



Publication Year	2018
Acceptance in OA	2021-01-15T15:20:35Z
Title	Gaia Data Release 2. Summary of the variability processing and analysis results
Authors	Holl, B., Audard, M., Nienartowicz, K., Jevardat de Fombelle, G., Marchal, O., Mowlavi, N., CLEMENTINI, Gisella, De Ridder, J., Evans, D. W., Guy, L. P., Lanzafame, Alessandro Carmelo, Lebzelter, T., Rimoldini, L., Roelens, M., Zucker, S., DISTEFANO, Elisa Maria Carmela, GAROFALO, Alessia, Lecoœur-Taïbi, I., Lopez, M., MOLINARO, Roberto, Muraveva, Tatiana, Panahi, A., Regibo, S., RIPEPI, Vincenzo, Sarro, L. M., Aerts, C., Anderson, R. I., Charnas, J., Barblan, F., Blanco-Cuaresma, S., Busso, G., Cuypers, J., De Angeli, F., Glass, F., Grenon, M., Juhász, Á. L., Kochoska, A., Koubsky, P., LANZA, Antonino Francesco, LECCIA, Silvio, Lorenz, D., MARCONI, Marcella, Marschalkó, G., Mazeh, T., MESSINA, Sergio, Mignard, F., Moitinho, A., Molnár, L., Morgenthaler, S., MUSELLA, ILARIA, Ordenovic, C., Ordóñez, D., PAGANO, Isabella, Palaversa, L., Pawlak, M., Plachy, E., Prša, A., Riello, M., Süveges, M., Szabados, L., Szegedi-Elek, E., Votruba, V., Eyer, L.
Publisher's version (DOI)	10.1051/0004-6361/201832892
Handle	http://hdl.handle.net/20.500.12386/29787
Journal	ASTRONOMY & ASTROPHYSICS
Volume	618

Gaia Data Release 2

Summary of the variability processing and analysis results

B. Holl^{1,2}, M. Audard^{1,2}, K. Nienartowicz³, G. Jevardat de Fombelle³, O. Marchal^{2,4}, N. Mowlavi^{1,2}, G. Clementini⁵, J. De Ridder⁶, D. W. Evans⁷, L. P. Guy^{2,8}, A. C. Lanzafame^{9,10}, T. Lebzelter¹¹, L. Rimoldini², M. Roelens^{1,2}, S. Zucker¹², E. Distefano¹⁰, A. Garofalo^{13,5}, I. Lecoeur-Taïbi², M. Lopez¹⁴, R. Molinaro¹⁵, T. Muraveva⁵, A. Panahi¹⁶, S. Regibo⁶, V. Ripepi¹⁵, L. M. Sarro¹⁷, C. Aerts^{6,18}, R. I. Anderson¹⁹, J. Charnas², F. Barblan¹, S. Blanco-Cuaresma^{1,20,21}, G. Busso⁷, J. Cuypers^{22,*}, F. De Angeli⁷, F. Glass¹, M. Grenon¹, Á. L. Juhász^{23,24}, A. Kochoska^{25,26}, P. Koubsky²⁷, A. F. Lanza¹⁰, S. Leccia¹⁵, D. Lorenz¹¹, M. Marconi¹⁵, G. Marschalkó^{23,28}, T. Mazeh¹⁶, S. Messina¹⁰, F. Mignard²⁹, A. Moitinho³⁰, L. Molnár²³, S. Morgenthaler³¹, I. Musella¹⁵, C. Ordenovic²⁹, D. Ordóñez², I. Pagano¹⁰, L. Palaversa^{7,1}, M. Pawlak^{32,33}, E. Plachy²³, A. Prša²⁵, M. Riggio⁷, M. Süveges³⁴, L. Szabados²³, E. Szegedi-Elek²³, V. Votrubá²⁷, and L. Eyer^{1,2}

(Affiliations can be found after the references)

Received 23 February 2018 / Accepted 28 May 2018

ABSTRACT

Context. The *Gaia* Data Release 2 (DR2) contains more than half a million sources that are identified as variable stars.

Aims. We summarise the processing and results of the identification of variable source candidates of RR Lyrae stars, Cepheids, long-period variables (LPVs), rotation modulation (BY Dra-type) stars, δ Scuti and SX Phoenicis stars, and short-timescale variables. In this release we aim to provide useful but not necessarily complete samples of candidates.

Methods. The processed *Gaia* data consist of the G , G_{BP} , and G_{RP} photometry during the first 22 months of operations as well as positions and parallaxes. Various methods from classical statistics, data mining, and time-series analysis were applied and tailored to the specific properties of *Gaia* data, as were various visualisation tools to interpret the data.

Results. The DR2 variability release contains 228 904 RR Lyrae stars, 11 438 Cepheids, 151 761 LPVs, 147 535 stars with rotation modulation, 8882 δ Scuti and SX Phoenicis stars, and 3018 short-timescale variables. These results are distributed over a classification and various Specific Object Studies tables in the *Gaia* archive, along with the three-band time series and associated statistics for the underlying 550 737 unique sources. We estimate that about half of them are newly identified variables. The variability type completeness varies strongly as a function of sky position as a result of the non-uniform sky coverage and intermediate calibration level of these data. The probabilistic and automated nature of this work implies certain completeness and contamination rates that are quantified so that users can anticipate their effects. This means that even well-known variable sources can be missed or misidentified in the published data.

Conclusions. The DR2 variability release only represents a small subset of the processed data. Future releases will include more variable sources and data products; however, DR2 shows the (already) very high quality of the data and great promise for variability studies.

Key words. stars: general – stars: oscillations – stars: solar-type – stars: variables: general – Galaxy: stellar content – catalogs

1. Introduction

The Coordination Unit 7 (CU7) of the *Gaia* Data Processing and Analysis Consortium (DPAC) is tasked to process the calibrated data of variable objects detected by *Gaia*. In *Gaia* Data Release 1 (DR1; [Gaia Collaboration 2016](#)), we published the light curves and properties of a sample of Cepheids and RR Lyrae stars detected in the Large Magellanic Cloud (LMC); see [Clementini et al. \(2016\)](#). We refer to [Eyer et al. \(2017\)](#) for a detailed description of the CU7 framework and processing pipeline. In this second data release (DR2; [Gaia Collaboration 2018a](#)), we extend the variability types for which we publish light curves to the following list of classes: Cepheids,

RR Lyrae stars, long-period variables (LPV), short-timescale variables, stars with rotation modulation, and δ Scuti and SX Phoenicis stars. For some of them, additional specific properties are inferred and published. This constitutes roughly 100 times as many variable sources as were published in DR1 ([Clementini et al. 2016](#); [Eyer et al. 2017](#)). Moreover, it covers the whole sky. The variability analysis for this release focused primarily on large-amplitude variable stars and rotation modulated stars. Other classes of variable stars are aimed to be introduced in later releases. We note that eclipsing binaries were identified as well, but will be treated separately and only delivered in future releases. We emphasise that specific selection criteria were applied to each variability type to limit contamination. Completeness within each variability type was not aimed at for DR2.

* Deceased on 28 February 2017.

The various variability type-specific methods that were used for the published DR2 results are discussed in full length in the following dedicated papers:

- DR2: variable stars in the *Gaia* colour-magnitude diagram (Gaia Collaboration 2018b).
- DR2: all-sky classification of high-amplitude pulsating stars (Rimoldini et al., in prep.).
- DR2: specific characterisation and validation of all-sky Cepheids and RR Lyrae stars (Clementini et al. 2018).
- DR2: rotation modulation in late-type dwarfs (Lanzafame et al. 2018).
- DR2: short-timescale variability processing and analysis (Roelens et al. 2018).
- DR2: the first *Gaia* catalogue of long-period variable candidates (Mowlavi et al. 2018).
- DR2: validation of the classification of RR Lyrae and Cepheid variables with the *Kepler* and K2 missions (Molnár et al. 2018).

This overview paper focuses on presenting the general properties of the exported sample of variable sources in the following way. Section 2 introduces the time-series photometric data and their relevant properties. Section 3 presents a summarised overview of the data processing and analyses. Section 4 gives an overview of the results, including our time-series filters, published time-series, classification, and Specific Object Studies (SOS) tables containing the results. Section 5 presents the final conclusions. Additional information regarding the *Gaia* archive queries used in this paper is provided in Appendix B.

The data are publicly available in the online *Gaia* archive¹ containing the “tables” and “fields” referred to in the rest of this article, as well as all of the “DR2 documentation” and catalogue “data model”².

