their Fig. 4). *Left panel:* 12 CCDs of the RVS focal plane laid out in three strips (we use the standard *Gaia* nomenclature and refer to columns as "strips") and four rows. The exposure time is fixed at 4.4 seconds per CCD (in TDI mode). The star images move in the along-scan (AL) direction (as indicated by horizontal arrows). During each transit, the star crosses all three CCDs on the row, and a spectrum is acquired in each of the three corresponding observational windows centred on the star. If the star has onboard $G_{RVS} \leq 7$, the telemetered window is 2D, of size 1260- or 1296-AL by 10-AC pixels (*upper right panel*). For fainter stars, the 10-AC pixels are summed during read-out to produce a 1D window (*lower right panel*). The FoV of both *Gaia* telescopes are projected onto the focal plane. The orientation of the field angles η and ζ is indicated in the lower left corner. The origin of the axes is in the astrometric focal plane and is different for the two FoVs.

preliminary. We note that in addition to RVS data for always fainter stars, each future release will include a complete reprocessing of data from the beginning of the mission, with improved calibrations and algorithms.

The RVS spectra processed for DR2 were collected during the first 22 months since the start of nominal operations on 25 July 2014. Only stars brighter than $G_{RVS} \sim 12$ have been processed for DR2, corresponding to ~5% of the spectra acquired by the RVS in this period. Still, the number of spectra treated exceeds 280 million. It is a major challenge to process such large amounts of data, and in addition to the reliability and robustness of the scientific processing software, the RVS pipeline has to cope with technical issues, such as computational time, data storage, backup, and I/O management for the scientific software and database access.

In this paper, we describe the data processing behind the radial velocities presented in DR2 and the performance achieved. A companion paper (Katz et al. 2018) presents a posteriori checks of the DR2 radial velocities and the dependence of the accuracy and precision of velocity measurements on stellar properties.

2. Overview and limitations of the spectroscopic pipeline

2.1. G_{RVS} magnitude

G_{RVS} is a narrow-band near-IR (NIR) magnitude, whose nominal passband (845–872 nm) is described in Jordi et al. (2010). The

impracticability of estimating uniformly the G_{RVS} magnitude for on-the-fly data acquisition, spectroscopic pipeline selection, and from the flux in the spectrum has led to different definitions of this quantity.

– The onboard G_{RVS} magnitude is estimated by the onboard software. It is used to decide whether to allocate an RVS window to a star (stars fainter than $G_{RVS} = 16.2$ do not get an observation window), as well as which window type to allocate (2D if $G_{RVS} \leq 7$, 1D otherwise; see Fig. 1). The onboard G_{RVS} magnitude is estimated using the flux in the portion of the red photometer (RP) spectrum covering the RVS wavelength range, or, when the RP spectrum is saturated, the flux in the astrometric image. The G_{RVS} magnitude is derived in each of these two cases using ad hoc formulae whose parameters were calibrated during commissioning.

– The onboard G_{RVS} magnitude described above is contaminated by straylight and instrumental effects. Another estimate of the G_{RVS} magnitude was therefore used to select the stars to process through the DR2 spectroscopic pipeline: the external G_{RVS} magnitude, noted G_{RVS}^{ext}. This is taken to be the magnitude listed in the initial Gaia source list (IGSL) published before the mission started (Smart & Nicastro 2014). For the 8% of stars detected by *Gaia* not found in the original IGSL, we take the G_{RVS}^{ext} magnitude to be the onboard one. Stars brighter than $G_{RVS}^{ext} = 12$ are processed through the DR2 spectroscopic pipeline, the brightest ones ($G_{RVS}^{ext} \leq 9$) being used for wavelength calibration. For reference, for most (~85%) stars in the IGSL, the G_{RVS}^{ext}

magnitude was estimated from Tycho-2 B_T and V_T and GSC23 B_J and R_F (Lasker et al. 2008) magnitudes using the formulae

$$G_{\rm RVS}^{\rm ext} = R_F - 0.0974 - 0.4830(B_J - R_F) - 0.0184(B_J - R_F)^2 - 0.0178(B_J - R_F)^3,$$
$$G_{\rm RVS}^{\rm ext} = V_T - 0.1313 - 1.3422(B_T - V_T) - 0.09316(B_T - V_T)^2 - 0.0663(B_T - V_T)^3.$$

This estimate was revealed to be imprecise, especially in case of red colour. A better estimate is obtained using V and I (see Sect. 4.6).

– Another quantity is computed internally by the pipeline directly from the spectra. This internal $G_{\rm RVS}$ magnitude, noted $G_{\rm RVS}^{\rm int}$, is computed based on the flux integrated in the RVS spectrum and a calibrated reference magnitude ($G_{\rm RVS}^{\rm ref}$, see Sect. 5.5) relying on V and I observations of about 113 000 Hipparcos stars (see Sects. 4.6 and 6.3).

2.2. Purpose of the spectroscopic pipeline

The radial velocity is estimated by measuring the Doppler shift of the spectral lines in the observed RVS spectrum compared to a synthetic template spectrum (at rest), selected to be as similar as possible to the observed spectrum.

The first goal of the *Gaia* spectroscopic pipeline is to measure the all-transit-combined radial velocity (V_R) for all the *Gaia* stars observed by the RVS (i.e. brighter than onboard $G_{\rm RVS} \sim 16$), for which an appropriate template could be found (the atmospheric parameters of the stars needed to find the template will be available from another pipeline). The precision depends on the spectral type, the magnitude of the star, and the number of observations. The pre-launch end-of-mission requirements on the precision of the radial velocity measurements (mission averaged, 40 transits) were set on three representative spectral types, B1V, G2V, and K1IIIMP (where MP stands for metal-poor, i.e. [Fe/H] ≈ -1.5 dex), depending on the magnitude,

– 1 km s^{-1}: B1V with $G_{\rm RVS} \leq 7.2$ ($V \sim 7$); G2V with $G_{\rm RVS} \leq 12.1$ ($V \sim 13$) and K1IIIMP with $G_{\rm RVS} \leq 12.3$ ($V \sim 13.5$);
– 15 km s^{-1}: B1V with $G_{\rm RVS} \sim 12.2$ ($V \sim 12$); G2V with $G_{\rm RVS} \sim 15.6$ ($V \sim 16.5$) and K1IIIMP with $G_{\rm RVS} \sim 15.8$ ($V \sim 17$).

The precision on the V_R measurements is expected to be low for early-type stars, which in the narrow NIR RVS wavelength range have shallow and broad spectral lines, and for late M-type stars, dominated by the molecular TiO band. For medium-temperature stars (FGK types), the precision is expected to increase towards cooler stars, where more neutral lines and more contrast with the continuum provide more information for comparisons with templates. Some RVS spectra of stars with different spectral types are shown in Cropper et al. (2018, their Fig. 17). The end-of-mission performance is modelled for several spectral types in Gaia Collaboration (2016, their Sect. 8.3[1]). With regard to systematic uncertainties, the pre-launch end-of-mission requirement was that after calibration, these should be less than 300 m s^{-1} (Cropper et al. 2018, Table 1).

Another goal of the spectroscopic pipeline is to provide the radial velocity at each transit through the RVS ($V_R^{\rm t}$) for the brightest stars to identify variable candidates, detect double-lined spectra, and estimate the velocity of the two components.

---

[1] See also: https://www.cosmos.esa.int/web/gaia/rvsperformance

The single-transit radial velocities are used by the downstream pipelines for variability and multiple-system studies and will be published in DR4. Rotational velocities will be estimated for a subset of bright stars. The all-transit-combined spectra are also produced and used by downstream pipelines to estimate stellar atmospheric parameters and abundances (Recio-Blanco et al. 2016). A subset of the combined spectra will be published in DR3.

The specific goal for DR2, with 22 months of mission data, was to provide the all-available-transit-combined (a minimum number of two transits is required) V_R for all the stars observed by the RVS brighter than $G_{\rm RVS}^{\rm ext} = 12$, for which an appropriate template could be found. The aim was to approach the end-of-mission requirements as best possible, and reach a precision close to ~1 km s^{-1}. The single-transit radial velocities, $V_R^{\rm t}$ and the combined spectra are also obtained, but not published. One problem in the DR2 pipeline was that estimates of the atmospheric parameters of observed stars, which are needed to associate synthetic templates, were not available for most stars. As a result, an additional goal of the pipeline was to provide (rough) estimates of these stellar parameters, to be able to associate appropriate template. For the hottest ($T_{\rm eff} \geq 7000$ K) and coolest ($T_{\rm eff} \leq 3500$ K) stars, however, the accuracy and precision of such estimates are not sufficient to allow selecting appropriate templates, and the radial velocities obtained for these stars have been removed from DR2. Moreover, no template is available for emission-line stars, which were removed from DR2.

Further goals of the pipeline are to automatically check whether the accuracy and precision of the computed radial velocities agree with expectations, and to self-check the behaviour of each processing step to detect any potential problem.

2.3. Pipeline overview

Figure 2 shows a flowchart illustrating the spectroscopic pipeline. We provide here an overview of the primary goals of each workflow (step) of this pipeline, which are described in detail in the following sections.

First, the `Ingestion` workflow extracts information from the input data (Sect. 3) and auxiliary data (Sect. 4) and groups this information into data packets that are processed by the downstream workflows: the position of the source and its coordinates in the focal plane (Fig. 1) are recorded, which are required for wavelength calibration; overlapping spectra are removed; the sourceId to which the spectrum belongs is identified and searched for in the IGSL to determine the associated $G_{\rm RVS}^{\rm ext}$. Stars fainter than $G_{\rm RVS}^{\rm ext} = 12$ are removed from the pipeline; the sourceId is also searched for in the auxiliary catalogues to identify standard stars (Sect. 4.1) and store their reference radial velocities, which are used in the `Calibration` workflow to fix the wavelength calibration zeropoint; the sourceId is also used to identify stars with atmospheric parameters available from the literature (Sect. 4.3), to which the appropriate synthetic spectrum is then associated (Sect. 4.4), and to identify stars with available $G_{\rm RVS}^{\rm ref}$ (Sect. 4.6) to be used as reference for the $G_{\rm RVS}^{\rm int}$ zeropoint calibration.

In the `Calibration Preparation` workflow (Sect. 6), stars brighter than $G_{\rm RVS}^{\rm ext} = 9$ are selected to be used as calibrator stars, and their spectra are prepared to be used in the `Calibration` workflow; the spectra are corrected for electronic bias and dark current; the gain is applied; the straylight background and cosmic rays are removed; the 2D windows are collapsed into 1D; a first-guess model of the wavelength calibration is applied; the atmospheric parameters (if not available from
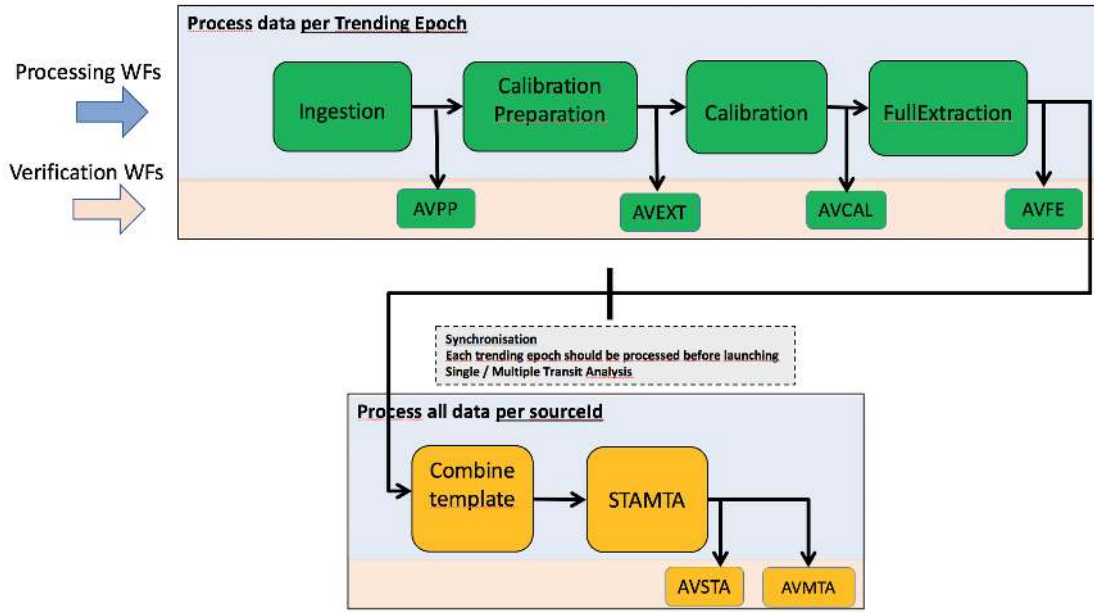
Fig. 2. Flow chart of the spectroscopic pipeline. The pipeline is composed of six processing workflows and six verification workflows. The green workflows process the data per transit (i.e. per observation). The yellow workflows process the data per source and collect the information produced upstream from each transit.

the literature) are estimated; stars with $4500 < T_{\rm eff} < 6500$ K are selected as calibrator stars for the wavelength calibration, together with standard stars; for the $G_{\rm RVS}^{\rm int}$ zeropoint calibration, only stars with available $G_{\rm RVS}^{\rm ref}$ are selected.

The `Calibration` workflow (Sect. 5) processes stars prepared for calibration (i.e. with $G_{\rm RVS}^{\rm ext} \leq 9$): the wavelength-calibration model is computed (Sect. 5.3); stars with $4500 < T_{\rm eff} < 6500$ K, along with radial-velocity standard stars, are used to fix the wavelength calibration zeropoint; the $G_{\rm RVS}$-zeropoint calibration model is also computed (Sect. 5.5), using stars with available $G_{\rm RVS}^{\rm ref}$ as photometric standards. The `Calibration` workflow also incorporates calibration parameters estimated off-line (as described in Sect. 5), to be applied to the spectra in the downstream workflows.

In the `FullExtraction` workflow, the spectra of all stars (including those with $9 < G_{\rm RVS}^{\rm ext} \leq 12$) are treated as in the `Calibration Preparation` workflow above, but this time using the calibration parameters obtained in the `Calibration` workflow. The stellar atmospheric parameters are determined, and a synthetic template is associated with each transit of a star.

In the `Combine Template` workflow (Sect. 6.6), the templates associated with different transits of a same star are combined to ensure that a single template per star is used to determine its radial velocity.

In the `STAMTA` workflow, the radial velocity of a star is computed separately for each transit, V_R^t (Sect. 7). Then, the median of the single-transit radial velocities obtained from all available transits is computed (Sect. 8) to obtain the final radial velocity of the star, V_R, published in DR2.

The purpose of the `Automated Verification` workflows (i.e. `AVPP`, `AVEXT`, `AVFE` and `AVSTA` and `AVMTA` in Fig. 2) is to check the products to be used by the downstream workflows: the data required for the verification diagnostics are processed in the pipeline to obtain the parameters to monitor; then, they are sent to an off-line software that produces diagnostic plots made available for validation via a web interface.

The entire process is managed by a software system, named SAGA, and the code is run in parallel on an Hadoop cluster system with 1100 cores and 7.5 TB memory. The processing took the equivalent of 630 000 h CPU time and needed 290 TB disc space.

2.4. Data products in DR2

The products of the spectroscopic pipeline published in *Gaia* DR2, the notation used in this paper, and their name in the *Gaia* Archive[2] are listed below.

- Median radial velocity (km s^{-1}): noted V_R; *Gaia* Archive name: *radial_velocity*. This is provided for most of the stars with $4 \leq G_{\rm RVS} \leq 12$. No variability detection is attempted, and all stars are treated as single stars. The detected SB2s (Sect. 7.5) and emission-line stars are removed from DR2.
- Radial velocity uncertainty (km s^{-1}): noted σ_{V_R}; *Gaia* Archive name: *radial_velocity_error*.
- Number of transits: the number of transits used to determine V_R, noted N; *Gaia* Archive name: *rv_nb_transits*.
- Template temperature (K): $T_{\rm eff}$ of the template used to determine V_R, noted $T_{\rm eff}^{\rm tpl}$; *Gaia* Archive name: *rv_template_teff*.
- Template surface gravity (dex): $\log g$ of the template used to determine V_R, noted $\log g^{\rm tpl}$; *Gaia* Archive name: *rv_template_logg*.
- Template metallicity (dex): [Fe/H] of the template used to determine V_R, noted [Fe/H]$^{\rm tpl}$; *Gaia* Archive name: *rv_template_fe_h*.

2.5. Limitations of the DR2 pipeline

As the other *Gaia* pipelines, the spectroscopic processing system will evolve and will be upgraded with new functionalities in the future data releases. In *Gaia* DR2, the radial velocities of stars brighter than $G_{\rm RVS} \leq 12$ are provided, and the pipeline has the

---

[2] https://gea.esac.esa.int/archive/

first-order functionality sufficient to treat these bright stars. The final V_R provided for each source is the median value of the radial velocities estimated from all source transit spectra.

In *Gaia* DR2, the expectation is to provide V_R for stars down to a magnitude $G_{RVS} \sim 14$. Then, second-order calibrations and the functionality to de-blend spectra will be included in the pipeline. The final V_R for each star will be computed with a more reliable model combining information from all single-transit spectra of the star.

In addition to the limitations from still-missing functionalities, the DR2 version of the RVS pipeline does not have access to the effective temperature (T_{eff}) published in DR2 (Andrae et al. 2018), which was produced by other processing pipelines running at the same time, nor to the G, G_{BP} and G_{RP} magnitudes (and the associated refined estimates of the G_{RVS} magnitude) from Gaia Collaboration (2018, Eq. (2)). As mentioned above, the T_{eff} needed to associate a spectral template with a star is taken from the literature when available, and otherwise estimated by the pipeline (Sect. 6.5). The G_{RVS}^{ext} magnitude used to select the stars to be processed is computed as described in Sect. 2.1.

Despite these limitations, the overall accuracy and precision of the radial velocities measured by the pipeline and published in DR2 are close to end-of-mission requirements (Sect. 9).

3. Input data

The time on board *Gaia*, the OBMT (onboard mission timeline), generated by the *Gaia* onboard clock, counts the number of six-hour spacecraft revolutions since launch. The events on board are given in OBMT (the notations REV, revolutions, are also used). The relation to convert OBMT into barycentric coordinate time (TCB) is provided by Eq. (1) in Gaia Collaboration (2018).

The spectra processed by the DR2 pipeline have been acquired by the RVS between OBMT 1078.3795 (25 July 2014) and OBMT 3750.5602 (23 May 2016).

In the focal plane (see Fig. 1), two directions are defined relative to the scan direction: AL (along scan) and AC (across scan). In the spectra, AL is the dispersion direction and AC is the spatial direction. The spectra are acquired with 1260 or 1296 AL pixels depending on the onboard software version used; to mitigate the straylight effect, an updated onboard software, VPU version 2.8 (see Cropper et al. 2018), was put in operation in June 2015, and the spectra acquired since then have 1296 samples. The wavelength range is 845–872 nm and the resolution element is ~3 AL pixels. The RVS instrument and the spectra produced are described in Cropper et al. (2018), and the in-flight performance of the *Gaia* CCDs in Crowley et al. (2016).

To be treated by the spectroscopic pipeline, the telemetered RVS spectra are reformatted by the initial data treatment (IDT) pipeline (Fabricius et al. 2016), and so is the associated information necessary for processing the spectra, such as the detection features: time, coordinate, field of view (FoV), CCD row, solar rotation phase, onboard magnitude, the AC position of the window on the CCD, its size and truncation status (the windows, originally of $AC = 10$ pixel size, can be truncated if in conflict with other windows), and the pre-scan pixel values necessary to estimate the electronic bias. The main input data to the spectroscopic pipeline are the IDT products called *SpectroObservation*, containing the three CCD spectra corresponding to one transit (i.e. one single observation) of the source. The IDT also provides the files *BaryVeloCorr* containing the value of the barycentric velocity correction calculated every 5 minutes. The barycentric velocity correction is added to the spectroscopic,

Gaia-centric, radial velocity measurements to obtain the radial velocities relative to the solar system barycentre.

In addition to the IDT, the spectroscopic pipeline depends on the data produced by the intermediate data update (IDU). The IDU cross-match permits source identifier (sourceId) to be associated to each SpectroObservation. The star coordinates from the astrometric global iterative solution (AGIS) and the spacecraft attitude permit calculating the field angles (η, ζ), i.e. the coordinates in the FoV reference system needed by the wavelength calibration. For information on the AGIS2.1 solution used in the spectroscopic pipeline, see Lindegren et al. (2018), and for a description of the field angles and of the FoV reference system, see Lindegren et al. (2012) and Fabricius et al. (2016).

The IGSL is also part of the inputs to obtain the information on the G_{RVS}^{ext} of the star.

3.1. Extraction of the input data information

In the `Ingestion` step (Fig. 2), the information contained in the input data that is relevant for the downstream processing is extracted and stored in the format needed by the spectroscopic pipeline data model. During this step, the data to be processed downstream are selected (see Sect. 3.2).

For each CCD spectrum, some additional information is computed using the IDT and IDU input data:

- The time t^{obs} is calculated for each of the three CCD windows in a transit. This is the time at the mid point of the window as it passes the CCD fiducial line. The fiducial line is the nominal mean position of those light sensitive TDI lines (one CCD AL pixel corresponds to one TDI and to 982.8 μs) that contribute to the signal integration. Based on the time, t^{obs}, of the mid-point of the window, the time t_s^{obs} of the samples of the window are calculated.
- The field angles, $\eta(t_s^{obs})$ and $\zeta(t_s^{obs})$, are calculated for each sample of the spectrum using the AGIS astrometric parameters $\alpha, \delta, \mu_\alpha, \mu_\delta, \varpi$, and the attitude OGA3 (Lindegren et al. 2018).
- The barycentric velocity correction is calculated at any transit time by linear interpolation of the *BaryVeloCorr* data provided by the IDT. This correction is added to the measured *Gaia*-centric radial velocity (Sect. 7) in order to obtain a barycentric radial velocity.

3.2. Selections applied to the Input data

The spectra that could not be processed by this version of the pipeline were excluded from the data flow. The result is that the large majority (more than 95%) of the spectra acquired by the RVS are excluded in DR2, but are awaiting inclusion in future releases. The following criteria were used to filter out the data:

- Spectra of sources fainter than $G_{RVS}^{ext} = 12$ are filtered out. This is the most important filter, as the large majority of the RVS data are fainter than 12 (the faintest RVS spectra have onboard $G_{RVS} = 16.2$). The value of $G_{RVS}^{ext} = 12$ was chosen based on general considerations: it corresponds to a $S/N \sim 2$ in a single CCD spectrum, assuming the median straylight level observed at the beginning of the spacecraft operations (~5 e$^-$ pix^{-1}).

  However, the G_{RVS}^{ext} is only a rough estimation, and many fainter stars (based on their flux observed in the window and their G_{RVS}^{int}) entered the processing despite this filter.
- Spectra with non-rectangular truncated windows are filtered out. This filter is at transit level: depending on the particular observation geometry during a transit, a source

spectrum may or may not overlap with one or more other source spectra. This filter excludes approximately 40% of the remaining spectra. The overlapping is more important in crowded regions, where 60–80% of the spectra have a truncated window.

Because some bright stars produce nearby spurious sources, they have a truncated but rectangular window, and are not filtered out. Their spectra exhibit a flux loss, which combined with non-perfect AC centring of the window and the tilt of the spectrum can result in an artificial spectral slope. Approximately 40% of the stars with onboard magnitude G_{RVS} between 7 and 9 are affected and have a window AC size of 5 pixels (instead of the normal size of 10). The spurious detection events decrease rapidly with the magnitude of the star, and at $G_{RVS} \sim 11$, only ~5% of the stars have a reduced size AC window. The 2D windows of the very bright stars are never truncated because of other source windows, and may be truncated only by the CCD borders.

– Sources without AGIS coordinates are filtered out. So are spectra acquired during pre-defined time intervals, where the data are known to be of poor quality. These intervals include the time during the decontaminations, the refocusing, the commissioning of the VPU 2.8, the time intervals where the AGIS residuals are high, and the time gaps in the IDT barycentric velocity correction that are longer than 10 minutes.

The total time covered by all the excluded intervals is ~200 revolutions or ~7.5% of the total observation time.

4. Auxiliary data

In this section, the auxiliary non-*Gaia* data used by the pipeline are described. Some of them play an important role in the calibration and in the determination of the V_R results.

4.1. Auxiliary radial velocities of standard stars

The radial velocity standard stars used by the wavelength-calibration modules have V_R that is constant in time and precise at the level of 0.1 km s^{-1}, as determined from high-precision ground-based observations. They are needed as external calibrators to fix the RVS V_R zeropoint. The auxiliary V_R catalogue, including the 2568 standard stars with the highest precision and accuracy used in wavelength calibration, is described in Soubiran et al. (2018). The V_R zeropoint of that dataset is set by the SOPHIE spectrograph (Perruchot et al. 2008). By being calibrated on this dataset, the RVS V_R zeropoint should agree with that in Soubiran et al. (2018).

The wavelength calibration software module, which works per CCD and FoV (Sect. 5.3), requires standard stars covering the AC dimension of each CCD in order to produce good results in each calibration unit (a calibration unit consists of a dataset covering 30 h of observations and containing the calibrator stars; see also Sect. 5.3). This is achieved, in general, by observing approximately 200–300 standard stars every 30 h.

An additional 5729 stars were used as standard stars. They were extracted from the master list in Crifo et al. (2010), where the photometric variables and known multiple-systems are excluded, and only the stars with F-G-K spectral types were kept. The list was then correlated with the Extended HIPPARCOS Compilation (XHip; Anderson & Francis 2012), and only stars with a V_R measurement with an uncertainty $\sigma_{V_R} < 1$ km s^{-1} and quality A and B were retained. Finally, the list of candidate standards was refined using the RVS observations obtained during

Table 1. External radial velocity catalogues used in the pipeline.

Cat Name	nb stars	nb transits	σ_{V_R} km s^{-1}	pipeline use
CU6GB-cal	2568	34630	<0.1	calibration
XHip	5729	83366	<1.0	calibration
CU6GB-val	1819	16819	<0.3	validation
RAVE	9542	86035	<1.5	validation
APOGEE	8584	76177	<0.5	validation
SIM	650	4885	<0.1	validation

Notes. The first two catalogues are used in wavelength calibration and are described in Sect. 4.1. The other catalogues (Sect. 4.2) are used to verify the RVS pipeline results. σ_{V_R} is the uncertainty associated with the radial velocity measurements provided by each catalogue.

the trial runs of the pipeline, together with the observations obtained by the daily pipeline. The stars with an RVS V_R different by more than 3 km s^{-1} from the one in XHip and the stars that varied between the RVS single-transit measurements >3 km s^{-1} were excluded from the list. We also removed stars with a double-lined spectrum.

Table 1 lists the standard star catalogue CU6GB-cal (derived from the name of the *Gaia* spectroscopic team; GB stands for ground-based) and XHip. Figure 3 shows the magnitude G_{RVS}^{ext} distribution of the stars belonging to these catalogues.

4.2. Auxiliary radial velocities of validation stars

The validation dataset consists of stable stars for which a ground-based radial velocity is available. Table 1 also lists the catalogues used for validation in the automated verification workflow (Sects. 7.8 and 9) and in the off-line validation (Katz et al. 2018). Figure 3 shows the G_{RVS}^{ext} magnitude distribution of the stars processed by the pipeline. The selection of the stable stars was based on the following criteria:

– CU6GB-val: All the stars from Soubiran et al. (2018) that were not used as CU6GB-cal were used as validation stars. These stars are presumed to be stable based on at least two high-precision ground-based observations over a minimum time baseline of 30 days and with a standard deviation <0.3 km s^{-1}. It includes observations with standard deviation <0.1 that do not fulfil all the RVS requirements for calibrations such as magnitude and spectral type, and some new observations qualified for calibration that could not be integrated in the pipeline processing workflows.

– RAVE: Stars from DR5 with a constant radial velocity within 1.5 km s^{-1} during at least two RAVE observations (Zwitter et al. 2016).

– APOGEE: Stars from DR3 that exhibit a constant radial velocity within 0.5 km s^{-1} (i.e. the scatter on the radial velocity, *vscatter* ≤ 0.5 km s^{-1}) during at least four APOGEE observations, and the χ^2 probability for non-constancy is < 0.1 (*stablerv_chi2_prob* <0.1).

– SIM: Stars from the Radial Velocities of the Southern Space Interferometry Mission Grid Stars Catalogue (Makarov & Unwin 2015) have a binarity probability <95%.

4.3. Auxiliary atmospheric parameters

The atmospheric parameters (T_{eff}, $\log g$ and [Fe/H]) of the observed stars are used for two tasks in the pipeline:

– to select the synthetic template that matches the stellar spectrum (the template is selected based on the minimum
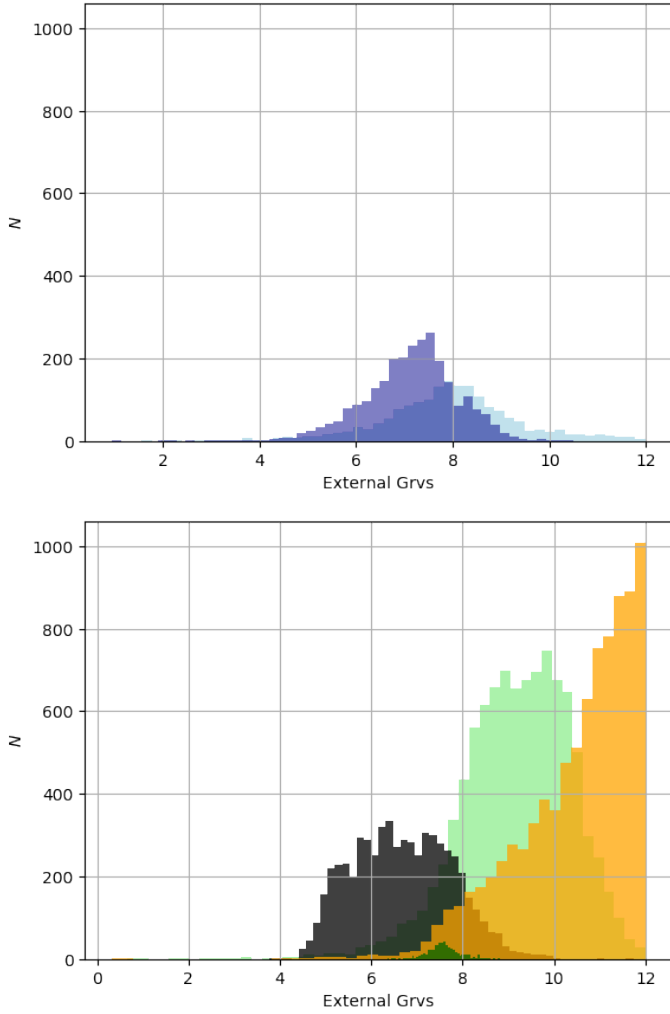
Fig. 3. Distribution on $G_{\mathrm{RVS}}^{\mathrm{ext}}$ of the stars belonging to the auxiliary radial velocity catalogues that are processed by the pipeline. *Top:* Magnitude distribution of the stars in CU6GB-cal (dark blue) and in CU6GB-val (light-blue). *Bottom*: Magnitude distribution of the stars in XHip (black), RAVE (light green), APOGEE (orange), and SIM (dark green).

distance between the atmospheric parameters of the star and those of the synthetic spectrum, Sect. 6.5);

– to select calibration stars with a suitable spectrum for the calibration in question. Specifically, the wavelength calibration needs spectra with pronounced Ca II lines, therefore the selection is $4500 < T_{\mathrm{eff}} < 6500$ K.

Because the T_{eff} measurements obtained with *Gaia* data in Andrae et al. (2018) were not available to the pipeline, a list of auxiliary atmospheric parameters was compiled. It contains parameters, mostly T_{eff}, for ~1.8 million stars.

The auxiliary catalogue, AuxAtmParams, contains ~800 000 values of T_{eff} and ~650 000 values of $\log g$ and [Fe/H] taken from the literature (Soubiran et al. 2014), and another ~1 million values of T_{eff} that have been derived for Tycho2 stars (with $4500 \le T_{\mathrm{eff}} \le 7500$ K) using the 2MASS photometry point-source catalogue (with the quality flags taken into account) and the Casagrande et al. (2010) effective temperature versus $J − K$ colour relations [T. Zwitter, intern. *Gaia* commun.].

Given that some stars appear in several catalogues, a priority selection was performed, following Soubiran et al. (2014). Priority was given to the spectroscopic over the

photometric estimation, and for the photometric catalogues, preference was given to those providing the three parameters instead of only T_{eff}.

Only some (i.e. ~15%) of the stars treated in DR2 are in the auxiliary atmospheric parameter files and have set atmospheric parameters. The treatment for the other stars is described in Sect. 4.5.

4.4. Auxiliary synthetic spectra

Synthetic spectra were used to generate the template spectra that simulate noiseless RVS spectra (Sect. 6.7). The templates were cross-correlated with the RVS spectra both for the V_{R} determination (Sect. 7.1) and in the wavelength calibration to determine the spectral line positions (Sect. 5.3).

The auxiliary synthetic spectra library used in the pipeline is composed of 5256 spectra selected from an updated version of the synthetic spectral library described in Sordo et al. (2011). The Sordo et al. (2011) spectral library was updated, and the selection of the grids for the spectroscopic pipeline is described in Blomme et al. (2017). The main improvement was the extension of the MARCS grid (Gustafsson et al. 2008) to the cooler stars. The following grids were selected:

– Spectra obtained with MARCS models:
   • T_{eff}: 2500–8000 K; step 100 K between 2500–3900 K and 250 K between 4000–8000 K;
   • $\log g$: −0.5 to +5, step 0.5;
   • [Fe/H]: −5.0, −4.0, −3.0, −2.5, −2.0, −1.5, −1.0, −0.75, −0.50, −0.25, 0.0, +0.25, +0.5, +0.75, +1.0;
– A-type spectra:
   • T_{eff}: 8500–15 000 K; step 500 K;
   • $\log g$: +0.5 to +5, step 0.5;
   • [Fe/H]: −0.5 to +0.25, step 0.25;
– OB-type spectra:
   • T_{eff}: 15 000–55 000 K, step 1000 K between 15 000–30 000 K, step 2500 K between 30 000–50 000 K;
   • $\log g$: highest value: +4.75, lowest: approximately linearly from 1.75 at 15 000K to 4.0 at 55 000 K;
   • [Fe/H]: −0.3, 0.0, +0.3.

The synthetic spectra are characterised by a number of parameters. The main parameters are effective temperature (T_{eff}), gravity ($\log g$), and metallicity ([Fe/H]). Secondary parameters include the abundance of the α elements ([α/Fe]), and the turbulent velocity (v_{turb}).

The selection was made to reduce the number of spectra to those that could be exploited by the pipeline, and to have only one spectrum per set of main parameters. The MARCS spectra are available for various values of the [α/Fe] parameter, and the lowest non-negative [α/Fe] for a given T_{eff}, $\log g$, and [Fe/H] was selected. The OB spectra are available with two values of v_{turb}, and those with $v_{\mathrm{turb}} = 2$ km s^{-1} were selected. The A-star and OB-star grids overlap at $T_{\mathrm{eff}} = 15\,000$ K. In that case, we selected the OB-star spectra because that grid is denser.

The current set of synthetic spectra does not include convective motions or gravitational redshift. At a later stage, the synthetic spectra are planned to be replaced by versions that do include convective motions (e.g. Chiavassa et al. 2011, 2018; Allende Prieto et al. 2013).

4.5. Synthetic template spectra

For those stars without atmospheric parameters in the auxiliary catalog (the majority), a pipeline module, called `DetermineAP` (Sect. 6.5) was used to determine their atmospheric parameters

by cross-correlation with the star spectrum. The template that matched the RVS spectrum best was selected, and the atmospheric parameters associated with this template were attributed to the star.

This restricted template spectral library, containing 28 spectra, was created for `DetermineAP`, and contains a subset of the synthetic spectra data (Sect. 4.4). The following arguments were used to construct this restricted synthetic template data set:

– The set should be relatively small, so that the computation time is short.
– The steps in the atmospheric parameters should be chosen based on changes in the spectrum morphology.
– The set should represent the RVS data treated by the pipeline. Because in DR2 we treat only $G_{RVS} \leq 12$ and because for bright stars, the morphology of the spectra is important, we used two sequences, one with [Fe/H] = 0 and T_{eff} = 3100, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 9000, 10 000, 15 000, 20 000, 25 000, 30 000 and 35 000 K and the other with [Fe/H] = −1.5 and T_{eff} = 3100, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, to represent cool stars with low metallicity.
– For the gravity we used Table 3.3 from Robin et al. (2012), suggesting that most stars are giants with $\log g \sim 3$ (but slightly depending on T_{eff}). The main-sequence value of $\log g = 4$ is used for the hotter stars.
– A rotational velocity = 0.0 km s^{-1} was assumed when converting these synthetic spectra into templates.