2. Data

The results in DR2 are based on the first 22 months of *Gaia* data, taken between 25 July 2014 and 23 May 2016. For the variability analysis we made use of the photometric time series in the G , G_{BP} , and G_{RP} bands, using field-of-view (FoV) averaged transit photometry. An exception to this was the analysis of short-timescale variables, for which some per-CCD G -band photometry was also analysed, but not published. The photometry is described in detail in Evans et al. (2018) and Riello et al. (2018). We emphasise that owing to the tight DR2 data-processing schedule, we only had access to a preliminary version of the DR2 astrometric solution (Lindegren et al. 2018), on which our results are based. No radial velocity data or astrophysical parameters were available during our processing.

2.1. Relative completeness

As a consequence of the *Gaia* scanning law, the sky coverage is rather non-uniform for the 22 months of DR2 data, resulting in sky-coverage gaps when selecting a certain minimum number of FoV transits. In our processing, we used cuts of ≥ 2 , ≥ 12 , and ≥ 20 G -band FoV transits, which we discuss in Sect. 3. Figure 1 shows the effect of these cuts on the sample of all available time series at the time of processing. Assuming that all sources with ≥ 2 FoV transits are bona fide sources, these figures show that a cut of ≥ 12 FoV transits results in a relative completeness of 80%, and a cut of ≥ 20 FoV transits results in a relative completeness

of only 51%. It can generally be expected that these relative completeness are an upper limit to the absolute (sky) completeness, which is assessed in detail in the follow-up variability papers mentioned in the introduction. Arenou et al. (2018) conclude that the catalogue is mostly complete, in the absolute sense, roughly between magnitudes 7 and 17, and incomplete beyond. For a first-order relative completeness assessment in certain general Galactic directions, we introduce the following regions and compute the *Gaia* source count in each of them:

- *Galactic plane* $30^\circ < l < 330^\circ$ & $|b| < 15^\circ$.
- *Galactic centre* $20^\circ > l > 340^\circ$ & $|b| < 15^\circ$.
- *High Galactic latitudes* $|b| > 45^\circ$, excluding a 10° radius around the SMC.
- *LMC* region within a radius of 7.5° from Galactic coordinates $l = 280.47^\circ$, $b = -32.89^\circ$.
- *SMC* region within a radius of 2.5° from Galactic coordinates $l = 302.81^\circ$, $b = -44.33^\circ$.

We repeat this for the (approximate) sky footprints of various external catalogues, list the results in Table 1 and display the regions in Fig. 1. It shows that a (maximum) completeness of between 70–85% for ≥ 12 G -band FoV transits and a relative completeness of between 50–70% for ≥ 20 G -band FoV transits can be expected. The bottom panel of Fig. 1 shows that various sky regions are not or only partially sampled for ≥ 20 FoV: the Galactic centre region, for example, has a relative completeness of only 14%.

We remark that the numbers in Table 1 and Fig. 1 do not correspond exactly to the number of sources published in the `gaia_source` table of DR2 because various source filters in the global *Gaia* processing chain were applied before and after the variability processing. The derived percentages match very well, however, and they are relevant for the presented discussion.

2.2. Outliers and calibration errors

Various issues are visible in the current photometric data. Some of them are caused by instrumental effects, while others are sky related. Improved calibration strategies and careful flagging of these events are being designed and will be implemented in the chain of processing systems that convert the raw data into calibrated epoch photometry for the next *Gaia* Data Release. Some of these issues are also described in Evans et al. (2018):

- hot pixel columns are not yet treated,
- poor background estimates exist for some observations,
- observations close to bright sources can have biased background estimates or might even be spurious detections due to diffraction spikes,
- spatially close sources can have scan-angle direction dependent outliers,
- isolated sources can still be affected through overlap of the two FoVs.

It is expected that a certain fraction of observations affected by these issues has not been flagged and hence users should be aware of such unflagged outliers in the published data. We discuss variability flagging in more detail in Sect. 4.1.

In addition to the flagging of observations, the *Gaia* data processing consortium has applied various filters that remove sources that are affected by specific calibration issues; see Gaia Collaboration (2018a) for more details.

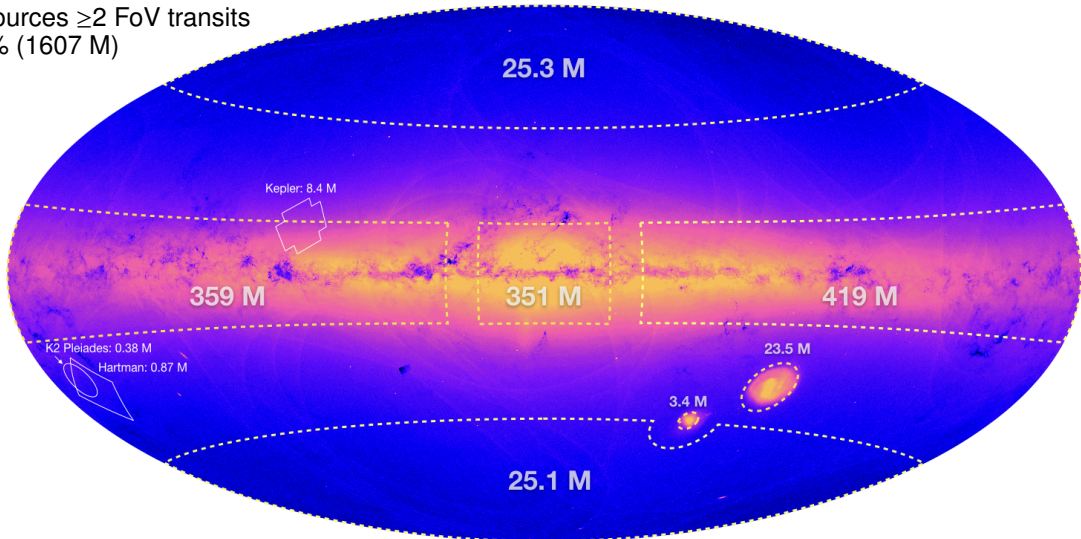
3. DR2 variability data processing and analysis

The general variability data processing has been described in detail in our DR1 paper (Eyer et al. 2017), to which we refer here.

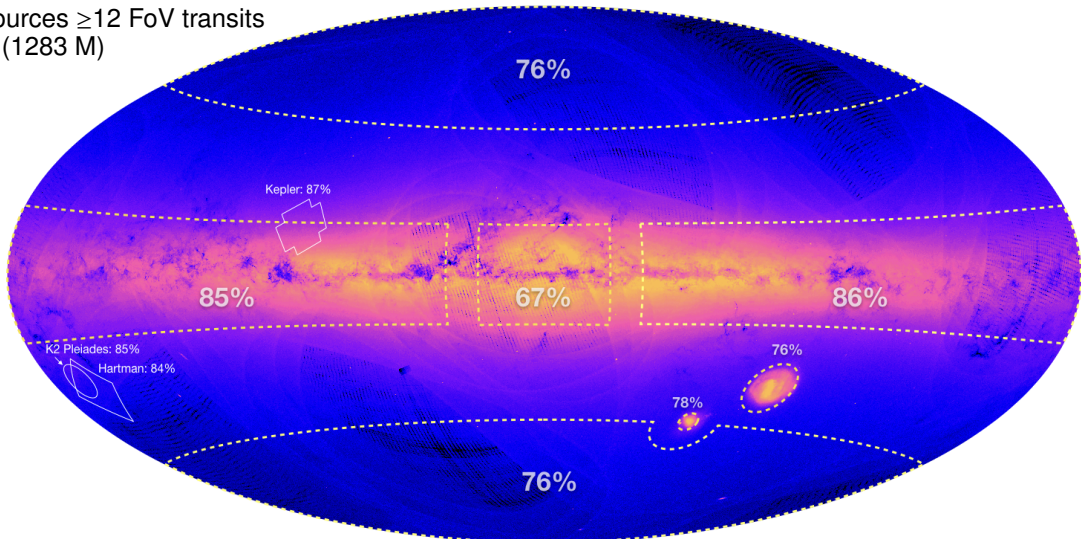
¹ <http://gea.esac.esa.int/archive/>

² <http://gea.esac.esa.int/archive/documentation/>

All sources ≥ 2 FoV transits
100% (1607 M)



All sources ≥ 12 FoV transits
80% (1283 M)



All sources ≥ 20 FoV transits
51% (826 M)