The atmospheric parameters of the 28 spectra that were selected from the auxiliary synthetic spectral library are also tabulated in Sect. 6.5.

The 28 synthetic spectra were transformed into templates to be used in `DetermineAP`. The transformation of the synthetic spectra into templates was made as described in Sect. 6.7, with the difference that here the spectra were convolved with a Gaussian LSF with $R = 11 500$, independent of the CCD and FoV. This simple template simulation is sufficient for the `DetermineAP` purposes, which is to compute the maximum value of the correlation peak for a template selection and not to derive a radial velocity.

4.6. Auxiliary G_{RVS}

The G_{RVS}^{ref} magnitudes of the stars contained in this file were used as the reference to estimate the G_{RVS}^{int} zeropoint (Sect. 5.5). We computed G_{RVS}^{ref} for 113 686 HIPPARCOS stars using the formula $G_{RVS}^{ref} = V - 0.0501 - 1.1667 * (V - I) + 0.0052 * (V - I)^2 + 0.0011 * (V - I)^3$, where V and I are the magnitudes in the HIPPARCOS catalogue (Jordi et al. 2010). The magnitude estimation using the V and I bands is more precise than the one that we have in IGSL, and for this reason, G_{RVS}^{ref} (and not G_{RVS}^{ext}) was used for the zeropoint estimation.

5. Calibration of the RVS instrument

The pipeline for this first data release provides the first-order calibrations necessary to treat bright star spectra that do not overlap with spectra in other windows. More calibrations will be necessary in future versions to treat fainter stars and blended spectra.

The most important calibration necessary for estimating V_R is the wavelength calibration. It is included in the pipeline, which calculates the associated calibration parameters and their spatial and temporal dependence. The payload module
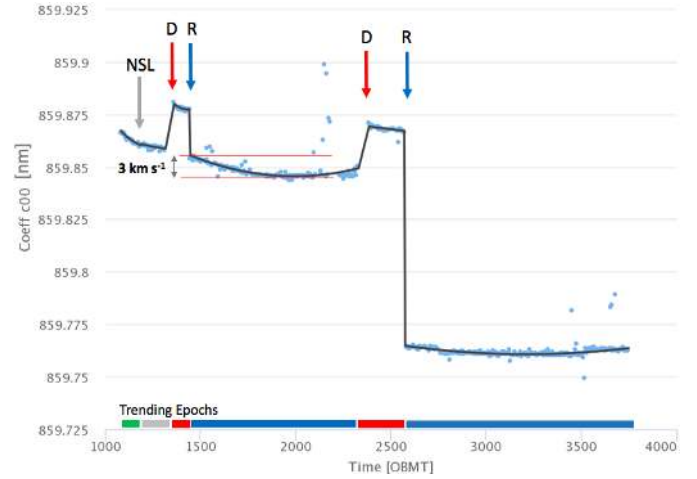
Fig. 4. Wavelength calibration model: Trending function of the coefficient C_{00}, representing the wavelength-calibration zeropoint, shown for the CCD in row 5, strip 16, and FoV1 (black line). The timescale on the abscissa is expressed in OBMT (revolutions, Sect. 3). The blue points are the values of C_{00} obtained at each calibration unit. The arrows indicate breakpoints. The most important discontinuities are due to optics decontamination (red arrows, at OBMT 1317 and 2330.6) and to the re-focusing events (blue arrows, at 1443.9 and 2574.6). An additional breakpoint was set at the transition between EPSL and NSL, at OBMT 1192.13 (grey arrow). The DR2 data set that covers the time between OBMT 1078 (25 July 2014) and 3750.5 (23 May 2016) was divided into six trending epochs that are indicated with colour bars along the abscissa: shortly after each decontamination, a re-focusing was needed, and the trending epochs between these two events (the red bars) are short. The two red horizontal lines delimit the variation of the wavelength zeropoint over one trending epoch. These variations are typically small, <3 km s^{-1} (~0.3 pixels).

decontamination and refocusing operations produce discontinuities in the basic angles that are reflected in the wavelength calibration coefficients. Figure 4 shows the temporal evolution of the coefficient C_{00} representing the wavelength calibration zeropoint. An additional smaller discontinuity is produced by the change in the *Gaia* scanning law from the ecliptic pole scanning law (EPSL) to the nominal scanning law (NSL). *Gaia* was operated in EPSL during the first weeks of operations (between 25 July and 21 August 2014), with the spin axis in the ecliptic plane and the FoV scan through both the south and north ecliptic poles (Gaia Collaboration 2016). Based on these discontinuities, five breakpoints were defined that correspond to these events, and the entire calibration data set was separated into six trending epochs. A calibration model was produced for each trending epoch.

In addition to the wavelength calibration, the bias non-uniformity (bias NU) effect (Sect. 5.1), which produces jumps in the spectra (see Fig. 10) that compromise the V_R estimation even in bright stars, needed to be calibrated. The calibration coefficients were computed by a dedicated RVS pipeline, and were shown to be stable for the DR2 purposes (Hambly et al. 2018), so no temporal variation was implemented.

The other calibrations that were necessary to treat the spectra but that had only second-order impact on the V_R estimation of bright stars were not included in the pipeline. These are the line-spread-function in the AL direction (LSF-AL), the CCD blemishes and saturation level, and straylight. The calibration model used for these was determined off-line using a limited data set, or the pre-launch data. We describe here the calibration

models we used to obtain the DR2 V_R, both those produced by the pipeline, and those produced off-line.

5.1. Electronic bias non-uniformity

Hambly et al. (2018; see also Cropper et al. 2018) described the bias NU effect, its calibration procedure, the resulting parameters and their temporal evolution in detail. A constant voltage offset is applied to the CCDs prior to digitalisation. This electronic bias is needed to avoid negative signal values and wrap around zero-digitised units (ADU) at low signal levels. This constant bias level is measured from the prescan data that are acquired every 70 minutes in bursts of 1024 samples. The constant (uniform) bias calibration consists of extracting the median signal from the bursts of prescan data for each CCD and of applying a linear interpolation to these data to obtain the bias offset that is to be subtracted from the spectra at any time.

In addition to this constant bias, owing to the manner in which the *Gaia* CCDs are operated, the electronic bias applied to an acquired sample prior to digitisation contains an unstable component such that the bias value applied to a given sample depends on the details of the read-out process, in particular on the number of flushes (fast skipping of the inter-window pixels before the sample), the position of the glitches caused by interruptions of the serial clocks during the operation of the parallel clocks, and the common baseline (the difference between the prescan bias and the image area bias). These features are known collectively as the bias NU anomaly and are described in detail by Hambly et al. (2018).

The calibration of the RVS bias NU is done by a dedicated pipeline using well-defined sequences of samples that are called special sequences and are acquired during special calibration periods about three to four times per year, using CCD gates to hold back photoelectrons. The result of the calibration is a set of coefficients that model the flushes and glitches for each CCD. These coefficients are used to correct the RVS spectra (Sect. 6).

The special sequences were acquired twice during the 22 months covered by the DR2 data, and the RVS bias NU effect appeared to be stable (Hambly et al. 2018). For this reason, only one set of coefficients (the most frequently tested one that was obtained during commissioning) was used to correct the entire data set, and any temporal variation of the bias NU anomaly was neglected.

5.2. Scattered light calibration

After launch, it was discovered that the readout windows of the RVS spectra were contaminated with a diffuse background dominated by solar straylight (Cropper et al. 2018). The straylight contamination varies over time and also with the position on the focal plane, but a large part of it follows a quite stable pattern based on the satellite rotation phase. The typical pattern is shown in Fig. 5.

The background level was measured using the virtual objects (VOs), which are windows containing only background signal (no sources) that are acquired every 2 seconds and cover the CCDs uniformly in the AC direction. The background level measurements obtained from the VOs are used to build a 2D map (Fig. 5) dominated by the solar straylight, which is called the scatter map. The scatter map is used to obtain the background level to subtract from the RVS spectra, depending on their AC position on the CCD and on the rotation phase of the satellite during their acquisition.

The scatter map used in the pipeline was produced off-line using the VOs acquired only during the first 28 days of *Gaia*
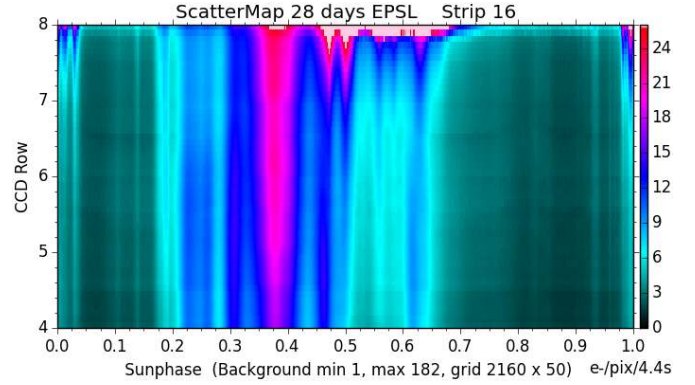

Fig. 5. Background map (scatter map) used in the pipeline for the four CCD rows in strip 16 (strip 16 in the *Gaia* focal plane is numbered 2 in the RVS focal plane). On the y-axis is the across scan (AC) direction plotted as CCD row number, while the x-axis is the *Gaia* solar phase angle. This scatter map is obtained using the VOs acquired during the 28 days of operations, when *Gaia* was in EPSL mode (first trending epoch). The scatter map is dominated by the solar straylight. The units are e^- pix^{-1} (not divided by the 4.4 s exposure time).

operation, when *Gaia* was in EPSL mode, in the first trending epoch. It is used to correct the entire data set, and temporal variations are not considered.

To produce the scatter map, the software proceeds in two steps:

1. `MakeScatter` is responsible for treating the single VOs. First, the VO spectra with truncated windows, 2D windows, or windows that are affected by a CCD defect are discarded to remove other signal contamination from the background measurements. Then, the VO spectra are corrected for the bias NU. The background flux level is calculated as the median flux of all valid samples in the VO spectrum after excluding the samples that are affected by cosmic rays or saturation. These results are stored in data files (one per VO), called *Scatter*, that also contain all the information needed to create the scatter map: the CCD and the AC position of the VO, including the CCD row as well as the solar phase of *Gaia* at the acquisition time.

2. `MakeScatterMap` uses all the accumulated Scatter data and combines them based on their position, to create one 2D map. A map covers one revolution time in x and all CCD AC positions in y. It is composed of a grid with 2160×50 cells (10 s AL and 157 AC pixel per grid cell), chosen in order to have a sufficient number of background Scatter measurements per cell (18 measurements). For each Scatter measurement, the solar phase angle is calculated using the satellite attitude. The AC cell index is calculated as $y = \frac{rowfrac - 4.}{4.} * ngridcells$, where *rowfrac* is the AC location, subtracting 14 prescan pixels and half a pixel to refer to pixel centre, then divided by 1966, which is the number of exposed CCD pixels. For each grid cell, the straylight level is then the median value of all corresponding Scatter background measurements, and the associated uncertainty is the standard error of the median.

The correction applied to the spectra is a simple subtraction of a constant level from the entire spectrum and does not significantly affect the V_R estimation for these bright stars. However, it affects the magnitude G_{RVS}^{int} (Sect. 6.3). The background level calibration was not a priority for this pipeline version and will be improved for DR3.
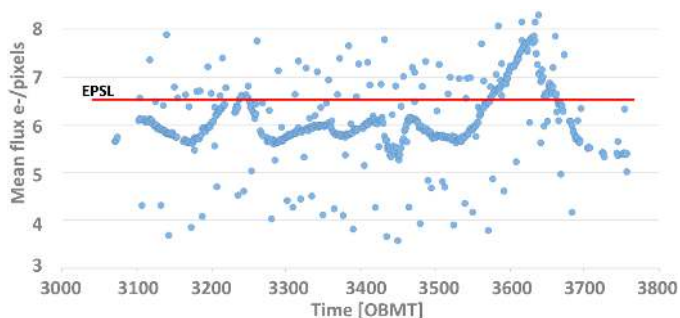
Fig. 6. Temporal evolution of the mean level of the scatter map in CCD row 6 strip 16 during the last period covered by the DR2 data set. The timescale is expressed in OBMT, and the time interval shown corresponds to 12 December 2015 and 23 May 2016. Each blue point represents the mean background level over 30 h. It is computed using the scatter map obtained with the VOs acquired in 30 h (outliers are from VO incompleteness). The units are e⁻/pix, not divided by the 4.4 seconds exposure time. The red horizontal line at 6.54 e⁻ pix⁻¹ is the mean level of the scatter map obtained with the EPSL VO data and is used to process the whole DR2 data. The peak at OBMT 3600 corresponds to a scanning of the Galactic plane by *Gaia* and arises from the contribution to the straylight from the high density of external (i.e. non-solar) sources in these scans.

In order to quantify the consequences of neglecting time variations and using the constant scatter map obtained off-line with the EPSL data to process all of the DR2 data, the average background level of the EPSL map was compared with the average level of the background over the last five months of data covered by DR2 . During this period, the daily pipeline (which started operation in December 2015, with the primary goal of tracking and verifying the status of the RVS) estimated and monitored the average background level every 30 h. Figure 6 shows the variation in average background level (in CCD strip 16) between December 2015 and May 2016 compared with the average EPSL scatter map level from August 2014. The figure shows that the mean background level during EPSL was higher than during the period between December 2015 and May 2016. This may be explained by the fact that EPSL receives a constant contribution from the galactic plane and/or zodiacal light that is due to the stable orientation of the *Gaia* spin axis (not doing any precession).

5.3. Wavelength calibration

The RVS is not equipped with wavelength-calibration lamps as these are difficult to implement in TDI operation. Hence a self-calibration approach is adopted, where some of the RVS observations themselves are used to calibrate the instrument (Guerrier et al. 2007). The basic principle of the RVS wavelength calibration is that the wavelength value associated to a sample can be expressed as a function of the FoV coordinates (η, ζ) of the source at the time when the sample crosses the CCD fiducial line (Sect. 3.1).

Each FoV and CCD of the RVS is calibrated independently and continuously using the data acquired during time intervals of 30 h in order to obtain one wavelength calibration every 30 h. This time period, over which the calibration stars and the characteristics of the instrument are assumed to remain constant, defines the wavelength calibration unit (CaU) duration.

The time evolution of the wavelength calibration (zero-point, dispersion, and ariations in the spatial direction) is then
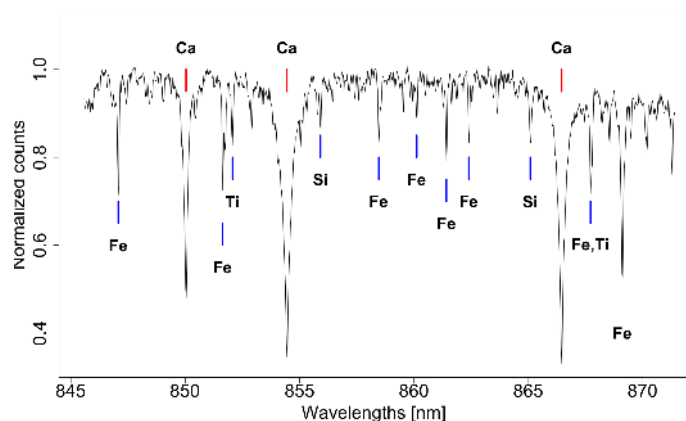
Fig. 7. RVS spectrum of a bright star used for wavelength calibration. The calcium triplet lines are indicated in red, and some Fe and Si lines are indicated in blue.

modelled in the trending module, using the information obtained in all CaUs.

5.3.1. Selection of calibrator stars

The first step of the wavelength calibration process is to select those star spectra from the RVS observations acquired during the CaU duration that present several well-detected narrow lines to be reference line candidates. In practice, we selected the spectra of stars with $T_{\rm eff}$ in the interval $4500 < T_{\rm eff} < 6500$ K, and $G_{\rm RVS} \leq 9$ (Fig. 7 shows a typical spectrum of a calibrator star). The calibration stars (those that satisfy the selection conditions) also include $V_{\rm R}$ standard stars with well-known, stable radial velocities from the auxiliary radial velocity catalogue (Sect. 4.1). These are used to fix the wavelength calibration zeropoint. The spectra have passed through the `Calibration Preparation` workflow (Fig. 2) and were cleaned (Sect. 6) and calibrated with the first-guess model. The wavelength range of each input spectrum was reduced to avoid including leading/trailing edge data that might confuse the algorithm. In addition, the suitable template spectrum was selected from the auxiliary synthetic spectral library (Sect. 4.4). The identified template lines were used to "predict" where the corresponding RVS spectrum lines might lie.

5.3.2. Identification of the reference lines

The second step is to identify the reference lines suitable for wavelength calibration in these calibration spectra and to associate the triplet $(\lambda_{\rm rest}, \eta, \zeta)$ with each of them. There is no definitive list of reference lines. Rather, the list of lines is determined by the template assigned to each calibrating spectrum. Blended lines and lines presenting cosmic rays or saturation are discarded. These are the tasks of a software module called `Centroiding`, which works in the following way:

– Detection of the deepest lines in the spectra. The 20 deepest lines of the spectrum are detected (they are sorted by decreasing depth), and estimates of the location of their centroids are evaluated. These first centroids are expressed in integer sample values, with the centre of the bluest sample of the spectrum associated with the value 0. The Ca lines are easily identified; they are normally the deepest lines in the spectra, but the code takes into account that sometimes the Fe I line, which lies close to the red wing, is observed as a deeper line than some of the triplets.