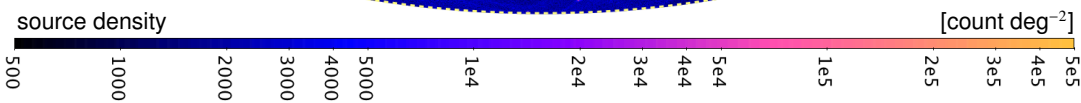
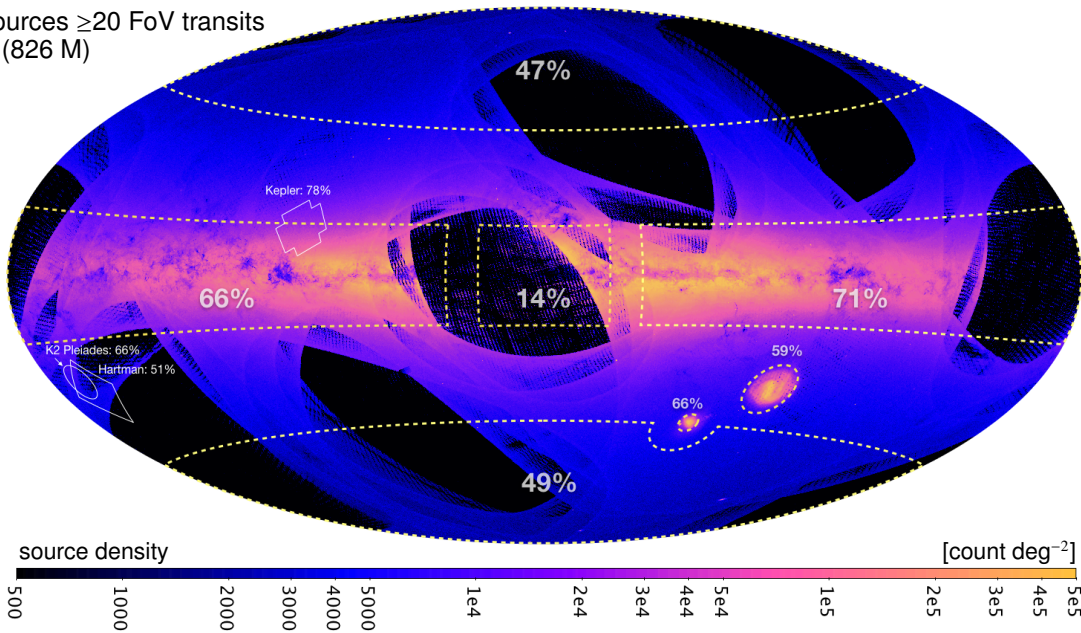


Fig. 1. Relative source counts with respect to sources having ≥ 2 *G*-band FoV transits, illustrating the drop in relative completeness for source sets with ≥ 12 and ≥ 20 *G*-band FoV transits, as applied in our processing (see Fig. 2). Regional values are listed in Table 1 and discussed in Sect. 2.1. Values for external catalogs are *Gaia* source counts in the approximate sky footprint, not source cross-match counts. Note that actual time series are published only for the 550 K variable sources in DR2 (see Fig. 3).

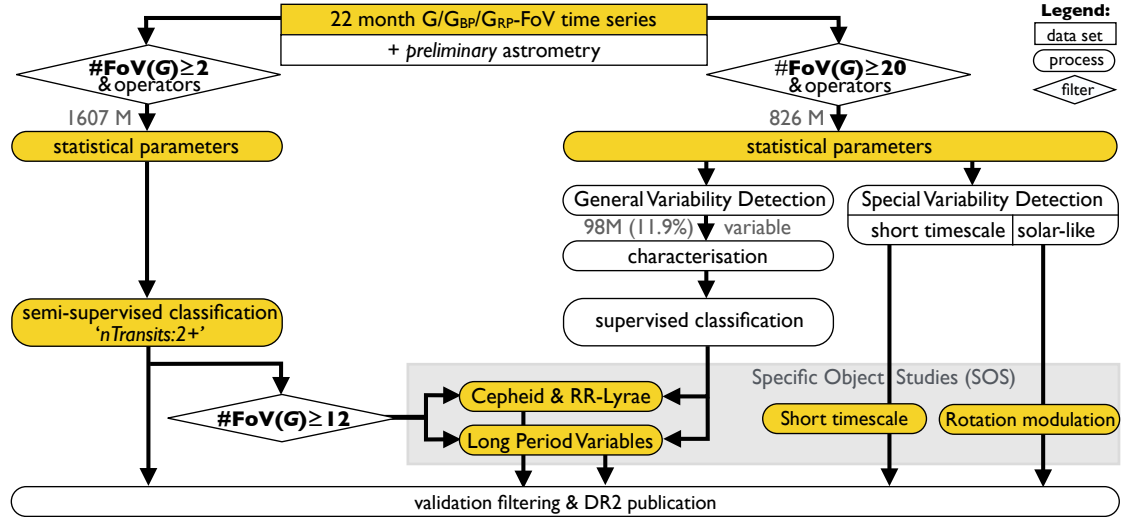


Fig. 2. DR2 variability processing overview. Data were published from the highlighted yellow boxes for the validation filtered sources. There were two main tracks: one starting from ≥ 2 G -band FoV transits (left panel), and the other from ≥ 20 G -band FoV transits (right panel). The former resulted in the published $nTransits:2+$ classifier in the `vari_classifier_result` table, and the latter resulted in the published SOS tables of `vari_short_timescale` and `vari_rotation_modulation`. The published SOS tables of `vari_rrlyrae`, `vari_cepheid` and `vari_long_period_variable` result from a mixed feed of the published $nTransits:2+$ classifier (for sources with at least ≥ 12 G -band FoV transits) and from the unpublished classifier of the ≥ 20 FoV track.

Table 1. Relative source counts for sources with ≥ 2 G -band FoV transits, illustrating the decrease in relative completeness for source sets with ≥ 12 and ≥ 20 G -band FoV transits, as applied in our processing.

Direction	≥ 2 FoV	≥ 12 FoV	≥ 20 FoV
ALL	100% (1607)	80% (1283)	51% (826)
Galactic plane	100% (778)	85% (665)	68% (532)
Galactic centre	100% (351)	67% (236)	14% (50.0)
High Galactic latitudes	100% (50.4)	76% (38.4)	48% (24.4)
LMC region	100% (23.5)	76% (17.9)	59% (13.9)
SMC region	100% (3.4)	78% (2.6)	66% (2.2)
Catalina DR2	100% (400)	80% (320)	53% (210)
Catalina DR1	100% (200)	80% (160)	50% (100)
OGLE-IV Bulge	100% (305)	66% (200)	10% (29)
OGLE-IV MCs	100% (33)	77% (25)	62% (20)
<i>Kepler</i>	100% (8.4)	87% (7.4)	78% (6.6)
Rebull et al. (2016) (K2 Pleiades)	100% (0.38)	85% (0.32)	66% (0.25)
Hartman et al. (2010) (Pleiades)	100% (0.87)	84% (0.73)	51% (0.44)

Notes. The top part specifies the numbers for some generic directions in the Galaxy, as defined in Sect. 2.1 and illustrated in Fig. 1. The bottom part specifies the numbers for several external catalogues, listing the *Gaia* source counts in their approximate sky footprints (not their source cross-match counts). In parenthesis, we provide the actual *Gaia* source counts in millions. Estimates of the absolute completeness of the published variable samples are presented in Tables 2 and 3.

We here only summarise the main aspects that are specific to the DR2 release. More exhaustive information is provided in the DR2 documentation.

The variability processing is summarised in Fig. 2, and the number of sources involved at various levels of the processing is indicated in the diagram. The processing followed two main paths: one starting from 1607 million sources with ≥ 2 G -band FoV transits (hereafter `geq2`), and the other from 826 million sources with ≥ 20 G -band FoV transits (hereafter `geq20`). The $\geq X$ means a G -band time series with X non-null and positive FoV flux observations, but before the variability filters were applied that are described in Sects. 4.1 and 4.2.

The first path resulted in the published $nTransits:2+$ classification results, as described in Sect. 4.3, as well as in a split-off path with ≥ 12 G -band FoV transits (hereafter `geq12`) that served

as one of the inputs for the SOS modules: SOS Cep and RRL (Clementini et al. 2016, 2018) and SOS LPV (Mowlavi et al. 2018).

The second path leads to the majority of the pipelines and SOS results, as described in Sect. 4.4. It had two different variability detection methods. The first is a “general” variability detection, defined by a classifier trained on known cross-matched constant objects³ from OGLE4 (Soszyński et al. 2012), HIPPARCOS (ESA 1997), the Sloan Digital Sky Survey (SDSS) (Ivezić et al. 2007), and classified constants, as well as many additional catalogues for variables (Rimoldini et al., in prep.). The second consists of multiple dedicated variability detection

³ The least variable sources found in <ftp://ftp.astrow.edu.pl/ogle/ogle4/GSEP/maps/>.

		Classification (vari_classifier_result)					Specific Object Studies (SOS)	
		best_class_name	Total				Total	Table name
8 882	δ Scuti & SX Phoenicis	DSCT_SXPHE	8 882					
228 904	RR Lyrae stars	RRAB, RRC, RRD, ARRD	195 780	88 120	107 660	33 124	140 784	vari_rrlyrae
11 438	Cepheids	CEP, ACEP, T2CEP	8 550	1 863	6 687	2 888	9 575	vari_cepheid
151 761	Long period variables	MIRA_SR	150 757	62 144	88 613	1 004	89 617	vari_long_period_variable
3 018	Short timescale variables						3 018	vari_short_timescale
147 535	Rotation modul. (BY Dra) stars						147 535	vari_rotation_modulation
+ 551 538 (550 737 unique source_id)		+ 363 969 (all unique source_id)					+ 390 529 (390 449 unique source_id)	

Fig. 3. Accounting of the 550 737 unique variable sources published in DR2. All have an entry in the *Gaia* archive `gaia_source` and `vari_time_series_statistics` tables. Their distributions in the additional variability tables are shown. The classification output types are mutually exclusive so that their sum matches the total number of entries in the published `vari_classifier_result` table. Eighty sources appear in more than one SOS table (not illustrated), which is detailed in Sect. 4.5. The overlap between similar-type classification and SOS is detailed in a Venn-like fashion showing the counts in the three possible subsets. A few hundred sources have different assigned types in classification and SOS tables (not illustrated), as explained in Sect. 4.6, hence the difference on the left between the total type count and the unique `source_id` count. No SOS module treated δ Scuti & SX Phoenicis stars.

methods: one tuned to detect short-timescale variables and another to rotation modulation effects, which were subsequently followed by their associated SOS modules: SOS short timescale (Roelens et al. 2018) and SOS rotation modulation (Lanzafame et al. 2018), analysing these sources in more detail. The general variability detection was followed by a generic multi-harmonic Fourier modelling based on the periodogram peak-frequency of an unweighted least-squares period search. Various classification attributes were derived from this model and the time-series statistical parameters, on which the classifier for the `geq20` path was trained. The classified RR Lyrae stars, Cepheids, and LPVs were subsequently fed into the corresponding modules SOS Cep and RRL and SOS LPV, being their second input source. All modules of the variability pipeline include validation rules to ensure that the output values are within acceptable ranges, which allowed us to identify issues early on in the processing.

An accounting of the number of sources published in the classification and SOS modules is given in Fig. 3. It shows that of the 550 737 unique variable sources, 363 969 (66%) appear in the classification results (Sect. 4.3) and 390 529 (71%) in SOS (Sect. 4.4). For several variability types, 203 681 sources (37%) appear in both, while the rest of the sources appear only in either classification (160 288, i.e., 29%) or SOS results (186 768, i.e., 34%). Note that the number appearing in both classification and SOS tables (203 681) is 721 more than the sum of overlapping types shown in Fig. 3 because these are cases of different assigned types in classification and SOS tables, which are omitted from the figure.

Depending on the type of object, there might only exist either a classification or SOS result. SOS results have generally lower contamination than the classification results and contain various type-specific astrophysical parameters that can be used to obtain more detailed information (such as the period), although this information might be available for a less sky-uniform sample due to the minimum number of *G*-band FoV observations of 12 or 20, depending on the SOS type. The classification contains only a label and “score” and has generally higher contamination rates, but it contains larger samples that are (more) uniformly distributed over the sky due to the minimum number of *G*-band FoV observations, which are 5 for RR Lyrae stars, 6 for δ Scuti and SX Phoenicis stars, 9 for Cepheids, and 12 for LPVs. These

differences between classification and SOS in contamination rate and sample size are part of the pipeline design in which classification results feed into most of the SOS modules, as shown in Fig. 2 and discussed in more detail in the DR1 paper (Eyer et al. 2017).

4. Results

4.1. Observation filtering and operators

The variability processing makes use of photometric data provided in units of fluxes (e/s), and then converts them into magnitudes using the magnitude zero-points defined in Evans et al. (2018). In our pipeline, this transformation is done by our `GaiaFluxToMagOperator` operator. Several additional operators were used to remove FoV transits by flagging and filtering them when they did not meet the required criteria, that is, when they were of insufficient quality. The chain of consecutive operators is described briefly here, for more details see the DR2 documentation:

- `RemoveNaNNegativeAndZeroValuesOperator` flagged and removed transits with NaN, negative, or zero flux values.
- `RemoveDuplicateObservationsOperator` flagged and removed pairs of transits with too close observation times (essentially one transit from the pair was incorrectly assigned to the source). This operator was responsible for the majority of the transit removals for *G* magnitudes brighter than ~ 8 ; see Mowlavi et al. (2018) for more details.
- `GaiaFluxToMagOperator` converted the *Gaia* fluxes into magnitude using the *Gaia* zero-point magnitudes.
- `ExtremeValueCleaning` flagged and removed transits with unrealistically faint magnitudes. The cut values were chosen to be $G \geq 25$, $G_{BP} \geq 24$, and $G_{RP} \geq 22$.
- `ExtremeErrorCleaningMagnitudeDependent` flagged and removed transits with extreme magnitude errors. For *G*, we used both a lower (0.01%) and upper (99.7%) threshold based on the observed distribution of transit magnitude errors (as a function of magnitude), while for G_{BP} and G_{RP} , we used only an upper threshold (99.9%).
- `RemoveOutliersFaintAndBrightOperator` flagged and removed the time-series outliers of the source. Different criteria were used to account for the distribution of magnitudes or of the magnitude errors in the time series.

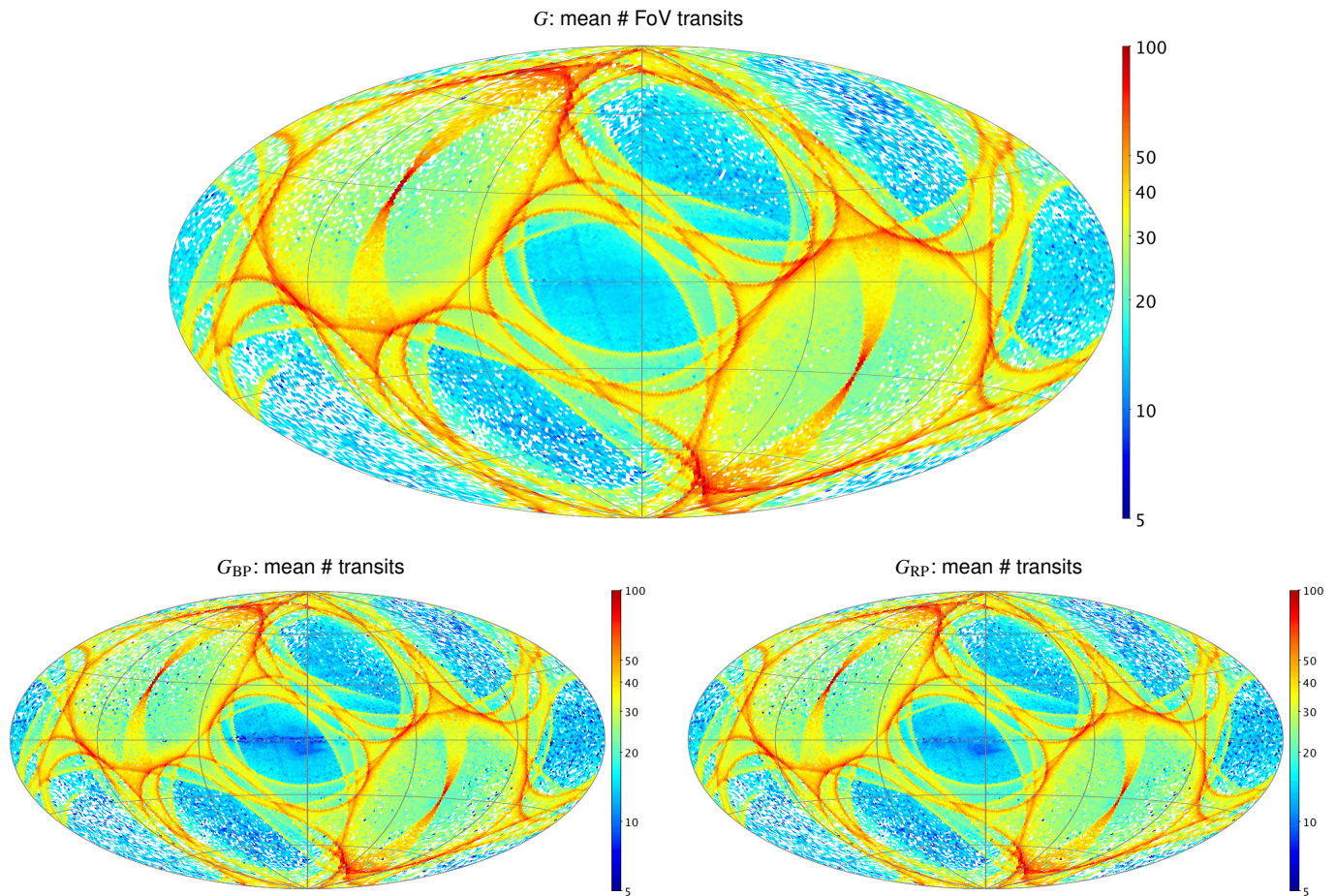


Fig. 4. Selected (i.e., not-rejected) number of transits in the time series of the G , G_{BP} , and G_{RP} photometric bands in Galactic coordinates for all 550 737 published sources. These numbers have been aggregated in sky bins of 0.84 deg^2 (level 6 HEALPix) and their mean value is shown. The colour scale is clipped at 100, although the number of transits reaches up to 242, as shown in Fig. 5. The choice of the rainbow colour scale has been motivated by the convenient colour divisions between 5, 10, 20, 30, 40, 50, and 100 transits. Galactic longitude increases to the left side.