- Identification of the reference lines based on their predicted positions. The prediction algorithm uses the Ca line sample positions identified during the line detection phase and the wavelengths of these lines in their template counterpart to compute a wavelength scale. Next, the algorithm takes each template line in turn and identifies a sample within the RVS spectrum that represents the notional centre of the matching line (s_0). It then assigns four samples from either side of this centre to construct a nine-sample wide line. If the centre sample is the deepest flux sample of the nine, then the line is deemed a reference line; otherwise, it is discarded.
- Deriving rest wavelength. The rest wavelength is derived via cross-correlation between the detected reference lines and their template counterparts. A local calibration relation is defined,

$$\lambda = k \times s + \delta\lambda = k \times s' + \lambda_0, \tag{1}$$

where s is the sample, λ is the wavelength corresponding to the sample s, k is the sampling factor which takes into account the spectral resolution element relative to the sample size, s' is the changed frame of the observed line ($s' = s - s_0$, where s_0 is the discrete centroid of the line), $\delta\lambda$ is the wavelength shift of the local calibration law, and λ_0 is the wavelength location of the discrete centroid ($\lambda_0 = k \times s_0 + \delta\lambda$). The best match between the template and the observed lines is found by a classical cross-correlation, that is, by computing the optimal parameter λ_0 associated with the sample s_0. Because the template is at rest, $\lambda_0 = \lambda_{\text{rest}}$. The rest wavelength is derived as follows:

- Shift the local calibration relation in wavelength by modifying the parameter λ_0 of the local calibration relation.
- Calibrate the observed line using the local calibration relation.
- Re-sample the at-rest template to the sampling of the observed line.
- Compute and analyse the correlation coefficient using the RVDir method (see Sect. 7.3).
- Using an interpolation algorithm, the centroid locations ($s_0, \lambda_{\text{rest}}$) are expressed in field angles (η_s, ζ_s).

At the end of this process, each reference line of each spectrum has a position and a rest wavelength: ($\lambda_{\text{rest}}, \eta_s, \zeta_s$).

5.3.3. Determination of the dispersion function

The third step of the wavelength calibration process is to determine the spectral dispersion for the CaU in question, using the reference line triplets ($\lambda_{\text{rest}}, \eta_s, \zeta_s$) produced in the previous step. The dispersion function is represented as a second-order bivariate polynomial,

$$\lambda_{\text{observed}} = \lambda_{\text{rest}}\left(1 + \frac{V}{c}\right) = \sum_{m=0}^{2}\sum_{n=0}^{1} C_{mn}\eta_s'^m\zeta_s'^n, \tag{2}$$

where

- $\lambda_{\text{observed}}$ is the wavelength associated with the centre of the sample;
- C_{mn} are the unknown calibration coefficients;
- V is the unknown *Gaia*-centric radial velocity of the calibration star;
- η_s' and ζ_s' are the field angle FoV coordinates of the calibration stars at the fiducial time of the CCD sample, shifted and scaled. Shifting and scaling are needed to improve the stability of the solution;

- η_s' is shifted and scaled η_s: $\eta_s' = s_\eta(\eta_s - \eta_0)$;
- s_η is the η scale factor, which is the same for all configurations: $s_\eta = 1000$;
- ζ_s' is shifted and scaled ζ_s: $\zeta_s' = s_\zeta(\zeta_s - \zeta_0)$;
- s_ζ is the ζ scale factor, which is the same for all configurations: $s_\zeta = 10$;
- η_0 and ζ_0 are the pivot points, they are fixed for each FoV and CCD and correspond to the projection of the middle of the readout register into the η, ζ plane for a wavelength of 860.5 nm.

To improve the stability of the solution, we set to zero the coefficients C_{21}, C_{02}, C_{21}, and C_{12} in Eq. (2), and obtain

$$\lambda_{\text{observed}} = \lambda_{\text{rest}}\left(1 + \frac{V}{c}\right) = C_{00} + C_{10}\eta_s' + C_{20}\eta_s'^2 + C_{01}\zeta_s' + C_{11}\eta_s'\zeta_s'. \tag{3}$$

To avoid the degeneracy because a shift in the dispersion law can be compensated for by a shift of the radial velocities of the calibrator stars, the auxiliary ground-based standard stars with known radial velocity (Sect. 4.1) are used to fix the zeropoint. The standard stars are also part of the calibrator stars because they satisfy the conditions required to be selected as calibrator. V^{ref} is their radial velocity, transformed into *Gaia*-centric radial velocity.

The calibration coefficients and the radial velocities (V^s) of the calibration stars are derived simultaneously by a least-squares fit, minimising the following function χ_k^2:

$$\chi_k^2 = \sum_{r=1}^{N_{\text{spectra}}(k)}\sum_{l=1}^{N_{\text{line}}}\left[\lambda_{r,l} + \left(\frac{\lambda_{r,l}V^{\text{ref}}}{c}\right) + \left(\frac{\lambda_{r,l}V^{r\neq\text{ref}}}{c}\right) - \sum_{m=0}^{2}\sum_{n=0}^{1} C(k)_{mn}\eta_{r,l}'^m\zeta_{r,l}'^n\right] \tag{4}$$

where

- $N_{\text{spectra}}(k)$ is the number of spectra observed during the calibration unit k;
- N_{line} is the number of reference lines in each spectrum;
- $C(k)_{mn}$ are the calibration coefficients for the calibration unit k (C_{02}, C_{21}, and C_{12} were set to zero);
- $\lambda_{r,l}$ is the measured rest wavelength of the reference line from `Centroiding` (dependent variable);
- V^{ref} is the known radial velocity of a ground-based standard; it is shifted to *Gaia*-centric by subtracting the *Gaia* barycentric velocity correction (dependent variable);
- V^r is the unknown *Gaia*-centric radial velocity of the star in the rth spectrum;
- $\eta_{r,l}'$ and $\zeta_{r,l}'$ are the FoV angular coordinates of the calibrator star (corresponding to the rth spectrum) at the time associated with the location of the centroid of the lth reference lines in the rth spectrum (independent variable).

The minimisation of the χ_k^2, equivalent of resolving a linear system is accomplished using the `Single Value Decomposition` algorithm, implemented in the `Efficient Java Matrix Library` [3].

5.3.4. Modelling temporal variations

In the last step, the temporal variations of the wavelength calibration are modelled. The wavelength calibration process is repeated for all the CaUs of the trending epoch. After the coefficients for the whole epoch are obtained, a trending module analyses the long-term variation of each calibration coefficient, and performs a curve fit that best describes the emerging trends.

---

[3] https://ejml.org/wiki/index.php?title=Manual

The type of function that best fits the temporal trend of C_{00} and C_{10} is a second-order polynomial, while a simple median value is used to model the trend of the other coefficients.

The trending functions are the final results of the wavelength calibration and are used to calibrate all the spectra acquired at any time and at any position. Figure 4 shows, as an example, the calibration values and the trending function (black line) of the coefficient C_{00} that represents the wavelength-calibration zero-point for one of the CCDs of the leading FoV, FoV1, during the time covered by the DR2 data.

5.3.5. Automated verification

The quality of the dispersion functions obtained for each CaU and for each FoV and CCD is verified by the `Automated Verification` module. First, the number of the standard stars and of the calibrator stars used to compute the solution is verified (we need to have a sufficient number of standard stars covering the ζ coordinate in the CCD). Then, the residuals between the wavelength of the reference lines obtained by the line centroid algorithm and those obtained by applying the dispersion law are computed: $\Delta_\lambda = \lambda_{calibration} - \lambda_{rest}(1 + \frac{V^s}{c})$, where

- $\lambda_{calibration}$ is calculated using the dispersion law (Eq. (9)) and the field angles η_s and ζ_s of the reference lines;
- λ_{rest} is the rest wavelength of the reference line;
- c is the velocity of light in vacuum;
- V^s is the velocity of the star, obtained with the wavelength calibration.

The temporal variations of the calibration coefficients and the trending functions are inspected by eye. The final quality estimation of the wavelength-calibration model is provided in Sect. 7.8, after the extraction of the radial velocities of the standard stars, by comparing the results obtained with the RVS data with the auxiliary ground-based standard values.

5.4. Line spread function (LSF-AL)

The *Gaia* point spread function (PSF) is approximated by the cross product of the AL and the AC line spread functions (LSF). The LSF-AL and LSF-AC calibrations are not implemented in this pipeline version. The LSF-AC calibration will be implemented for DR3 and will be used for deblending the spectra and to estimate the flux loss out of the windows. The LSF-AL, on the other hand, is needed in this pipeline version. It contains information on the resolution of the RVS spectra and is used to convolve the synthetic spectra to generate the templates. The LSF-AL calibration was therefore estimated offline, using the 28 days of EPSL data, and was used for the data acquired before the first decontamination at OBMT 1317. For the remaining data, the LSF-AL calibration was derived using on-ground data. In future data releases, the LSF calibration will be included in the pipeline and the estimation will be improved.

The RVS resolution element is monitored daily in each configuration in the first-look (FL) pipeline (Fabricius et al. 2016) since the beginning of the *Gaia* operations in order to detect defocusing. In the FL, the resolution element is determined by measuring the width of some unresolved Fe lines via correlation with a mask. The FL diagnostics show that the resolution element is in general very close to the nominal value (3 pixels, corresponding to $R = 11\,500$), with a maximum degradation of ~3% owing, in general, to gradual de-focus. The degradation of the resolution element appeared much more important, ~20%, at the beginning of the operation before OBMT 1317, when the *Gaia* optics suffered from high contamination by water
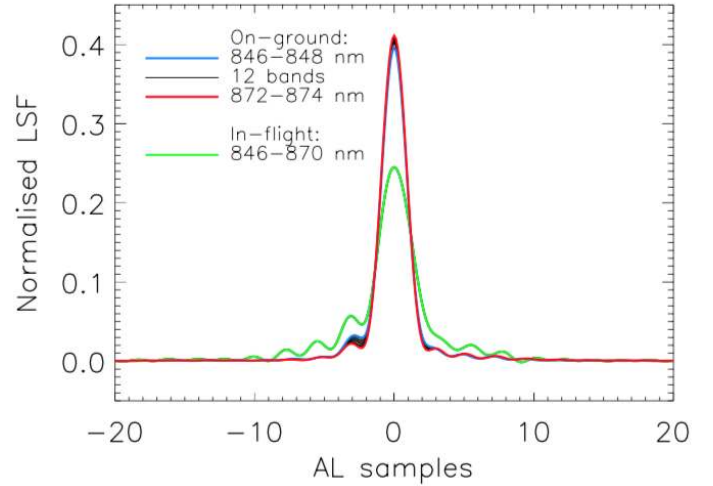
Fig. 8. LSF-AL model used in the pipeline for the CCD in row 6 strip 16 and FoV1. The in-flight LSF (green line) is computed using RVS data obtained before the decontamination at OBMT 1317 and shows a significant degradation compared to the nominal, on-ground LSF. The resolution degradation was ~20% and was recovered after the decontamination. The on-ground LSF is estimated in 15 wavelength bands. The shortest and longest wavelength bands are coloured blue and red, respectively; the two curves almost overlap, showing the weak dependence of the RVS LSF-AL on wavelength, which is neglected in the in-flight model.

ice. The high contamination is also shown by bright values of the magnitude zeropoint in Fig. 9.

Two LSF-AL models were estimated, both using principal component analysis (Lindegren 2009): one is the on-ground LSF-AL, which was used to generate the templates of all the RVS spectra acquired after the decontamination at OBMT 1317; the other is the in-flight LSF-AL, which was calculated using the EPSL data (before the 1317 decontamination). This was also used to generate the template for the RVS spectra acquired before OBMT 1317 to take into account the strong contamination of the optics. No other time variation was considered. The two LSF models are shown in Fig. 8.

The on-ground LSF-AL model profiles that were used after OBMT 1317 were produced for each CCD and FoV. They were based on 15 wavebands, each of 2 nm, covering the range 846-874 nm (there are 15 LSFs per CCD and per FoV in total). Time variations and inter-CCD AC variations were neglected.

The in-flight LSF, before the 1317 decontamination, was derived for each CCD and FoV. It was based on one waveband centred at 858 nm and covering 24 nm (there is one LSF per CCD and per FoV in total). It is independent of wavelength, and time variations and inter-CCD variations were neglected. It was modelled using the PCA as a linear combination of eight basis functions:

$$LSF = \sum_{n=0}^{7} h_n H_n, \tag{5}$$

where h_n are the linear combination coefficients of the basis functions H_n. The basis functions were selected from a set of on-ground basis functions (Lindegren 2009) in order to model the nominal LSF profiles adequately (which were chosen to represent all the data configurations).

Deriving the observed LSF is equivalent to finding the coefficients h_n. To do this, we performed a least-squares fit of the

linear combination of the basis functions convolved with a high-resolution ground-based spectrum of the observed star. Ground-based spectra of about 1200 bright, constant stars were collected for this purpose. They were extracted from the NARVAL and the ESPADON archives[4] to represent the RVS observations of these stars with high fidelity. For the purpose of the LSF-AL calibration, the auxiliary synthetic spectra cannot be used, as they do not provide a sufficiently good match with the real spectra.

The observed RVS spectra (O) were modelled as a convolution of the LSF with the relevant ground-based high-resolution spectrum with high signal-to-noise ratio (S/N) S shifted to the radial velocity of the RVS observations and resampled to the RVS observation,

$$p_\lambda O = S * LSF(h_0, h_1, ..), \tag{6}$$

where: p_λ is a wavelength-dependent scaling factor to allow for flux and also slope and curvature differences; $LSF(h_0, h_1, \ldots)$ is the LSF function, and the asterisk is the convolution operator.

Using Eq. (5), we can rewrite Eq. (6) as

$$p_\lambda O = \sum_{n=0}^{7} h_n X_n, \tag{7}$$

with

$$X_n = S * H_n. \tag{8}$$

As Eq. (7) is of the form of a general linear least-squares fit, we can solve for the linear combination coefficients h_n with X_n as the basis functions of the fit. p_λ is a second-degree polynomial of the form $p_\lambda = a_0 + a_1\lambda + a_2\lambda^2$, and minimises the differences between the auxiliary spectrum and the RVS observation. The a_n are obtained by linear least-squares solving for each sample i in the RVS spectrum the set of equations $(S * H_0)_i/O_i = a_0 + a_1\lambda_i + a_2\lambda_i^2$.

5.5. $G_{\rm RVS}$ zeropoint

The pipeline provides the calibration and temporal evolution of the $G_{\rm RVS}^{\rm int}$ zeropoint ($ZP_{G_{\rm RVS}}$), which is useful for monitoring the contamination during the period covered by the DR2 data set. The $ZP_{G_{\rm RVS}}$ is estimated for each CCD and FoV and for each CaU (30 h). The calibration is made using the HIPPARCOS stars listed in the auxiliary file as reference (Sect. 4.6), for which we have estimated the reference magnitudes $G_{\rm RVS}^{\rm ref}$.

- First, the RVS spectra of the calibration stars (Sect. 4.6) were selected; they were already cleaned and calibrated in wavelength. Spectra with a window size AC < 10 were discarded to limit the flux loss, which typically left about 2000 stars per CaU.
- The ZP is estimated for each spectrum by $ZP^{\rm spec} = G_{\rm RVS}^{\rm ref} + 2.5\log(TotFlux)$, where $G_{\rm RVS}^{\rm ref}$ is the magnitude coming from the auxiliary file, and $TotFlux$ is the integrated flux between 846 and 870 nm, divided by the exposure time for one CCD (4.4 s).
- The zeropoint, $ZP_{G_{\rm RVS}}$, for each CCD and FoV is the median of the values $ZP^{\rm spec}$, obtained with the spectra observed in the relevant configuration. The associated uncertainty is given by their robust dispersion:
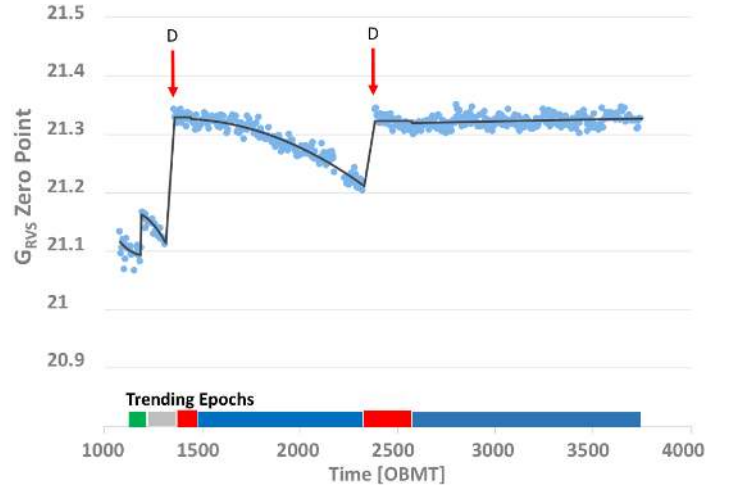
Fig. 9. $G_{\rm RVS}^{\rm int}$ zeropoint temporal evolution (for the CCD in row 6, strip15 and Fov1). The red arrows indicate the decontamination events that resulted in an improvement of the $G_{\rm RVS}^{\rm int}$ zeropoint. After the decontamination at OBMT 2330.6, there is no degradation of the RVS transmission, and the decontamination events become rarer over mission time. The trending epochs are indicated in the same way as in Fig. 4. During the first trending epoch, a wrong catalogue was used, and the zeropoint data obtained in this epoch should be offset by ~ + 0.08 magnitudes.

$\sigma_{ZP} = \frac{P(ZP^{\rm spec}, 84.15) - P(ZP^{\rm spec}, 15.85)}{2}$, where $P(ZP^{\rm spec}, 84.15)$ and $P(ZP^{\rm spec}, 15.85)$ are the 84.15th and the 15.85th percentiles of the distribution of $ZP^{\rm spec}$.

The temporal evolution of the $ZP_{G_{\rm RVS}}$ was modelled with second-degree polynomial trending functions (as was done for the wavelength calibration), see Fig. 9, and the value of the $ZP_{G_{\rm RVS}}$ can be obtained at any transit time. It is used to estimate the magnitude $G_{\rm RVS}^{\rm int}$ (Sect. 6.3).

6. Cleaning and reducing the RVS spectra

This section describes the process responsible for transforming the RVS windows into cleaned and calibrated spectra. This process is applied to the spectra to be used for calibration (i.e. in the workflow `Calibration Preparation` of Fig. 2) using initial, first-guess calibrations, and is then applied to all the spectra, this time using the final calibrations (this is done in the `FullExtraction` workflow of Fig. 2 in charge of preparing the spectra to be used to extract the radial velocities).