The time series following the above last operator was then the input for the different variability modules, and all transits that were flagged and filtered in the operator chain can be identified in the published time series by the `rejected_by_variability` flag (see Sect. 4.2). The resulting number of selected (i.e., not-rejected) transits can be found in the `num_selected_g_fov/bp/rp` fields of the `vari_time_series_statistics` table. In Figs. A.1–A.3 the rejected transits are plotted as crosses. However, some SOS modules (e.g., SOS Cep and RRL, SOS rotation modulation) applied individual stricter conditions that were tuned for their analysis. We note that the SOS short timescale module also used another operator, `RemoveOutlierPerTransitOperator`, to remove CCD outliers per transit.

The sky distribution of the selected number of transits in the G , G_{BP} , and G_{RP} time series are shown in Fig. 4 for all 550 737 published sources. The distribution of the mean values per sky-pixel is dominated by the scanning law, which is also visible in the large masked regions of the ≥ 20 FoV panel of Fig. 1. Figure 5 shows a histogram of the time-series lengths. The minimum number of transits in G is 5.

4.2. Time series, statistics, and variability flag

The full list of variable stars published in DR2 can be easily identified in the `gaia_source` table by the field

`phot_variable_flag`, which is set to `VARIABLE`, while for all other sources it is set to `NOT_AVAILABLE`. The time series of all variable stars are published in G , G_{BP} , and G_{RP} . In contrast to DR1, the time series are no longer provided in a separate archive table, but instead are accessible in a Virtual Observatory Table⁴ (VOTable) linked⁵ via the field `epoch_photometry_url` in the `gaia_source` table. For various reasons, a fraction of the transit observations was not photometrically processed in DR2, resulting in null values. The variability processing flags per FoV transit (see previous section) are provided in the column `rejected_by_variability` of the VOTable epoch photometry. Additional flags from the photometric processing are available for transits in the column `rejected_by_photometry`. Unchanged from DR1 is the expression of observation time in units of barycentric JD (in TCB) in days -2455197.5 ; see DR2 documentation or Eyer et al. (2017) for more details.

All time series are provided in both flux and magnitude, where the flux is provided with the associated standard error. The magnitude-transformed error value is omitted given its non-symmetric nature⁶. The Vega-system flux to magnitude zero-points are defined in Evans et al. (2018).

⁴ See <http://www.ivoa.net/documents/VOTable/>.

⁵ For tutorials on using this datalink, see the *Gaia* archive help webpage <http://gea.esac.esa.int/archive-help>. Time-series bulk download is possible from <http://cdn.gea.esac.esa.int/Gaia/>.

⁶ Practical approximation: $\text{mag}_{\text{error}} \approx (2.5/\ln(10)) \cdot \text{flux}_{\text{error}}/\text{flux}_{\text{value}}$.

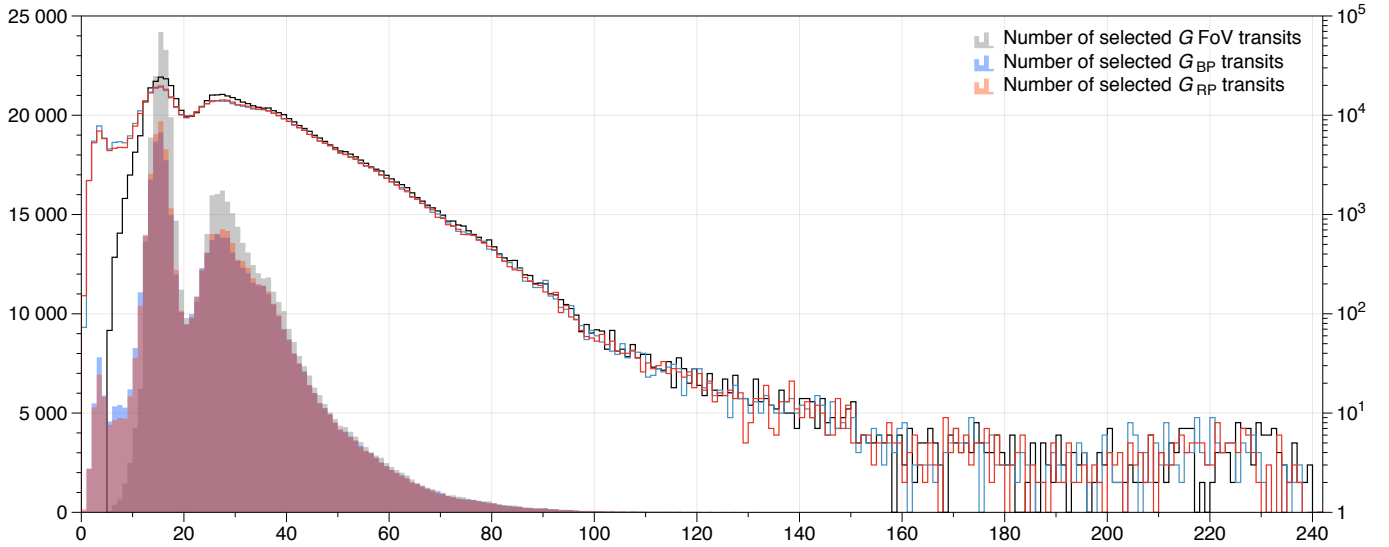


Fig. 5. Histogram of selected (i.e., not-rejected) number of transits in time series of the G , G_{BP} , and G_{RP} photometric bands for the 550 737 published variable sources. The filled bars correspond to the linear scale on the left side, while the solid lines corresponds to the logarithmic scale on the right side of the plot. The bin size is 1 FoV transit.

Several basic statistical parameters of the cleaned⁷ time series are listed in the `vari_time_series_statistics` table. The variable stars will furthermore have an entry in at least one of the `vari_*` tables, see Fig. 3 for their numbers. An overview of time-series statistics per *Gaia* band and types is provided in Tables A.1 and A.2. The mean magnitudes provided in `gaia_source` (`phot_g/bp/rp_mean_mag`) differ from the mean magnitudes in `vari_time_series_statistics` (`mean_mag_g_fov/bp/rp`) because they are calculated differently (see Evans et al. 2018) and because the filtering of the light curves was different; furthermore, the `gaia_source` mean photometry can be absent, but present in the `vari_time_series_statistics` table (together with the related time series in `epoch_photometry_url`) because various filters were applied by the photometry pipeline. The sky and median magnitude distributions of these sources are shown in more detail in Figs. 6 and 7.

4.3. Classification

In DR2 we introduce the output of the semi-supervised classifier targeted to identify high-amplitude variable stars over the whole sky, resulting from the `geq2` path described in Sect. 3. This is the aim of the `nTransits:2+` classification. Owing to the limited 22 months of data and rejected observations, the number of FoV transits per source can be very small, prohibiting the use of Fourier modelling parameters for a large portion of the sources. Previous studies such as Ivezić et al. (2000) and Hernitschek et al. (2016) have shown that it is still possible to identify good RR Lyrae star candidates with just a few observations. Our classifier was trained to identify high-amplitude variables of the type of RR Lyrae stars (anomalous RRd, see Soszyński et al. 2016; RRd, RRab, RRc), LPV (Mira-type and semi-regulars), Cepheids (anomalous Cepheids, δ Cepheids, type-II Cepheids), δ Scuti and SX Phoenicis stars, amongst the many other (non-)variable source types that exist. These classes can be found in the *Gaia* archive `vari_classifier_result` table with `best_class_name` = ARRd, RRd, RRAB, RRc, MIRA_SR, ACEP, CEP, T2CEP, and DSCT_SXPHE, respectively. Each

⁷ Selecting FoV transits with `rejected_by_variability=false`.

entry has an associated `best_class_score` between 0 and 1. Additional (non-published) classes were included in the prediction types (see Rimoldini et al., in prep. for all details). Descriptions of the classifier and the published classes can be found in the *Gaia* archive `vari_classifier_definition` and `vari_classifier_class_definition` tables, respectively.