The spectra entering this process have already undergone the on-board processing (Cropper et al. 2018), and the IDT window reconstruction (Fabricius et al. 2016; Sect. 3). They have also passed the `Ingestion` step (Sect. 3.1), where all the relevant information for the downstream processing is extracted from the input data.

Figure 10 shows a spectrum at two processing stages: at input, before bias NU correction, and at output, cleaned, and calibrated in wavelength, with the edges cut and normalised.

6.1. Raw spectra cleaning

6.1.1. Electronic bias correction

The first step is to correct the flux in the raw spectrum, still in ADU, for the uniform bias derived from the prescan data and for
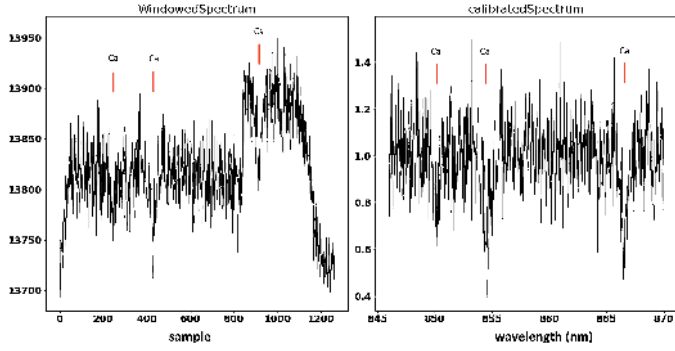
Fig. 10. Different reduction stages of a spectrum from a 4.4 seconds exposure on a single CCD. *Left*: Raw spectrum. Y is in ADU (analogue-to-digital units), X is in samples (one sample is 1 AL × 10 AC pixels), the biasNU effect is visible as an offset after sample 820. The Ca II triplet lines are indicated by red lines. *Right*: Spectrum calibrated in wavelength, normalised to the pseudo-continuum, and samples on the leading/trailing edges are discarded. The wavelength range of the calibrated spectra is [846–870] nm. The atmospheric parameters that have been determined by DetermineAP (Sect. 6.5) and associated with the star are $T_{\rm eff}$ = 6000 K, log g = 3.5, and [Fe/H] = 0. The internal magnitude is $G_{\rm RVS}^{\rm int}$ = 10.2 (Sect. 6.3), and the $S/N \sim 9$.

the non-uniform bias offset (Sect. 5.1). The bias NU correction procedure involves the reconstruction of the readout timing for each sample of the CCD, see Hambly et al. (2018). In Fig. 10 we show a spectrum that is affected by bias NU.

During the bias correction procedure, the saturated samples (ADU = 65 535) are identified and flagged. The spectra containing saturated samples are then excluded from the processing, which removes ~0.3% of the spectra.

6.1.2. Gain and dark current correction

After the bias removal, the spectral flux is transformed into photoelectrons using on-ground gain values. The nominal (on-ground) dark current level of 2.80×10^{-4} electrons/pixel/s is then subtracted.

6.1.3. Background removal

The background level is subtracted from each spectrum using the ScatterMap obtained as described in Sect. 5.2. Based on the AC location of the RVS window, and the *Gaia* Solar phase angle, the position of the window in the ScatterMap is found and the corresponding straylight level is subtracted from the spectrum. The spectra falling in ScatterMap regions where the straylight level is higher than 100 e⁻ pix⁻¹ in the 4.4 seconds exposure are flagged and excluded from the processing. In total, another ~0.3% of spectra have been excluded for this reason.

The straylight level is assumed here not to vary with time. The consequence of neglecting time variations, and using, for the entire data set, the ScatterMap obtained using the data obtained only during EPSL, is that, in general, the level of straylight subtracted from the spectra is too high (see Fig. 6), and many spectra, at the faint magnitude end had a negative total flux after the straylight subtraction. They are excluded from the processing. Approximately ~9% of spectra are excluded because of negative total flux.

Note that only the offset is subtracted, but the noise in the signal, induced by the straylight, can not be subtracted and has the effect of reducing the effective S/N of the spectra. The S/N degradation is more important for the faint stars. As an example,

considering only poisson and readout noises, the S/N per sample in one CCD spectrum of a $G_{\rm RVS}$ = 12 star, would be of ~4.5 without straylight and it is degraded by a factor of 2, when the straylight level is 7 e⁻ pix⁻¹ and by a factor of 3 when the straylight level is of 22 e⁻ pix⁻¹ (these are typical values of the ScatterMap in Fig. 5).

6.1.4. Cosmetic defects

The list of pre-launch cosmetic defects is taken from a list provided by the CCD manufacturers, e2v[5]. No attempt has been made to detect new defects in this pipeline version, nor to measure the column response non uniformity (CRNU). The RVS spectra with pre-launch CCD defects are flagged and excluded from the processing, and ~1% of the spectra have been excluded. Except for the defects, the CCD column response is considered to be uniform. This approximation is justified for the purposes of this pipeline version by the pre-launch measurements showing that the CCD CRNU is typically less than 1%.

6.1.5. Spectra extraction and cosmic ray removal

The 2D window spectra are extracted and corrected for cosmic rays using the Horne optimal extraction algorithm (Horne 1986). The 1D spectra are extracted on board (Cropper et al. 2018). The cosmic ray correction algorithm compares the three CCD spectra of each transit to identify and correct the cosmic ray hits. The spectra are calibrated in wavelength (Sect. 6.2) and normalised to their pseudo-continuum (Sect. 6.4), in order to be directly comparable. The wavelength samples of the 3 spectra are compared to detect the outliers above 5 σ from the median level and flagged as cosmic ray hits. The flux in the cosmic ray samples is substituted with interpolated values of the nearest samples from the spectra of the other two CCDs.

6.2. Spectra wavelength calibration

This algorithm computes the wavelength scale that is to be applied to each spectrum. The wavelength calibration coefficients C_{mn} are computed at any time and position, using the trending functions (Sect. 5.3, Fig. 4). Each spectrum sample has associated its FoV (η, ζ) coordinates and uncertainty (Sect. 3.1). The wavelength λ_s of the sample s is

$$\lambda_s = C_{00} + C_{10}\eta' + C_{20}\eta'^2 + C_{01}\zeta' + C_{11}\eta'\zeta', \qquad (9)$$

where η' and ζ' are the shifted and scaled field angles as in Eq. (2).

The uncertainties on λ_s resulting from the propagation of the uncertainties on the coefficients and on the field angles are found to be underestimated. We therefore estimated the wavelength calibration uncertainty a posteriori based on the median dispersion of the single-transit $V_{\rm R}^{\rm t}$ results for the bright constant stars in the auxiliary data. This uncertainty is estimated to ~400 m s⁻¹ (see Sect. 7.8), corresponding to ~0.0011 nm. It is propagated to the single-transit $V_{\rm R}^{\rm t}$ uncertainty, but is not used to compute the uncertainty on the all-transits combined $V_{\rm R}$, published in DR2. Instead, the uncertainty is estimated using the dispersion among the individual single-transit $V_{\rm R}^{\rm t}$.

After applying the wavelength calibration, the wavelength range of each spectrum is reduced to avoid including the wings of

---

[5] e2v is called Teledyne e2v as of March 2017.

the band-pass filter, which might disturb the V_R cross-correlation algorithms. All samples falling outside the wavelength range (846–870 nm) are discarded. Figure 10 (right) shows an example of a calibrated spectrum.

6.3. Internal magnitude estimation

The magnitude G_{RVS}^{int} is estimated for each CCD spectrum by

$$G_{RVS}^{int} = -2.5 \log(TotFlux) + ZP_{G_{RVS}}, \qquad (10)$$

where
- $TotFlux$ is the total flux from the star measured in the spectrum (cleaned as described in Sect. 6.1), between 846 and 870 nm, divided by the CCD exposure time.
- $ZP_{G_{RVS}}$ is the zeropoint (Sect. 5.5) computed at the transit time in the same CCD and FoV of the observed spectrum. The zeropoint at the transit time is computed using the trending function.

The estimation of $TotFlux$ on the RVS spectra is limited by the missing appropriate background level estimation (Sect. 5.2), which needs to be subtracted to estimate the flux level of the source, and by the LSF-AC calibration, which is necessary to estimate the source flux loss outside the readout window. As a result, the magnitude G_{RVS}^{int} that is produced has an insufficient quality for publication in DR2. Nevertheless, it is used to exclude stars from DR2 whose flux is too low (i.e. with $G_{RVS}^{int} \geq 14$).

6.4. Continuum normalisation

The spectrum is normalised to its pseudo-continuum, which is determined using a polynomial fitting. First the spectrum is fitted with a polynomial of degree two. The stellar lines are iteratively rejected by sigma clipping, that is, the pseudo-continuum is set by the pixels within the given interval $[-3, +10]\,\sigma$ from the polynomial. The polynomial fit and sigma clipping are iterated until convergence on the status of the pixels, as to whether they belong to the pseudo-continuum or not. If convergence is reached, the spectrum is divided by the polynomial.

Spectra with a strong gradient in their pseudo-continuum (e.g. cool stars with a molecular band) are difficult to normalise in this way. If a positive gradient is detected in the spectrum (cool stars typically exhibit positive gradients) and the windows of the spectrum are not truncated (because when the window is truncated, there is some flux loss outside the window that can result in a gradient), or if the spectrum is too noisy, no pseudo-continuum is computed, and the spectrum is divided by a constant that is the 90th percentile of the fluxes.

6.5. Atmospheric parameters and template selection

The synthetic spectrum that is associated with each RVS spectrum is selected among the 5256 synthetic spectra of the auxiliary spectral library (Sect. 4.4). The selection is based on the weighted minimum distance between the atmospheric parameters T_{eff}, $\log g$, and [Fe/H] that are associated with the star and those associated with the synthetic spectrum, T_{eff}^{tpl}, $\log g^{tpl}$ and [Fe/H]^{tpl}. The distance to minimise is:

$$D = \frac{|T_{eff} - T_{eff}^{tpl}|}{100} + \frac{|\log g - \log g^{tpl}|}{3} + \frac{|[Fe/H] - [Fe/H]^{tpl}|}{2} \qquad (11)$$

where T_{eff} is in K. The weights of 100, 3, and 2 are an ad hoc estimate to give more weight to a difference in T_{eff} (on which the morphology of the spectrum depends most) than in $\log g$ or [Fe/H].

An important task of the pipeline is then to associate appropriate parameters with the stars observed by the RVS. When the atmospheric parameters of the star are known from the literature and are stored in the auxiliary file (Sect. 4.3), they are associated with the star. When $\log g$ or [Fe/H] are not in the auxiliary parameter file, the default values $\log g = 4.5$ and [Fe/H] $= 0$ are set.

The majority of the RVS observations, about 85% in this release, do not have auxiliary associated parameters. For these stars, a pipeline module called `DetermineAP` is used to estimate their atmospheric parameters. It works by cross-correlating (Pearson correlation in direct space) the observed RVS spectrum with the 28 templates described in Sect. 4.5. The template that gives the highest cross-correlation peak provides the atmospheric parameters to be associated with the star[6]. The atmospheric parameters are determined for each transit of the star, and at the end, the parameters that were found for the majority of the transits are associated with the star (Sect. 6.6).

To constrain the results of the module `DetermineAP` and for a rough estimate on the uncertainties on the parameter determination and the consequent uncertainties induced on the radial velocity estimations, we computed off-line the atmospheric parameters using `DetermineAP` for a sample of stars for which we also had the parameters from the literature (Sect. 4.3). A poor estimate of the atmospheric parameters implies that a poor template is associated with the stars and may result in larger systematic shifts in the radial velocity measurements. We then obtained the radial velocity of the stars using the template corresponding to the parameters found by `DetermineAP` and compared them with the radial velocity obtained using the template corresponding to the parameters from the literature.

For this test, the RVS spectra of stars with parameters from the PASTEL compilation (Soubiran et al. 2010), which are included in the auxiliary file in Sect. 4.3, were selected from the DR2 data set and processed in `DetermineAP`: 13 000 stars were observed in ~150 000 transits.

In general, given that T_{eff} is the driving parameter for the presence and strength of spectral lines, the mismatches on T_{eff} between the stars and the templates are expected to produce larger uncertainties on the V_R determination than the mismatches on [Fe/H] and on $\log g$. All the transits were processed with `DetermineAP`. We obtained for each of the 28 templates the residuals $\Delta T_{eff} = T_{eff}^{tpl} - T_{eff}^{cat}$, between T_{eff}^{tpl} from `DetermineAP` and T_{eff}^{cat} taken from the PASTEL compilation, and the residuals $\Delta V_R^t = V_R^{tpl} - V_R^{cat}$, between the radial velocity obtained using a template with the parameters from `DetermineAP` (V_R^{tpl}), and the radial velocity obtained using a template with the parameters from PASTEL (V_R^{cat}). The results of the test are shown in Table 2, where
- $Md(\Delta T_{eff})$ is the median of the residuals ΔT_{eff} and indicates the accuracy of the estimation of the effective temperature T_{eff}^{tpl} of the star obtained by `DetermineAP` (assuming that the PASTEL T_{eff} represents the actual star temperature);

---

[6] At fainter magnitudes, $G_{RVS} > 12$, which are not included in this data release, the module does not attempt the cross-correlation, and default parameters are assigned.

Table 2. Performance of the module `DetermineAP` and template mismatch errors.

Template			$Md(\Delta T_{\rm eff})$	$\sigma(\Delta T_{\rm eff})$	$Md(\Delta V_{\rm R}^{\rm t})$	$\sigma(\Delta V_{\rm R}^{\rm t})$	nb
K	dex	dex	K	K	km s^{-1}	km s^{-1}	
3100	3.0	0.0	-1959	3870	1.23	116.46	113
3100	3.0	-1.5	-1262	1465	1.23	2.67	61
3500	3.0	0.0	-1435	10 896	-0.07	27.40	58
3500	3.0	-1.5	-1413	814	-0.14	1.35	62
4000	3.0	0.0	88	500	-0.05	0.39	1251
4000	3.0	-1.5	-200	426	-0.1	0.57	261
4500	3.0	0.0	117	324	0.0	0.33	4857
4500	3.0	-1.5	113	302	0.12	0.4	1789
5000	3.0	0.0	145	222	-0.03	0.28	17 829
5000	3.0	-1.5	-1050	713	0.05	0.30	5622
5500	3.5	0.0	-480	314	-0.07	0.30	53 491
5500	3.5	-1.5	-345	791	-0.06	0.49	853
5500	4.5	0.0	145	302	-0.14	0.3	6294
6000	3.5	0.0	-508	217	0.06	0.66	33 297
6000	3.5	-1.5	674	726	-0.24	0.52	1185
6500	3.5	0.0	-280	211	-0.12	1.47	11 805
6500	3.5	-1.5	-425	845	-1.0	0.92	203
7000	3.0	0.0	-3.0	591	-3.28	11.08	3499
7000	3.0	-1.5	-2348	5886	4.31	8.50	234
7500	3.0	0.0	200	743	-9.02	19.40	919
8000	3.0	0.0	-1247	1099	5.57	2.58	252
9000	3.0	0.0	-1680	9708	0.14	14.66	84
10 000	3.0	0.0	-4380	4883	0.74	2.90	243
15 000	4.0	0.0	4714	5957	-3.09	9.14	4403
20 000	4.0	0.0	-1100	4357	-2.24	9.22	91
25 000	4.0	0.0	2753	9574	-8.79	48.92	77
30 000	4.0	0.0	24 340	12 273	3.17	285.1	17
35 000	4.0	0.0	5000	12 350	-1.16	46.90	122

Notes. Each of the 28 templates in the first column is identified with its atmospheric parameters ($T_{\rm eff}^{\rm tpl}$, $\log g^{\rm tpl}$, and [Fe/H]$^{\rm tpl}$); nb is the number of transits for which the template has been selected.

- $\sigma(\Delta T_{\rm eff})$ is the robust standard deviation of the residuals $\Delta T_{\rm eff}$ and indicates the precision of the estimation of the effective temperature of the star by `DetermineAP`. The robust standard deviation is defined as

$$\sigma(\Delta T_{\rm eff}) = \frac{P(\Delta T_{\rm eff}, 84.15) - P(\Delta T_{\rm eff}, 15.85)}{2}, \qquad (12)$$

  where $P(\Delta T_{\rm eff}, 15.85)$ and $P(\Delta T_{\rm eff}, 84.15)$ are the 15.85*th* and 84.15*th* percentiles of the distribution of the residuals $\Delta T_{\rm eff}$, respectively.
- $Md(\Delta V_{\rm R}^{\rm t})$ is the median of the residuals $\Delta V_{\rm R}^{\rm t}$ and indicates the shift between the radial velocity obtained with the template found by `DetermineAP` ($V_{\rm R}^{\rm tpl}$) and the radial velocity obtained with the template with the PASTEL parameters ($T_{\rm eff}^{\rm cat}$); it indicates the systematic uncertainty in the estimation of $V_{\rm R}^{\rm tpl}$ that is introduced by the template mismatch.
- $\sigma(\Delta V_{\rm R}^{\rm t})$ is the robust standard deviation of the residuals $\Delta V_{\rm R}^{\rm t}$ calculated as in Eq. (12) and indicates the random uncertainties in the estimation of the $V_{\rm R}^{\rm tpl}$ that are introduced by the template mismatch.

The radial velocities obtained using the template with the parameters from `DetermineAP` and those obtained with the template with the parameters from the catalogue are in good agreement for $4000 \le T_{\rm eff}^{\rm tpl} \le 6500$ K, despite sometimes large difference in $\Delta T_{\rm eff}$. Intermediate-temperature star spectra are dominated by the Ca II triplet, and provided the template and the star temperature are both in this range, the template mismatch error in

the estimation of $V_{\rm R}^{\rm tpl}$ is small. RVS spectra of stars from spectral type B2 to M6 are shown in Cropper et al. (2018, Fig. 17). The `DetermineAP` results degrade for the cool stars where the molecular TiO band appears, and for the hotter stars where in addition to $T_{\rm eff}$, $\log g$ mismatches play an important role in the morphology of the spectrum because of the appearance of the H lines, which are more pronounced in dwarfs near to the Ca II lines. A description of a more detailed study on the template mismatch errors is in preparation by Blomme et al. (in prep.).

In the next data release, the spectroscopic pipeline will include the atmospheric parameters produced with *Gaia* data and more appropriate templates will be assigned to the stars to improve the $V_{\rm R}$ estimation of hot and cool stars.

6.6. Combine template atmospheric parameters

`DetermineAP` estimates the atmospheric parameters for each transit of the star. Then, for each star, the parameters found most frequently are selected and the synthetic spectrum with these parameters is used to derive the $V_{\rm R}$ of the star.

These parameters are published in DR2 (Sect. 2) and provide information on the synthetic spectrum that was used to obtain the star $V_{\rm R}$.

The transit information is combined in the workflow `Combine Template` (shown in Fig. 2). Even though it contains only simple functionality, this workflow is technically important because it is the first that uses information from all the transits. It runs after all the data were processed by the other workflows, and together with the following `STAMTA` workflow, works per source.

6.7. Template generation

The `GenerateTemplate` module starts from a synthetic spectrum (Sect. 4.4). The spectrum is convolved with the instrumental profile (Sect. 5.4), using the LSF-AL corresponding to the CCD and FoV coordinates of the spectrum. Then, the spectrum is resampled to a wavelength step of 0.00747 nm, which is ten times finer than the nominal resolution element of the RVS. The module can convolve the spectrum with a rotational profile, but for DR2 , the projected rotational velocity was set to zero.

The wavelength range for the template is [843, 873] nm. This is slightly larger than that of the object spectrum in order to ensure than the shifted template in the velocity range ± 1000 km s^{-1} can always be resampled on the object spectrum. The template spectrum is then normalised (Sect. 6.4) in the same way as the observed spectra.

## 7. Deriving the single-transit radial velocity

The three CCD spectra corresponding to each FoV transit of a star, with the corresponding synthetic template spectra associated, are analysed by the single-transit analysis (`STA`) set of modules to determine the radial velocity of the star at each FoV transit. This is the radial velocity of the star with respect to the *Gaia* satellite, referred to as the spectroscopic radial velocity, $sV_{\rm R}^{\rm t}$.

The $sV_{\rm R}^{\rm t}$ computed by STA is then corrected for the *Gaia* barycentric velocity to provide the radial velocity with respect to the barycentre of the solar system,

$$V_{\rm R}^{\rm t} = sV_{\rm R}^{\rm t} + BaryVeloCorr. \qquad (13)$$

The $V_{\rm R}^{\rm t}$ obtained at each transit of the source are then combined (Sect. 8) to provide the median barycentric velocity of the source $V_{\rm R}$. In DR2, $V_{\rm R}$ are provided, but not $V_{\rm R}^{\rm t}$.
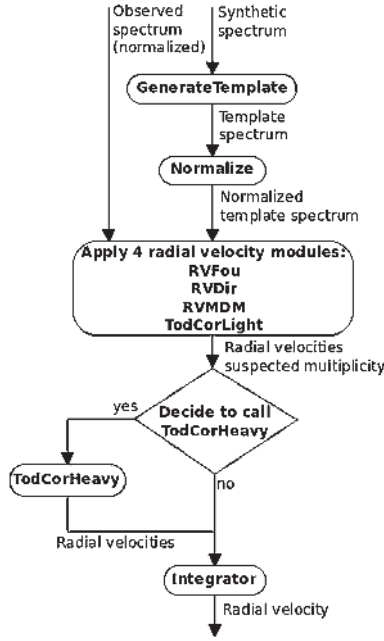
Fig. 11. Flowchart showing the main steps of the Single Transit Analysis pipeline.

7.1. Single and double stars per transit analysis

Figure 11 gives a flowchart of the modules that make up the STA part of the processing. Names and acronyms on this figure are explained below. The main inputs to the STA are the observed cleaned and calibrated spectra (a set of three spectra for one transit) and the synthetic spectrum (Sect. 4.4) that is chosen to correspond to the astrophysical parameters determined in Sect. 6.6. From the synthetic spectrum, the templates are generated (Sect. 6.7) and normalised (Sect. 6.4) in the same way as the observed spectra.

The normalised observed spectra and template are then processed in sequence by four or five different radial velocity modules. RvDir (Sect. 7.3) uses the Pearson correlation coefficient, RvFou (Sect. 7.2) the standard method of cross-correlation in Fourier space, and RVMDM (Sect. 7.4) the minimum distance method. In order to search for binaries, TodCorLight (Sect. 7.5) applies a technique equivalent to a 2D cross-correlation, assuming that the astrophysical parameters of the secondary are the same as those of the primary. If TodCorLight finds that the spectrum is double-lined, a range of astrophysical parameters is explored by TodCorHeavy (Sect. 7.5). For this range, the set of 28 templates listed in Table 2 is used. The four or five radial velocity modules each provide their separate results. These are then passed through Integrator (Sect. 7.6), which combines them into a single result.

Each radial velocity module was implemented independently, with only a few common rules. This approach was chosen to guarantee that the implementation contains the best features appropriate to it, unhampered by too many constraints on the programming details.

All of the radial velocity modules handle the three CCDs corresponding to one transit, and in the process of determining the spectroscopic radial velocity of the star during the transit, sV_R^t, they also determine the spectroscopic radial velocity in each of the three CCDs (sV_R^{CCD}). All modules have been designed with the ability to handle the case where information from one or two CCDs is missing.

7.2. RVFou

RvFou implements the standard cross-correlation method for Doppler-shift measurement using the Fourier transform (David et al. 2014, Sect. 2.3.1). For cross-correlation, the observed spectrum and the template must be resampled on a grid with constant step in log(wavelength) and normalised to their continuum. Furthermore, ideally, both of their edges should be featureless, but this cannot be guaranteed in an automated environment. Therefore, to avoid problems with the Gibbs effect on the Fourier transforms, we apply edge apodisation (5%) with a cosine bell function after normalisation. This ensures that the spectra behave like continuous periodic functions for the Fourier transform.

Fourier transforms are computed using the fast Fourier transform (FFT) algorithm found in the APACHE COMMONS MATHS java library. The FFT technique requires that the number of flux bins in the two data segments (i.e. observed spectrum and template) be a power of two; within this constraint, the number of samples was optimised and chosen to be a factor of twice the power of two, which gives the closest (but slightly larger) number of wavelength bins in the observed spectrum.

The maximum position of the cross-correlation function (CCF) defines the value of the spectroscopic radial velocity for the RVS CCD spectrum in question sV_R^{CCD}. It is estimated by fitting a parabola to the peak of the function. Since we cannot refine the wavelength grid at will, we locate the peak by combining the results of a three- and four-point parabola fit to reduce the impact of discretisation, as proposed by David & Verschueren (1995).

Internal uncertainties on the radial velocity measurements are estimated using the expression proposed by Zucker (2003, Sect. 2.3), that is,

$$\sigma^2_{sV_R^{CCD}} = -\left(N \; \frac{C''}{C} \; \frac{C^2}{1-C^2} \right)^{-1} , \tag{14}$$

where N is the number of wavelength bins, C represents the value of the CCF, and C'' the value of its second derivative, both evaluated at its maximum.

These operations are carried out separately on the three CCD spectra recorded in a transit to obtain three sV_R^{CCD}; for a transit-combined sV_R^t measurement and uncertainty estimate, the three correlation functions are merged into one, as proposed in Zucker (2003).

7.3. RVDir

In the module RvDir, the radial velocity of the input spectrum on a single CCD, sV_R^{CCD}, is obtained from the maximum value of the Pearson correlation function (PCF). The PCF is defined as

$$C_{pc}(v) = \frac{\sum_{n=1}^{N}(f_n - \bar{f})(t_n(v) - \bar{t}(v))}{\sqrt{\sum_{n=1}^{N}(f_n - \bar{f})^2 \sum_{n=1}^{N}(t_n(v) - \bar{t}(v))^2}} , \tag{15}$$

where f_n is the flux in wavelength bin n of the observed spectrum, and $t_n(v)$ is the flux of the template spectrum shifted with radial velocity v and resampled to the wavelength grid of the observed spectrum. Object and template spectrum are both normalised to the continuum. N is the number of wavelength bins of the object spectrum; $\bar{f}$ and $\bar{t}(v)$ are the averaged fluxes of the object spectrum and of the shifted template spectrum, respectively. The radial velocity shift of the input spectrum corresponds to the maximum of the PCF.

For each CCD, the PCF is computed in a first step on a coarse velocity grid ranging from -1000 km s^{-1} to $+1000$ km s^{-1} with a step $\Delta v_{\text{coarse}} = 10$ km s^{-1} and its highest value is identified. In a second step, the PCF is computed with a finer velocity grid $\Delta v_{\text{fine}} = 0.5$ km s^{-1} over a reduced range that extends to ± 50 km s^{-1} from the maximum of the coarse array. Thus the PCF for each CCD is obtained as an array defined on an irregular grid (steps are either 10 km s^{-1} or 0.5 km s^{-1}).

In order to achieve a sub-step accuracy, a parabola is fitted through the highest three points of the "fine" PCF sampling. The maximum of that parabola defines the value of $sV_{\text{R}}^{\text{CCD}}$ for the spectrum at hand.

An internal uncertainty estimate is obtained using the same expression as for RvFou, specifically Eq. (14), where C now represents the value of $C_{\text{pc}}(v)$ and C'' its second derivative with respect to v, both evaluated at the radial velocity $sV_{\text{R}}^{\text{CCD}}$ corresponding to the maximum value of the PCF.

To construct the combined PCF, we merge the three single-CCD velocity grids into one that contains all of their points. There may be grid points in the combined grid for which fewer than three values are directly available from the previously calculated arrays; such missing values are filled in by linear interpolation. Finally, the combined PCF is obtained as the mean of the former three PCFs. The same procedure is applied to the combined PCF to obtain sV_{R}^{t} and $\sigma_{sV_{\text{R}}^{\text{t}}}$.

7.4. RVMDM

The RVMDM module determines the radial velocity shift that minimises the "distance" between the observed spectrum and the template that was Doppler-shifted and resampled to the observed wavelength grid. The theoretical background for this technique is discussed in David et al. (2014, Sect. 2.3.3 and Appendix B). For convenience, both spectra are normalised. Using the same notation as in Sect. 7.3, we define the χ^2 distance function as

$$C_{\text{md}}(v) = \sum_{n=1}^{N} \frac{(f_n - t_n(v))^2}{\sigma_n^2}, \tag{16}$$

where σ_n is the uncertainty on the observed flux.

The function is first evaluated on a coarse velocity grid (-1000 to $+1000$ km s^{-1}, with a step of 20 km s^{-1}). A parabola is fitted to the lowest three points, its minimum providing the Doppler shift as measured on that grid. Then the step is halved; we define a new search range centred on the grid point nearest to the previous measurement, with twice the previous step size as width, and repeat the Doppler-shift measurement. This refinement is iterated eight times so that the smallest grid-step size is about ~ 0.08 km s^{-1}, which corresponds to approximately 1% of a wavelength bin.

An internal uncertainty estimate on this measurement is obtained by starting from the best-fit value $C_{\text{md,min}}$ and changing the velocity until $C_{\text{md}}(v) = C_{\text{md,min}} + 1.0$, noting v_1 and v_2 for the velocities where this occurs. The threshold 1.0 is the $\Delta\chi^2$ value that defines a 68.3% confidence region for a fit with one parameter (v). We define the error bar as the maximum of $|v_1 - sV_{\text{R}}^{\text{t}}|$ and $|v_2 - sV_{\text{R}}^{\text{t}}|$ to mimic a classical one-σ error bar.

The function $C_{\text{md}}(v)$ for the combined CCDs is obtained trivially by extending Eq. (16) to a sum over the three CCDs, but this requires a common velocity grid whose definition is complicated by the occurrence (in the single-CCD grids) of nine different step sizes in regions that need not be common, and by the fact that the minimum of the combined $C_{\text{md}}(v)$ may occur in a region where none of the single-CCD grids was refined. Thus the

combination may require additional shift and resampling operations on part of the data, as well as careful tracking of all v-values involved. At the end of the calculation, both $sV_{\text{R}}^{\text{CCD}}$ for each CCD, and the transit combined sV_{R}^{t} are provided with their uncertainties.

7.5. TodCor

The three STA modules described above (Sects. 7.2–7.4) are designed to measure the unique Doppler shift exhibited by single-lined spectra. However, gravitationally bound systems (binaries or multiple stars) might generate composite spectra; this is also the case for accidental confusion on the RVS line of sight. Moreover, two FoVs are observed simultaneously in the same focal plane, which may give rise to false composite spectra. TodCor is dedicated to the detection of composite double-lined spectra.

2The module was initially designed as an implementation of the TodCor method first described by Zucker & Mazeh (1994). However, our STA algorithm, still called TodCor, evolved from this initial design to implement the following functionalities: i) derive the uncertainties on the measured radial velocities and on the brightness ratio, and ii) assign a probability that the observed spectrum is double-lined rather than single-lined. A more detailed description of this new algorithm is in preparation by Damerdji et al. (in prep.). Its actions can be described as follows:

– Process the input spectrum as a single-star spectrum for deriving a single-star model goodness-of-fit.
– Process the input spectrum as a double-star spectrum for deriving a double-star model goodness-of-fit.
– Compare the goodness-of-fit, check the significance of the difference (taking into account the different degrees of freedom) and decide the nature of the observed spectrum. Thereafter, output the computed velocities (V_R or V_{R1} and V_{R2}) of the most likely model.

For each transit and for each model, an approximate solution is found for the individual and combined CCD spectra by assuming equal flux uncertainties. The input spectra to this method (observed and template) have normalised continua. A minimum χ^2 is derived in Fourier space, thanks to the Parseval equality (Press et al. 1993, Sect. 12.1).

The second part of the algorithm refines the approximate solution taking into account the flux uncertainties. It consists of a Levenberg-Marquardt minimisation (Press et al. 1993, Sect. 15.5.2). For the single-star model, the observed spectrum is assumed to be $S(\lambda_i) = P_n(\lambda_i) * T_1(\lambda_i, V_{\text{R1}}, v_{\text{rot1}})$, while it becomes $S(\lambda_i) = P_n(\lambda_i) * [T_1(\lambda_i, V_{\text{R1}}, v_{\text{rot1}}) + \alpha\, T_2(\lambda_i, V_{\text{R2}}, v_{\text{rot2}})]$ in the double-star model, where $P_n(\lambda)$ is a polynomial function with degree n, linked to the magnitude of the spectrum ($n \leq 2$). Such a polynomial function is added to the fit to model the effect of tilted spectra in window or stellar reddening. The brightness ratio α is part of the parameters of the fit, while the polynomial coefficients are optimised.

The refined solution (single- or double-star model) contains the set of parameters, their uncertainties, and the goodness-of-fit χ_D^2. The uncertainties on the parameters are given by the diagonal elements of the variance-covariance matrix (Press et al. 1993, Sect. 15.5.2).

The TodCor-like algorithm comes in two flavours: TodCorLight, and TodCorHeavy.

– TodCorLight is responsible for the provisional identification of the composite spectra, which are flagged as suspected multiple. It assumes that the candidate secondary star has

atmospheric parameters identical to the primary star (i.e. both templates are the same).

– `TodCorHeavy` is triggered if `TodCorLight` decides that the star is a suspected multiple. `TodCorHeavy` loops over a list of atmospheric parameters for both the primary and the secondary stars by assuming these parameters to be equal for better numerical stability. This will change in the future version of the pipeline, when Todcor will use all the transits together. `TodCorHeavy` returns the atmospheric parameters of the primary-secondary pair that best fit the observed spectrum, together with the derived radial velocities and rotational velocities, their uncertainties, and the brightness ratio (α).

The decision on the binary nature of the observed star is based on the comparison of χ_S^2 and χ_D^2. $\frac{\chi_S^2 - \chi_D^2}{\chi_D^2}$ is assumed to follow an F-distribution. The binary model is accepted above a threshold of 0.9 and 0.99865 of the F-distribution CDF for `TodCorLight` and `TodCorHeavy`, respectively.

For this first version of `TodCor`, some additional detection limits were estimated off-line to minimise false detections. The binary model was not accepted, and the star in this transit was considered as single in the following cases:

– when the separation of the two component radial velocities is $|V_{R1} - V_{R2}| < 20$ km s^{-1} or $|V_{R1} - V_{R2}| > 500$ km s^{-1};
– when $\alpha < 0.25$ and $|V_{R1} - V_{R2}| > 40$ km s^{-1};
– when $\alpha < 0.35$ and $30 < |V_{R1} - V_{R2}| < 40$ km s^{-1};
– when $\alpha < 0.5$ and $20 < |V_{R1} - V_{R2}| < 30$ km s^{-1};

where α is still the brightness ratio.

7.6. Integrator

The `Integrator` is responsible for combining the results obtained by the various methods into a single result and provides a single spectroscopic radial velocity, sV_R^t, estimate for each star transit. For single stars, there are three different determinations of the sV_R^t of the star by the modules `RvDir`, `RvFou,` and `RVMDM`. As discussed in David et al. (2014, Sect. 6.3), even when in some cases a method provides better results than another, no simple general conclusion can be drawn about the algorithms' performance versus astrophysical parameters. Thus, the simplest approach was chosen, and the final sV_R^t was computed as the median of the sV_R^t derived by modules `RvDir`, `RvFou,` and `RVMDM`. The internal uncertainty on the measurement selected as the median was then also the internal uncertainty on the final sV_R^t, unless there are only two valid radial velocities, in which case their uncertainties were quadratically averaged.

For binary stars, the situation was simple since only one module, `TodCorHeavy`, determined the radial velocity. The `Integrator` module checked whether this module was launched and had provided two radial velocities (and two uncertainties on them), one per component. When this was the case, `Integrator` stored them in the data model.

The results for binary stars are not published in DR2, but are used to exclude the binary candidates from the data release.