The sky and magnitude distribution of the published classes, grouped by main type, is shown in the left panels of Figs. 6 and 7, and some time-series examples are shown in Figs. A.1 and A.2.

The published source samples were cleaned to reduce contamination levels caused by data and processing artefacts, as well as by genuine variability of other types. The results and details of this procedure are fully detailed in Rimoldini et al. in prep.

The output of this classifier for sources containing at least 12 measurements in their G -band time series was used as one of the two inputs to the SOS of RR Lyrae stars, Cepheids, and LPV, as shown in Fig. 2, hence accounting for the overlap between classification and SOS results in Fig. 3. Some SOS modules may reclassify the type in the SOS tables, see Sect. 4.6 An overview of the inferred completeness and contamination with respect to several cross-match catalogues of the different types is provided in Table 2.

4.4. Specific object studies

In the `geq20` path described in Sect. 3, potential variable type candidates that are identified in the supervised classification and special variability detection (SVD) parts of the pipeline were analysed in SOS modules, which often derived a model and/or additional astrophysical properties. The SOS modules can also decide whether an object is not of the expected type and as a result not produce any output for it. We recall that a sub-sample of the sources processed in the `geq2` path was processed by some SOS modules (see Fig. 2). We present short descriptions of the SOS modules used for DR2 below.

The sky and magnitude distribution of each published type is shown in the right panels of Figs. 6 and 7, and example light curves are presented in Figs. A.1–A.3. Because SOS modules

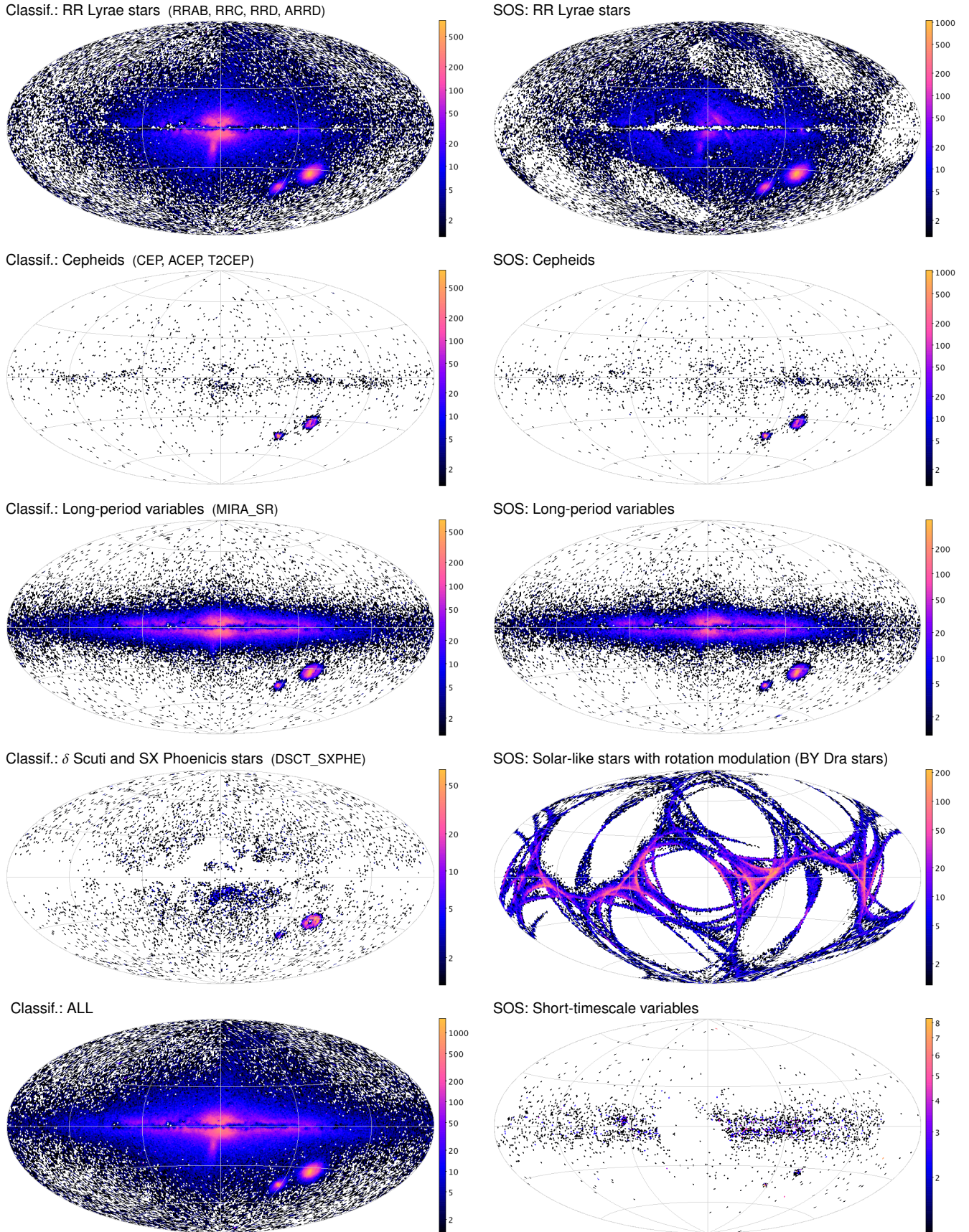


Fig. 6. Sky source densities (count deg^{-2}) in Galactic coordinates of the published sources in the classification table (*left column*; see Sect. 4.3) and SOS tables (*right column*; see Sect. 4.4). In the classification plots, the `best_class_name` entries were grouped by main type, as listed in parentheses. Galactic longitude increases to the left side.