7.7. Flagging

For the sV_R^t, the flag *isAmbiguous* is set by each of the methods (Sects. 7.2, 7.3 and 7.4). For this purpose, all radial velocity differences between the three CCDs are checked. If these are all smaller than 10 km s^{-1}, the transit is not ambiguous. Otherwise, if at least one radial velocity difference is significantly larger than the uncertainty on that difference, the *isAmbiguous* flag is set to
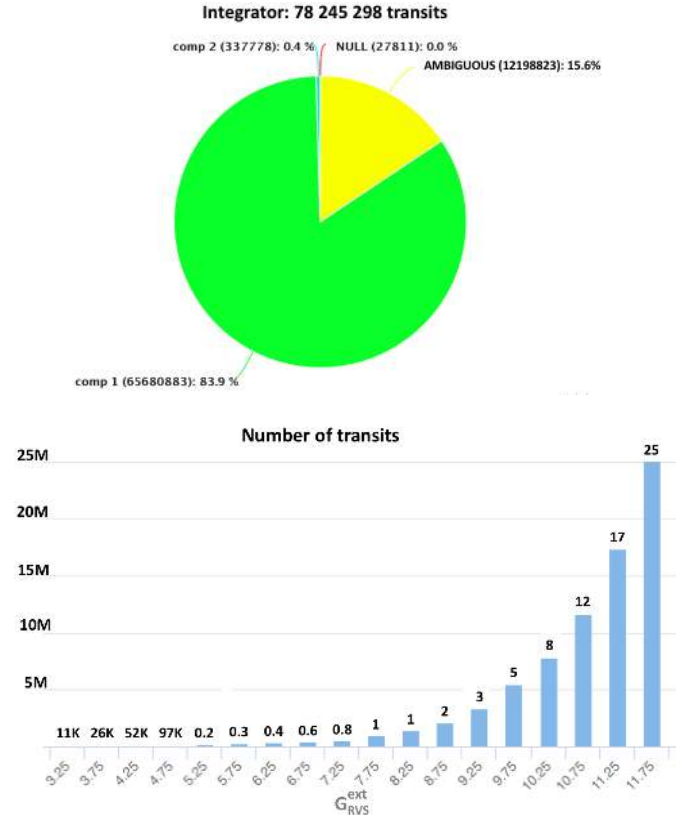
Fig. 12. STA statistics. The total number of transits for which a sV_R^t is obtained is ∼78 million (with ∼28 000 null, because invalid in the three STA methods `RvFou`, `RvDir,` and `RVMDM`). *Top:* Pie chart showing that 16% of the sV_R^t obtained have been flagged as ambiguous, 84% of the transits have been detected as single components (comp 1), and in 0.4% of the transits, multiple-lines have been detected and the transit was flagged as two components (comp 2). *Bottom:* Histogram of the distribution of the number of transits as a function of G_{RVS}^{ext}.

true. A flag *isValid* is set to false by any of the modules if a computational problem is encountered. `Integrator` combines these results and sets *isAmbiguous* in the case of single stars when at least two of the three methods have provided ambiguous results, and when two of the three methods have provided an invalid result. It sets the sV_R^t to null when all three methods have provided an invalid result.

7.8. Automated verification of `STA`

Figure 12 shows the number of transits processed by STA and their distribution as a function of G_{RVS}^{ext}. AVSTA (Fig. 2) verifies the quality of the single-transit radial velocities V_R^t (i.e. sV_R^t corrected for the barycentric velocity as in Eq. (13)). For this purpose, a verification dataset is extracted from the ∼78 million of V_R^t obtained by the pipeline (Fig. 12), which contains only the stars belonging to the auxiliary radial velocity catalogues (Table 1) and covers the magnitude range of the DR2 data (Fig. 3). These stars are expected to be constant. The radial velocities obtained by the pipeline (V_R^t) are compared with those in the catalogues (V_R^{ref}).

7.8.1. Comparing the results of the `STA` methods

The first check of `AVSTA` is to ensure that the radial velocities sV_R^t obtained by the three `STA` methods that are combined by

Integrator (Sect. 7.6) do not present significant systematic differences that might degrade the combined result. The median (Md) of the differences between the sV_R^t obtained in the three methods ($\Delta sV_R^t = sV_R^{method1} - sV_R^{method2}$) is computed for the stars in the validation data set. The following results confirm that the systematic differences between the STA methods are not significant:

RvDir−RvFou : $Md(\Delta sV_R^t) = -0.02$ km s^{-1};
$\quad \sigma(\Delta sV_R^t) = 0.10$ km s^{-1};

RvDir−RVMDM : $Md(\Delta sV_R^t) = -0.05$ km s^{-1};
$\quad \sigma(\Delta sV_R^t) = 0.12$ km s^{-1};

RvFou−RVMDM : $Md(\Delta sV_R^t) = -0.04$ km s^{-1};
$\quad \sigma(\Delta sV_R^t) = 0.16$ km s^{-1}.

7.8.2. Monitoring the quality of the upstream processing

The verification of the radial velocities permits the indirect verification of the performance of the upstream processing, in particular, of the wavelength-calibration performance. For this check, the subset of the verification-dataset containing only the stars belonging to CU6GB-cal (Table 1) is used. The V_R^{ref} of the stars in CU6GB-cal are assumed to be perfectly determined (their uncertainties of <0.1 km s^{-1} are neglected), and being used to calibrate the wavelength-calibration zeropoint, should show no systematic shift with the RVS velocities. Any difference is then introduced by the processing, including the upstream AGIS pipeline on which the spectroscopic pipeline is dependent (Sect. 3).

The precision of the single-transit radial velocities, indicative of random uncertainties, is quantified by the robust standard deviation of the residuals: $\sigma(\Delta V_R^t)$ (see Eq. (12)). Figure 13 shows the robust standard deviation of the residuals $\sigma(\Delta V_R^t)$ calculated over time bins of six revolutions and plotted as a function of time. This diagnostics is used to monitor the precision of the wavelength-calibration over the time covered by the DR2 data set. The peak in the interval OBMT [2324; 2331] was due to poor attitude data from the AGIS solution, resulting in a poor estimate of the field angles of the stars. The V_R^t obtained in this interval were not used to produce the combined V_R published in DR2. For the remaining time, the precision of the V_R^t estimation for the CU6GB-cal stars is ~0.4 km s^{-1}.

The accuracy of the single-transit radial velocities, which is indicative of systematic uncertainties, is quantified by $Md(\Delta V_R^t)$, where $\Delta V_R^t = V_R^t - V_R^{ref}$ are the differences (residuals) between the V_R^t and the V_R^{ref} in CU6GB-cal. Figure 14 shows the median value of the residuals, $Md(\Delta V_R^t)$ calculated over time bins of six revolutions and plotted as a function of time to monitor the systematic shifts of the radial velocity zeropoint, which reflect corresponding shifts in the wavelength calibration zeropoint. No explanation has been found for the discontinuities at OBMT 2751.3 and 3538.8. The shifts are smaller than 500 m s^{-1}, and being at transit level, do not significantly impact the median V_R of the stars affected. Two break-points corresponding to these discontinuities will be set in DR3 to correct for these shifts.

7.8.3. Dependence on the instrumental configuration

The subset of the verification-dataset, containing only stars belonging to CU6GB-cal, is used to estimate the systematic differences in the V_R^t depending on the CCD row and on the FoV of
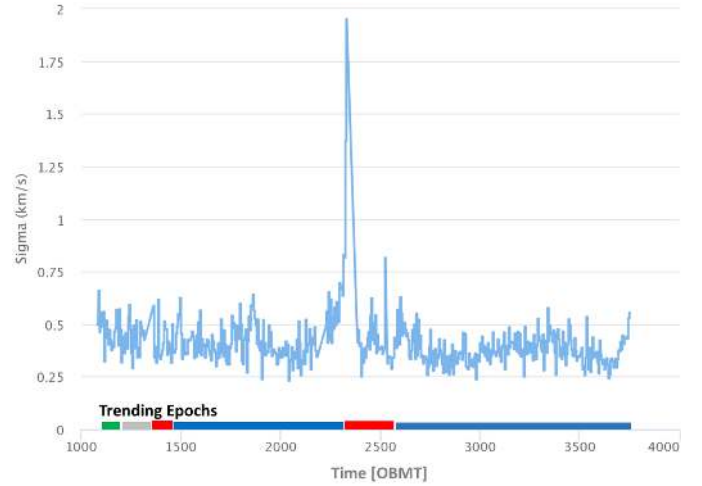
Fig. 13. Overall precision of the single-transit radial velocity, $\sigma(\Delta V_R^t)$, as a function of the observation time (OBMT revolutions) for the standard stars in CU6GB-cal (Table 1). These are bright FGK-type stars with $G_{RVS}^{ext} \leq 9$. The trending epochs are indicated on the x-axis. The y-axis shows the robust standard deviation of the residuals, $\sigma(\Delta V_R^t)$, calculated over bins of six OBMT.
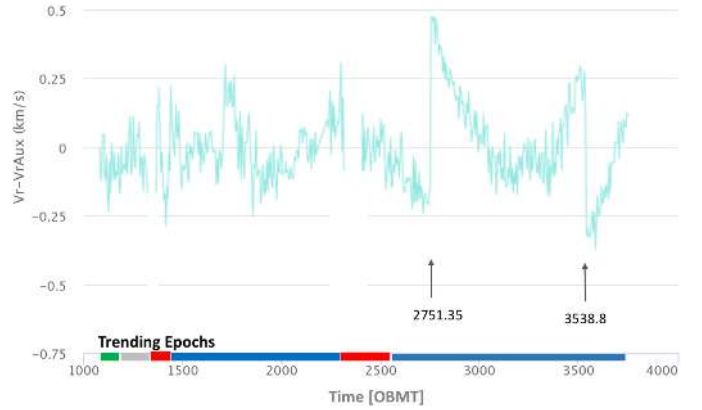
Fig. 14. Median value of the residuals, $Md(\Delta V_R^t)$, as a function of time for the standard stars in CU6GB-cal. This plot is used to monitor the systematic shifts in the wavelength-calibration zeropoint. The trending epochs are indicated on the x-axis. The median is calculated over bins of six OBMT (revolutions).

the transit (the stars are observed in FoV1 and FoV2, the FoVs of the two telescopes, and in general, in a different CCD row at each transit).

The mean, median, and robust standard deviations of the residuals ΔV_R^t were obtained over the entire period covered by DR2 for each configuration FoV-CCDrow. Table 3 shows the results. The median values $Md(\Delta V_R^t)$ indicate the systematic shift of V_R^t depending on the configuration of the transit. When no selection was made on the transit configuration (first row of Table 3), the overall small systematic shift is approximately ~−0.01 km s^{-1}, showing that no zeropoint shift between the RVS V_R^t and the V_R^{ref} in CU6GB-cal is introduced by the processing overall; the temporal evolution of the shift for all configurations together is shown in Fig. 14. Systematic shifts are present between the two FoVs (in the opposite sense) and between the different rows. These shifts are small and acceptable for the purposes of this data release. The final radial velocities are the result of the combination between various transits in different configurations, and the shifts are averaged. The robust standard deviation ($\sigma(\Delta V_R^t)$ in Table 3) is indicative of the

Table 3. Overall uncertainty depending on the configuration of the observation for the stars in CU6GB-cal.

FoV-row	$Md(\Delta V_R^t)$ km s^{-1}	mean(ΔV_R^t) km s^{-1}	$\sigma(\Delta V_R^t)$ km s^{-1}
All	-0.01	-0.04	0.42
FoV1-All	0.15	0.15	0.37
FoV1-row4	0.18	0.20	0.37
FoV1-row5	0.10	0.13	0.39
FoV1-row6	0.13	0.12	0.36
FoV1-row7	0.17	0.14	0.35
FoV2-All	-0.19	-0.23	0.40
FoV2-row4	-0.08	-0.15	0.43
FoV2-row5	-0.15	-0.22	0.45
FoV2-row6	-0.20	-0.17	0.33
FoV2-row7	-0.31	-0.41	0.36

Notes. The units are km s^{-1}. Median, mean, and standard deviation of the residuals $\Delta V_R^t = V_R^t - V_R^{ref}$(km s^{-1}) over the time covered by the DR2. FoV-row are the possible configurations of a star transit (FoV1 and FoV2 are the FoVs of the two telescopes, and rows 1–4 are the CCD rows, see Fig. 1).

precision of the measurements V_R^t for one transit in a given FoV and row for the stars of spectral type and magnitude typical of CU6GB-cal.

7.8.4. Comparison with the auxiliary catalogues

The entire verification dataset is used (and not only the subset of CU6GB-cal stars) to estimate the performance of the V_R^t measurements. RAVE and APOGEE cover fainter magnitudes than CU6GB-cal (Fig. 3), and permit verifying the accuracy and precision of V_R^t up to $G_{RVS}^{ext} \sim 12$. The accuracy and the precision of the V_R^t are quantified by $Md(\Delta V_R^t)$ and $\sigma(\Delta V_R^t)$, respectively, as described in Sect. 7.8.2. The results for the stars belonging to APOGEE are an overall accuracy of ~0.2 km s^{-1}, an overall precision of ~1.8 km s^{-1}, and for the stars belonging to RAVE, an overall accuracy of ~0.3 km s^{-1} and an overall precision of ~1.4 km s^{-1}. The uncertainties on the APOGEE and RAVE V_R^{ref} are larger than those of CU6GB-cal and contribute to the estimation of the accuracy. These results permitted verifying that overall, the pre-launch requirement on the accuracy on the radial velocities of ~300 m s^{-1} (Cropper et al. 2018, their Table 1) was met.

8. Combining the single-transit radial velocities

The multi-transit analysis (MTA) has the task to combine for each observed star the single-transit radial velocities, V_R^t, to produce the V_R published in DR2. V_R^t is combined by applying a median value,

$$V_R = Md(V_R^t),\qquad(17)$$

where

- V_R is produced only if the number, N, of V_R^t (i.e. the number of valid transits) is not smaller than two;
- the transits for which multiple-component lines have been detected (i.e. those indicated as comp-2 in Fig. 12) are excluded from the combination;
- for duplicated transits within 1 s of each other, the transit with the fainter spectrum is excluded. Duplicated transits (i.e. two transits nearby associated with the same source)
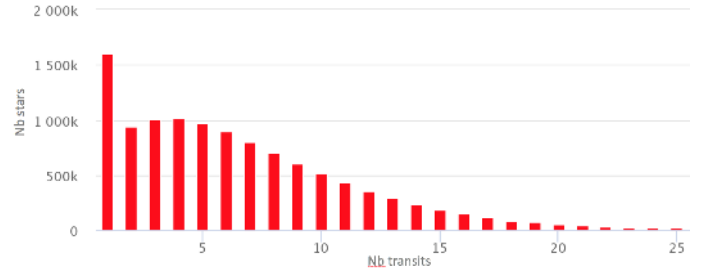
Fig. 15. Distribution of the number of stars as a function of their number of transits. The *x*-axis shows the number of transits, and the *y*-axis shows the number of stars. 1.6 million sources had only one valid transit, and for these sources, V_R is not computed.

may occur for bright stars and are the result of false detections due to PSF features.

Figure 15 shows the distribution of the number of stars as a function of the number of their valid transits. The uncertainty computed by the pipeline is the standard deviation on the median,

$$\sigma_{V_R}^{MTA} = \sqrt{\frac{\pi}{2}} \frac{\sigma(V_R^t)}{\sqrt{N}},\qquad(18)$$

where $\sigma(V_R^t)$ is the standard deviation of the set of V_R^t measurements.

In a post-processing stage, a constant shift of 0.11 km s^{-1} was later added to take into account a calibration floor contribution. This shift was estimated by comparing the internal (Eq. (18)) and the external precision (Sect. 9) for a set of constant stars from CU6GB-cal whose radial velocity V_R^{ref} is known with the best precision. Then, the uncertainty (km s^{-1}) associated with the V_R measurements in DR2 is

$$\sigma_{V_R} = \sqrt{(\sigma_{V_R}^{MTA})^2 + 0.11^2}.\qquad(19)$$

In addition to V_R, other information is computed for each star by combining the single-transit information. This is not published in DR2, but is used to infer the quality of the V_R and to discard poor-quality measurements (Katz et al. 2018).

9. Radial velocity results

Various diagnostics are implemented in the automated verification AVMTA (Fig. 2) to automatically verify the median radial velocity results produced by the MTA for each star from the total of its transits. In the same way as for the single-transit radial velocities, a verification dataset is defined that contains only stars for which we also have a ground-based radial velocity in the auxiliary external catalogues (Table 1). The verification dataset covers the magnitude range of the RVS data (Fig. 3), but does not contain early-type or late-M stars. Table 4 lists the measurements performed by AVMTA.

9.1. Overall accuracy

The accuracy indicates the systematic uncertainty. It is estimated using the verification dataset by calculating the difference between the RVS and the external catalogue measurements. In Table 4, $Md(\Delta V_R)$ is the median value of the residuals, calculated over the entire period covered by DR2 for the stars

Table 4. Median radial velocities compared to external catalogues.

Cat Name	$Md(\Delta V_R)$ km s^{-1}	$\sigma(\Delta V_R)$ km s^{-1}	$Md(\sigma_{V_R}^{MTA})$ km s^{-1}	$\sigma(\sigma_{V_R}^{MTA})$ km s^{-1}
CU6GB-cal	−0.01	0.20	0.14	0.07
XHip	0.35	0.51	0.13	0.10
CU6GB-val	−0.01	0.51	0.23	0.23
RAVE	0.29	1.03	0.37	0.27
APOGEE	0.22	0.89	0.71	0.67
SIM	0.27	0.29	0.23	0.13

Notes. The units are km s^{-1}. The `MTA` V_R (Eq. 17) are compared with the external catalogues described in Sect. 4.1 and 4.2. $Md(\Delta V_R)$ quantifies the overall systematic shifts between the RVS and the radial velocities of other catalogues. $Md(\sigma_{V_R}^{MTA})$ quantifies the overall precision of the RVS radial velocities for the stars belonging to each catalogue. The precision is worse for the stars in APOGEE, which are the faintest.
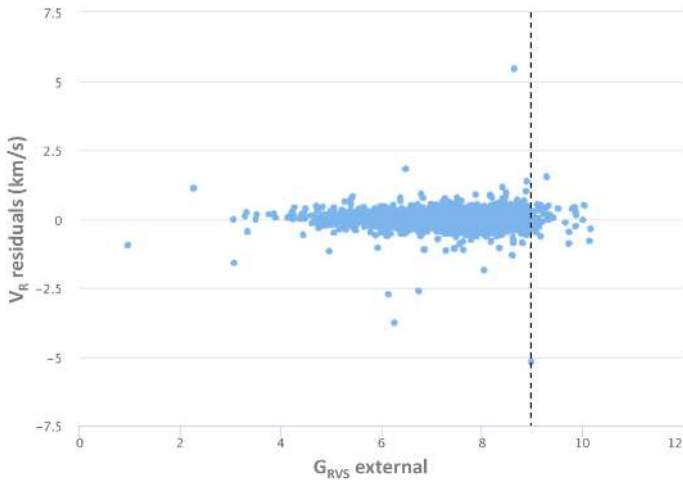
Fig. 16. Residuals ΔV_R for the stars in CU6GB-cal as a function of G_{RVS}^{ext}. The overall median value is −0.01 km s^{-1} and the dispersion is 0.20 km s^{-1} (Table 4). The magnitude distribution of the CU6BG-cal stars is shown in Fig. 3. The stars with $G_{RVS}^{ext} \leq 9$ are used as standard stars in the wavelength calibration. The pipeline processes stars with $G_{RVS}^{ext} \leq 12$. There is no bright-end limit in the selection of the stars to process, but the pipeline removes the saturated spectra (Sect. 6.1.1).

belonging to each catalogue in the verification dataset. The residuals $\Delta V_R = V_R - V_R^{ref}$ are the differences between the median radial velocity obtained by the pipeline (Eq. (17)) and those of the external catalogues (V_R^{ref}) and contain both the catalogue and the RVS uncertainties. They are in good agreement with the median residuals $Md(\Delta V_R^t)$ obtained with the single-transit radial velocities V_R^t and indicate the overall shift between the RVS and the other catalogues.