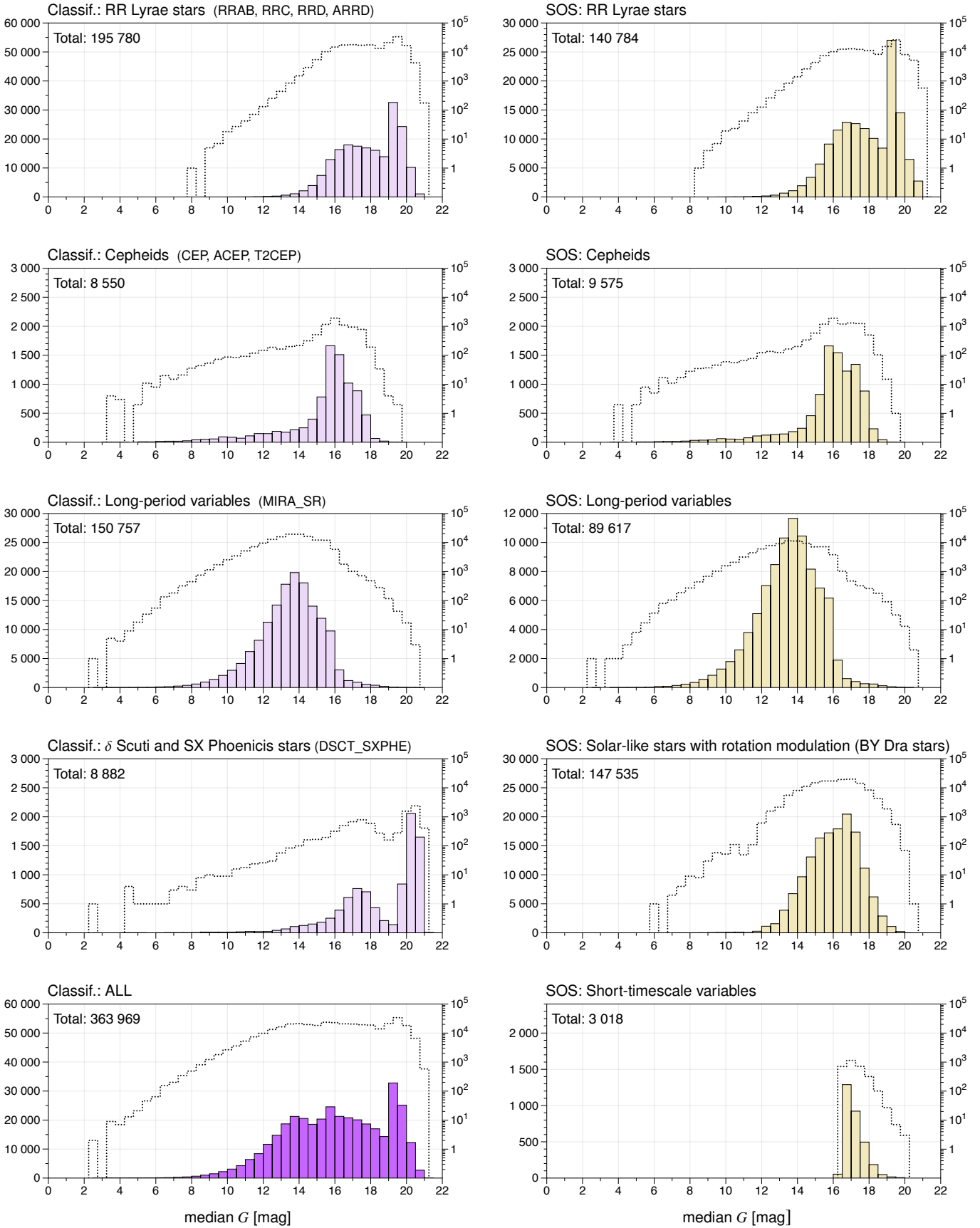


Fig. 7. Histogram counts of the median *G*-band magnitude distribution of the published classification classes (*left panels*; see Sect. 4.3) and SOS tables (*right panels*; see Sect. 4.4). In the classification plots, the `best_class_name` entries were grouped by main type, as listed in parentheses. The filled bars correspond to the linear scale on the left side of the plots (which varies in scale), while the dotted line corresponds to the logarithmic scale on the right side of the plots (always in the same scale). The bin size is fixed to 0.5 mag.

Table 2. Classification (absolute) completeness and contamination estimates with respect to available cross-matched reference catalogues.

Group	Class (best_class_name)	Catalogue (and region)	Completeness	Contamination
Cepheids	CEP, ACEP, T2CEP	OGLE-IV (LMC)	72%	>5%
Cepheids	CEP, ACEP, T2CEP	OGLE-IV (SMC)	68%	>5%
δ Scuti and SX Phoenicis	DSCT_SXPHE	Catalina (Drake et al. 2014)	99%	>13%
RR Lyrae stars	RRAB, RRC, RRD, ARRD	OGLE-IV (LMC)	60%	>9%
RR Lyrae stars	RRAB, RRC, RRD, ARRD	OGLE-IV (SMC)	64%	15%
RR Lyrae stars	RRAB, RRC, RRD, ARRD	OGLE-IV (Bulge)	49%	44%
RR Lyrae stars	RRAB, RRC, RRD, ARRD	Catalina (Drake et al. 2014)	63%	>5%
LPV	MIRA_SR	OGLE-III ^a (LMC+SMC)	40–50%	<5%
LPV	MIRA_SR	<i>Gaia</i> DR2 with ^b $\varpi/\sigma_\varpi > 10$	–	~7%

Notes. For more details on the validation of the classification results, we refer to (Rimoldini et al., in prep.). ^(a)Estimates are restricted to sources with peak-to-peak amplitude >0.2 mag. ^(b) ϖ and σ_ϖ refer to the DR2 parallax and associated error.

Table 3. SOS (absolute) completeness and contamination estimates with respect to cross-matched reference catalogues.

Processing path	SOS type	Catalogue (and region)	Completeness	Contamination
geq12 & geq20	Cepheids	OGLE-IV (LMC, limited region)	74%	~5% ^a
geq12 & geq20	Cepheids	OGLE-IV (Bulge)	3%	^b
geq12 & geq20	RR Lyrae stars	OGLE-IV (LMC, limited region)	64%	~9% ^a
geq12 & geq20	RR Lyrae stars	OGLE-IV (Bulge)	15%	^b
geq12 & geq20	LPV	OGLE-III ^c (SMC+LMC)	~30%	<few%
geq12 & geq20	LPV	OGLE-III ^c (SMC+LMC)	~30%	<few%
geq20	Short timescale	OGLE-II, III, IV (MCs)	0.05%	10–20%
geq20	Rotational modulation	The Pleiades (Hartman et al. 2010)	0.7–5.0%	<50% ^d

Notes. Processing paths are based on the number of selected FoV transits, as shown in Fig. 2 and discussed in Sect. 3. See Sect. 4.4 for additional contamination rate estimates for random and sky-uniform samples. For more details on the validation of the SOS type results, see Sect. 4.4 and the references therein. ^(a)Estimated from non-cross-match random samples inspection discussed in Sect. 4.4, extending partially beyond OGLE-IV footprint. ^(b)No reliable contamination rate for the Bulge has been derived, but in the Galactic disk footprint (see Udalski 2017), contamination of the SOS Cepheids and RR Lyrae stars might be of the order of several tens of percent (A. Udalski, priv. comm. based on unpublished OGLE-IV data). ^(c)Estimates are restricted to sources with a peak-to-peak amplitude >0.2 mag and periods longer than 60 days. ^(d)The 50% “contamination” quantifies the non-occurrence in the reference catalogue; the actual contamination of grazing binaries is estimated to be <1% (see Sect. 4.4.4).

can be fed by (partially) independent pipeline paths, it is not excluded that a source is listed in multiple SOS tables, as discussed in Sect. 4.5, or has a different classification and SOS type, as discussed in Sect. 4.6. Despite extensive validation efforts, the nature of our processing is an automated and statistical characterisation of the sources, hence identified sources that are not yet known in the literature should be considered to be candidates, unless they are specifically validated in one of the SOS articles referenced below.

An overview of the inferred (absolute) sky completeness and contamination with respect to existing catalogues of the different types is provided in Table 3; see sections below and the references therein for more details. Many of the listed reference catalogues have a (very) limited sky coverage, and hence the contamination might be locally biased. To assess contamination in the whole SOS published samples, we performed the following additional visual inspection for each SOS output separately: we took a random sample of 500 sources from the published results, as well as a sky-uniform⁸ sample of 500 sources (non-overlapping with the random sample). The former sample follows the general sky distribution as shown in Fig. 6, while the latter achieves an almost sky-uniform

sampling and hence also draws samples from very low density regions. The results are summarised below for each SOS result:

- *Cepheids*: 5% of the random sample was visually rejected, of which 3% might be due to binaries or ellipsoidals. Most of this sample is located in the Magellanic Clouds. Of the sky-uniform sample, 15% was visually rejected, of which 11% might be due to binaries or ellipsoidals.
- *RR Lyrae stars*: 9% of the random sample was visually rejected, of which 4% are faint ($20.0 < G < 20.7$ mag). Of the sky-uniform sample, 12% was visually rejected, of which 5% are faint ($19.9 < G < 20.7$ mag).
- *Long-period variables*: 0.1% (1) was visually rejected; it looks like a young stellar object, confirmed by its position in the Hertzsprung–Russel (HR) diagram.
- *Short-timescale variability*: 0.3% was visually rejected, and about 24% seem to exhibit short-timescale periodic variability, but this can currently not be confirmed with a high confidence.
- *Rotation modulation*: 0.3% was visually rejected.

These identified contamination rates are compatible⁹ with the cross-match sample contamination rates found in Table 3.

⁸ Approximate sky-uniform sampling is achieved by first grouping the sources in 12 288 bins (level 5 HEALPix; see Górski et al. 2005) of about 3.6 deg² and then randomly drawing non-empty pixels and extracting a source at random from it.

⁹ For Cepheids and RR Lyrae stars, the contamination rates quoted in Table 3 were taken from these analyses because contaminants from the OGLE-IV catalogues were already removed from the published results; (see Clementini et al. 2018 for more details).

4.4.1. Cepheids and RR Lyrae stars

The SOS Cepheids and RR Lyrae module (SOS Cep and RRL; Clementini et al. 2016, 2018), confirmed, characterised, and sub-classified in type 140 784 RR Lyrae stars and 9575 Cepheids of the candidates provided by the classifiers. The module can reclassify the type provided by the classifiers from Cepheid to RR Lyrae star, and vice versa. Amongst the RR Lyrae stars, the following number of sub-types were identified: 98 026 fundamental mode pulsators (RRab), 40 380 first-overtone pulsators (RRc), and 2378 double-mode pulsators (RRd). Amongst the Cepheids, there were 8890 δ Cepheids, 100 anomalous Cepheids, and 585 type II Cepheids, of which 223 were classified as BL Herculis, 253 as W Virginis, and 109 as RV Tauri.