When we assume that the uncertainties in the external catalogue are negligible, $Md(\Delta V_R)$ is an indicator of the overall accuracy (systematic zeropoint shift) in the RVS V_R. The low value for the stars in CU6GB indicates that overall, the RVS V_R zeropoint is in agreement with that of the standard stars, which implies that the pipeline processing did not introduce any significant systematic shift in general. Figure 16 shows the residuals as a function of G_{RVS}^{ext} for the CU6GB-cal stars, with no evident magnitude-dependent zeropoint shift.

The DR2 data do exhibit an overall shift of $\sim +300$ m s^{-1} compared with the other catalogues (Table 4). These shifts are

acceptable for DR2 since the pre-launch end-of-mission requirement on systematic uncertainties was ≤300 km s^{-1} (see Cropper et al. 2018, Table 1).

9.2. Overall precision

The precision indicates random uncertainties. In Table 4 the overall precision of the RVS V_R measurements is quantified using the verification dataset with $\sigma(\Delta V_R)$ and $Md(\sigma_{V_R}^{MTA})$ calculated over the 22 months covered by DR2. $\sigma(\Delta V_R)$ is the robust dispersion (see Eq. (12)) of the residuals for the stars belonging to the different catalogues of the verification data set. When we assume negligible uncertainties in the external catalogues, it is an indicator of the overall precision of the RVS measurements and is called external precision (this includes the precision of the external catalogue measurements). $Md(\sigma_{V_R}^{MTA})$ is the overall internal precision, which is also shown in Table 4, where $\sigma_{V_R}^{MTA}$ is estimated in the pipeline (Eq. (18)). The internal precision is independent of the external catalogue uncertainty and depends on the dispersion of the RVS measurements. It also depends on the magnitude of the stars: it is higher for CU6GB-cal and XHip, which are dominated by bright stars (see Fig. 3) and lower for RAVE and APOGEE, which contain fainter stars.

The CU6GB-cal V_R^{ref} are more precise ($\sigma < 0.1$; Table 1) and are also expected to be more accurate than the other external catalogues. For these stars the agreement between the external and the internal precision is reached by adding 0.11 km s^{-1} in quadrature to the internal estimation. This shift is interpreted as a calibration floor. After the pipeline processing was completed, it was added quadratically to the MTA uncertainty (Eq. (19)) to improve their estimation in a post-processing phase.

Figure 17 shows the internal precision as a function of G_{RVS}^{ext} for the ~10 million V_R produced by the pipeline (not only those in the verification dataset: SB1 and variable stars are included). The median, $Md(\sigma_{V_R}^{MTA})$, is calculated over 0.25 magnitude bins. The precision is even better than the pre-launch end-of-mission requirement of 1 km s^{-1} for stars brighter than $G_{RVS}^{ext} \sim 10.5$.

9.3. Off-line validation

When the processing of the data was completed and `Automated Verification` confirmed that the overall accuracy and precision of the V_R met the requirements, a validation campaign started on the ~10 million V_R produced by the pipeline. Its purpose was to analyse the global properties of the dataset and identify the V_R without sufficient quality to be published in DR2 (Katz et al. 2018). The exclusion criteria include those mentioned in previous sections, which are hot ($T_{eff}^{tpl} \geq 7000$ K) and cool stars ($T_{eff}^{tpl} \leq 3500$ K), see also Sect. 6.5, suspected SB2 detected as double-lined in 90% of the transits, emission-line stars, and stars that are too faint ($G_{RVS}^{int} > 14$), stars for which the V_R^t has been detected as *ambiguous* (Sect. 7.7) in all transits. Excluded from DR2 are also the V_R for which $\sigma_{V_R} \geq 20$ km s^{-1}, and ~400 outliers $V_R \geq 500$ km s^{-1} that were obtained with poor-quality spectra.

After the above filters were applied, an exhaustive validation study of the V_R pipeline products was performed, which is described in the accompanying paper (Katz et al. 2018). It includes an analysis of the accuracy and the precision depending on the star nature, magnitude, number of transits, and sky distribution. Then, another validation campaign (Arenou et al. 2018) was carried out on the entire DR2 data, including the products of the other pipelines, and resulted in the exclusion of some stars.
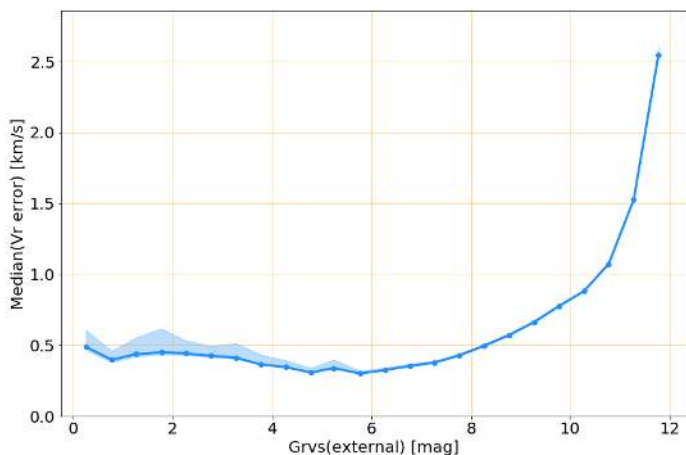
Fig. 17. Precision ($Md(\sigma_{V_R}^{MTA})$ over magnitude bins) of the full V_R data set produced by the pipeline: 98 16 603 stars. After an off-line validation campaign in which poor-quality V_R have been excluded, ~7.2 million stars are left for publication in DR2, and the internal precision was improved by ~30%

In the end, ~2.6 million V_R were excluded, and the number of *Gaia* stars with a V_R in DR2 is ~7.2 million.

10. Conclusions

We have presented the spectroscopic pipeline that was used to process the *Gaia* RVS data to produce the radial velocity measurements that are released in the *Gaia* DR2. The pipeline processed ~280 million individual RVS spectra of stars with apparent G_{RVS} magnitude brighter than ~12, distributed throughout the entire celestial sphere.

We described the calibration of the RVS, the reduction of the spectra and the determination of the radial velocity V_R, together with the overall V_R performance estimated by the pipeline. The V_R recorded for each star is the median of all radial velocities obtained from individual observations of that star. The radial velocity is measured through a fit of the RVS spectrum with a synthetic template spectrum. To select the appropriate template, a rough estimate of the stellar atmospheric parameters was performed by the pipeline. The large uncertainties affecting such estimates for the hottest ($T_{eff} \geq 7000$ K) and coolest stars ($T_{eff} \leq 3500$ K) implied large uncertainties in the resulting V_R because of template mismatch errors. The V_R of these stars were removed from DR2.

The overall accuracy (systematic uncertainties) of the radial velocity products, estimated automatically through comparisons with ground-based radial velocity catalogues, is better than 0.3 km s^{-1}. The overall precision (random uncertainties), also estimated automatically using stars known to be stable from ground-based observations, is better than 1 km s^{-1}. The best precision of <0.2 km s^{-1} is obtained for the stars brighter than $G_{RVS} \sim 7.5$. Although this is an early stage of data processing of only 22 months of *Gaia* RVS data, the radial velocities we obtained already approach or exceed the pre-launch end-of-mission requirements for bright stars

DR2 contains the median V_R of ~7.2 million stars. For bright stars, it provides the third component of the velocity vector, complementing the 2D proper motion information. The spectroscopic pipeline will be improved for the future data releases. This will include the V_R of fainter stars, more accurate and precise V_R for bright stars, variability information, rotational velocities,

calibrated spectra, and at a later stage, individual-transit observation data.

References

Allende Prieto, C., Koesterke, L., Ludwig, H.-G., Freytag, B., & Caffau, E. 2013, A&A, 550, A103
Anderson, E., & Francis, C. 2012, Astron. Lett., 38, 331
Andrae, R., Fouesneau, M., Creevey, O., et al. 2018, A&A, 616, A8 (*Gaia* 2 SI)
Arenou, F., Luri, X., Babusiaux, C., et al. 2018, A&A, 616, A17 (*Gaia* 2 SI)
Blomme, R., Edvardsson, B., Eriksson, K., et al. 2017, Gaia Data Processing and Analysis Consortium (DPAC) technical note GAIA-C6-TN-ROB-RHB-005, http://www.cosmos.esa.int/web/gaia/public-dpac-documents
Casagrande, L., Ramírez, I., Meléndez, J., Bessell, M., & Asplund, M. 2010, A&A, 512, A54
Chiavassa, A., Bigot, L., Thévenin, F., et al. 2011, J. Phys. Conf. Ser., 328, 012012
Chiavassa, A., Casagrande, L., Collet, R., et al. 2018, A&A, 611, A11
Crifo, F., Jasniewicz, G., Soubiran, C., et al. 2010, A&A, 524, A10
Cropper, M., Katz, D., Sartoretti, P., et al. 2018, A&A, 616, A5 (*Gaia* 2 SI)
Crowley, C., Kohley, R., Hambly, N. C., et al. 2016, A&A, 595, A6
David, M., & Verschueren, W. 1995, A&AS, 111, 183
David, M., Blomme, R., Frémat, Y., et al. 2014, A&A, 562, A97
Fabricius, C., Bastian, U., Portell, J., et al. 2016, A&A, 595, A3
Gaia Collaboration, (Prusti, T., et al.) 2016, A&A, 595, A1
Gaia Collaboration, (Brown, A., et al.) 2018, A&A, 616, A1 (*Gaia* 2 SI)
Guerrier, A., Katz, D., Turon, C., & Crifo, F. 2007, in SF2A-2007: Proceedings of the Annual Meeting of the French Society of Astronomy and Astrophysics, eds. J. Bouvier, A. Chalabaev, & C. Charbonnel, 485
Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, A&A, 486, 951
Hambly, N., Cropper, M., Boudreault, S., et al. 2018, A&A, 616, A15 (*Gaia* 2 SI)
Horne, K. 1986, PASP, 98, 609
Jordi, C., Gebran, M., Carrasco, J. M., et al. 2010, A&A, 523, A48
Katz, D., Sartoretti, P., Cropper, M., et al. 2018, A&A, submitted (*Gaia* 2 SI)
Lasker, B. M., Lattanzi, M. G., McLean, B. J., et al. 2008, AJ, 136, 735
Lindegren, L. 2009, Gaia Data Processing and Analysis Consortium (DPAC) Technical Note GAIA-C3-TN-LU-LL-084, http://www.cosmos.esa.int/web/gaia/public-dpac-documents
Lindegren, L., Lammers, U., Hobbs, D., et al. 2012, A&A, 538, A78
Lindegren, L., Hernández, J., Bombrun, A., et al. 2018, A&A, 616, A2 (*Gaia* 2 SI)
Makarov, V. V., & Unwin, S. C. 2015, MNRAS, 446, 2055
Perruchot, S., Kohler, D., Bouchy, F., et al. 2008, in Ground-based and Airborne Instrumentation for Astronomy II, Proc. SPIE, 7014, 70140J
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1993, Numerical Recipes in FORTRAN; The Art of Scientific Computing, 2nd edn. (New York,: Cambridge University Press)
Recio-Blanco, A., de Laverny, P., Allende Prieto, C., et al. 2016, A&A, 585, A93
Robin, A. C., Luri, X., Reylé, C., et al. 2012, A&A, 543, A100
Smart, R. L., & Nicastro, L. 2014, A&A, 570, A87
Sordo, R., Vallenari, A., Tantalo, R., et al. 2011, in J. Phys. Conf. Ser., 328, 012006

Soubiran, C., Le Campion, J.-F., Cayrel de Strobel, G., & Caillo, A. 2010, A&A, 515, A111
Soubiran, C., Le Campion, J. L., & Chemin, L. 2014, GAIA-C6-TN-LAB-CS-011
Soubiran, C., Jasniewicz, G., Chemin, L., et al. 2018, A&A, 616, A7 (*Gaia* 2 SI)
Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, A&AS, 143, 9
Zucker, S. 2003, MNRAS, 342, 1291
Zucker, S., & Mazeh, T. 1994, ApJ, 420, 806
Zwitter, T., Zerjal, M., Matijevic, G., et al. 2016, Gaia Data Processing and Analysis Consortium (DPAC) technical note GAIA-C6-TN-LU-TZ-005, http://www.cosmos.esa.int/web/gaia/public-dpac-documents

[1] GEPI, Observatoire de Paris, Université PSL, CNRS, 5 Place Jules Janssen, 92190 Meudon, France
 e-mail: paola.sartoretti@obspm.fr
[2] Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK
[3] Royal Observatory of Belgium, Ringlaan 3, 1180 Brussels, Belgium
[4] Laboratoire Univers et Particules de Montpellier, Université Montpellier, CNRS, Place Eugène Bataillon, CC72, 34095 Montpellier Cedex 05, France
[5] CNES Centre Spatial de Toulouse, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France
[6] CRAAG - Centre de Recherche en Astronomie, Astrophysique et Géophysique, Route de l'Observatoire, Bp 63 Bouzareah, 16340 Alger, Algérie
[7] Institut d'Astrophysique et de Géophysique, Université de Liège, 19c, Allée du 6 Août, 4000 Liège, Belgium
[8] Universiteit Antwerpen, Onderzoeksgroep Toegepaste Wiskunde, Middelheimlaan 1, 2020 Antwerpen, Belgium
[9] F.R.S.-FNRS, Rue d'Egmont 5, 1000 Brussels, Belgium
[10] Thales Services for CNES Centre Spatial de Toulouse, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France
[12] Leibniz Institute for Astrophysics Potsdam (AIP), An der Sternwarte 16, 14482 Potsdam, Germany
[13] Laboratoire d'astrophysique de Bordeaux, Université de Bordeaux, CNRS, B18N, allée Geoffroy Saint-Hilaire, 33615 Pessac, France
[14] Laboratoire Lagrange, Université Nice Sophia-Antipolis, Observatoire de la Côte d'Azur, CNRS, 34229, 06304 Nice Cedex, France
[15] Instituto de Astrofísica de Canarias, 38205 La Laguna, Tenerife, Spain
[16] Universidad de La Laguna, Departamento de Astrofísica, 38206 La Laguna, Tenerife, Spain
[17] IPAG, Université Grenoble Alpes, CNRS, IPAG, 38000 Grenoble, France
[18] Max Plank Institute für Sonnensystemforschung, Justus-von-Liebig-Weg 3, 37077 Göttingen, Germany
[19] Unidad de Astronomía, Fac. Cs. Básicas, Universidad de Antofagasta, Avda. U. de Antofagasta RCH-02800, Antofagasta, Chile
[20] ATOS for CNES Centre Spatial de Toulouse, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France
[21] Max Planck Institute for Extraterrestrial Physics, High Energy Group, Gießenbachstraße, 85741 Garching, Germany
[22] Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
[23] LNE-SYRTE, Observatoire de Paris, Université PSL, CNRS, Sorbonne Universités, 61 avenue de l'Observatoire, 75015 Paris, France
[24] Faculty of Mathematics and Physics, University of Ljubljana, Jadranska ulica 19, 1000 Ljubljana, Slovenia
[25] INAF-National Institute of Astrophysics, Osservatorio Astronomico di Padova, Osservatorio Astronomico, 36012 Asiago (VI), Italy
[26] Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia
[11] Department of Astronomy, University of Geneva, Chemin d'Ecogia 16, 1290 Versoix, Switzerland

Appendix A: Acronyms

Table A.1. Acronyms used in this paper.

Acronym	Description
AC	Across-scan (direction)
ADU	Analogue-to-digital unit
AGIS	Astrometric global iterative solution
AL	Along-scan (direction)
AP	Atmospheric parameters
AV	Automated verification
AVEXT	Automated verification of extraction (i.e. calibration preparation workflow)
AVPP	Automated verification of pre-processing (i.e. ingestion workflow)
AVFE	Automated verification of full extraction
AVMTA	Automated verification of MTA
AVSTA	Automated verification of STA
CaU	Calibration unit
CCD	Charge-coupled device (detector)
CCF	Cross-correlation function
CPU	Central processing unit
CRNU	(CCD) Column response non-uniformity
CTI	Charge transfer inefficiency
CU6GB	Coordination Unit 6 ground-based radial velocity catalogue
DR2	(*Gaia*) Data Release 2
EPSL	Ecliptic pole scanning law
FL	First look
FoV	Field of view (also denoted FOV)
GACS	*Gaia* archive core systems
GSC23	the second guide star catalogue version 2.3
IDT	Initial data treatment
IDU	Intermediate data update
IGSL	Initial *Gaia* source list
LSF	Line spread function
MTA	Multi-transit analysis
NIR	Near-infrared
NSL	Nominal scanning law
NU	Non-uniformity
OBMT	On-board mission timeline (in *Gaia* revolutions; 1 revolution = 6 h)
OGA3	On-ground attitude from AGIS
PCA	Principal components analysis
PCF	Pearson correlation function
PEM	Proximity-electronics module (CCD)
PSF	Point spread function
RAVE	Radial velocity experiment
RMS	Root mean square
RP	Red photometer
RVDir	Radial velocity determination in direct space, using PCF
RVFou	Radial velocity determination with Fourier technique, using CCF
RVMDM	Radial velocity determination using minimum distance method
RVS	Radial Velocity Spectrometer
SAGA	System of accomodation of *Gaia* algorithms
SB2	Double-lined spectroscopic binary
SIM	Space interferometry mission
STA	Single-transit analysis
TDI	Time-delayed integration (CCD)
VO	Virtual object
VPU	Video processing unit