A rough estimate of completeness and contamination was obtained by comparing our samples with OGLE catalogues of RR Lyrae stars and Cepheids in the LMC. We selected an area of the LMC centred at $\alpha = 83^\circ$ and $\delta = -67.5^\circ$ and extending from 67.5 to 97.5° in right ascension and from -63.5 to -73° in declination. Completeness of our RR Lyrae sample in this area is 64%, and contamination is less than 0.1%. In the same area, the Cepheid sample is 74% complete with 100% purity. A more in-depth description and analyses of the results, as well as the origin of the Bulge completeness rates quoted in Table 3, can be found in Clementini et al. (2018). Of the 599 Cepheids and 2595 RR Lyrae stars that were identified in DR1 (Clementini et al. 2016; Eyer et al. 2017), 533 Cepheids and 2517 RR Lyrae stars are also available in this DR2 release.

4.4.2. Long-period variables

Long-period variables are red giants that are characterised by long periods and variability amplitudes of up to a few magnitudes. They are identified by classification as MIRA_SR. On the basis of data available for DR2, CU7 selected 151 761 LPV candidates that fulfil the selection criteria for LPVs, which included a minimum of 12 points in G , a $G_{BP} - G_{RP}$ colour higher than 0.5 mag, a variability amplitude (quantified with the 5–95% trimmed range in G) larger than 0.2 mag, and a correlation between G and $G_{BP} - G_{RP}$ higher than 0.5. Of these, 150 757 are published in the classification table, and 89 617 have SOS results. The detailed distribution into these two tables is indicated in Fig. 3. The SOS LPV module processes LPV candidates when their period is longer than 60 days. The steps applied include period search, the determination of bolometric correction from G_{BP} and G_{RP} fluxes, the identification of red supergiants, and the computation of absolute magnitude. It must be noted that all LPV specific attributes published in DR2 that depend on the parallax must be recomputed based on the published parallaxes, because these LPV-specific attributes were computed based on an older version of the parallaxes that was available at the time of variability processing within the consortium. We refer to Mowlavi et al. (2018) for additional information. The light curves of all LPV candidates, whether identified in the classification table or in the SOS table, are published in DR2.

The completeness and contamination numbers for the LPV candidates towards the Magellanic Clouds are computed relative to a sub-sample of the OGLE-III catalogue of LPVs (Soszyński et al. 2009, 2011). The sub-sample consists of all OGLE-III LPVs with I -band amplitude larger than 0.2 mag for the classification completeness and contamination estimates, and is further restricted to the OGLE-III LPVs with periods longer than 60 d for the SOS completeness and contamination estimates. Additionally, an all-sky contamination estimate is provided for

Gaia LPV candidates that have small relative parallax uncertainties based on their position in the HR diagram. The main contaminants in this sample consist of young stellar objects. Details of SOS LPV processing and estimates of completeness and contamination can be found in Mowlavi et al. (2018), where the DR2 data set of LPV candidates are presented and further analysed.

4.4.3. Short-timescale variability

Short-timescale variable sources are astrophysical objects showing any variability in their optical light curve with a characteristic timescale of the variation shorter than 1 d (Roelens et al. 2017). A variety of astronomical objects are known to exhibit such fast photometric variations, including both periodic and transient variability, involving amplitudes from a few millimagnitudes to a few magnitudes, with different variability characteristics and phenomena at the origin of variation, from pulsations to flares to eclipsing systems. The diverse variability types targeted by the short-timescale module range from short-period binary stars to pulsating white dwarfs (e.g., ZZ Ceti stars or V777 Herculis stars) and cataclysmic variables such as AM Canum Venaticorum stars. However, for the *Gaia* DR 2, the short-timescale variability processing is oriented towards periodic variability, and the exact variable type of the detected short-timescale candidates is not yet determined. Both the investigation of transient variability (Wevers et al. 2018) and the further classification of the candidates are foreseen for the future *Gaia* data releases.

As mentioned previously, the identification of the short-timescale variable candidates published in *Gaia* DR2 involves particular variability detection methods that are specifically tuned to detect fast photometric variability, complementarily to the general variability detection approach (which may miss some of these events because they are so fast). These methods are applied to a preselected subset of the *Gaia* data set (Eyer et al. 2017): for *Gaia* DR 2, the short-timescale variability analysis is restricted to relatively faint sources ($G \sim 16.5$ – 20 mag) whose G per-CCD time series indicates that they are likely to show rapid variations (Roelens et al. 2018). Consequently, the candidates resulting from the short-timescale module are expected to overlap with other short-period SOS variables such as RR Lyrae and δ Scuti stars with periods shorter than 1 d, even though not there is no complete overlap with these SOS variable candidates.

The SOS short-timescale module identified 3018 short-timescale, suspected periodic candidates. These bona fide, short-timescale, suspected periodic candidates were identified based on the combination of the variogram analysis (Eyer & Genton 1999; Roelens et al. 2017), least-squares frequency search (Zechmeister & Kürster 2009), analysis of the environment of candidate sources over the sky, and a series of selection criteria from various statistics, such as the Abbe or inter-quartile range (IQR) values in the three bands (G , G_{BP} , and G_{RP}).

Completeness and contamination numbers for the short-timescale variable candidate sample, estimated by comparing OGLE-II, III, IV, and *Gaia* data in the Magellanic Clouds, are listed in Table 3. Furthermore, about 50% of the known OGLE-III and IV variables belonging to the short-timescale sample in these regions are longer period variables (i.e., with period >1 d). However, most of these contaminant sources have periods of a few days and quite high amplitudes, hence although they are no true short-period variables, their presence in the short-timescale published sample is justified. A more detailed description and analysis of the results can be found in Roelens et al. (2018).

4.4.4. Rotational modulation

Stellar flux modulation induced by surface inhomogeneities and rotation is searched for in a region of the HR diagram, broadly embracing stars of spectral type later than F on the main sequence. This type of variability, often indicated in the literature as BY Draconis stars (BY Dra) regardless of possible binarity, is an indication of active regions (dark spots and bright faculae on the stellar photosphere) that are produced by stellar magnetic activity. The evolution of stellar magnetic fields, similar to solar fields, produces variability phenomena on a wide range of timescales (see e.g., [Lanza et al. 2004, 2006](#); [Distefano et al. 2012](#), and references therein); one of the most prominent phenomena is the rotational modulation itself, from which the stellar rotation period can be inferred.

The variability pipeline comprises two packages dedicated to the detection and characterisation of solar-like stars with rotational modulation. First, solar-like variable candidates are selected, and if they are confirmed, they are studied in more detail to determine stellar rotation periods and other properties. Figure A.3 shows one of the published light curves together with the segments in which a significant (similar) period was detected.

We detected some 7×10^5 periodic variables in the pre-selected HR diagram region, which excludes areas populated by pulsating variables. The remaining expected main contaminants are eclipsing binaries and spurious detections derived from an incomplete sampling. To filter out these cases as well as possible, we applied filters that exclude sources whose period-folded light curves have a significantly uneven distribution in phase, with significant gaps, and are far from being sinusoidal (see [Lanzafame et al. 2018](#), for details). The final DR2 clean sample contains 147 535 rotational modulation candidates and fills 38% of the whole sky when divided into bins of $\approx 0.84 \text{ deg}^2$ (level 6 HEALPix); see for example Fig. 6.

An estimate of the final completeness and contamination of the DR2 rotational modulation variables is hampered by the fact that the occurrence rate of the BY Dra phenomenon is largely unknown so far. It is expected that all low-mass dwarfs are magnetically active to some extent, and the rotational modulation detectability essentially depends on the instrumental sensitivity, and also on the active region distribution and the phase of the magnetic cycle at the epoch of observations. New low-mass dwarfs displaying rotational modulation are continuously detected with the ever-wider span and increasing sensitivity of modern surveys, none of which has the full sky coverage capabilities of *Gaia*, however. At this stage, it is possible to perform some meaningful comparison only with the [Hartman et al. \(2010\)](#) and [Rebull et al. \(2016\)](#) observations of the Pleiades, with which, nevertheless, the DR2 geq20 sky coverage still overlaps only marginally. Assuming that the [Hartman et al. \(2010\)](#) catalogue lists all the BY Dra in its FoV down to $G \approx 14.5$, we estimate that the completeness of the BY Dra sample is 14% down to $G \approx 16.5$ in the overlapping field. Assuming that this value is uniform over the whole sky, we estimate a completeness upper limit of 5%. At the other extreme, we may assume that all low-mass dwarfs are BY Dra variables. Then, comparing with all stars observed by *Gaia* in the same sky region and magnitude range, we estimate a lower completeness limit of 0.7%.

We may roughly estimate an upper limit for contamination by assuming that the *Gaia* BY Dra detections that are not present in the [Hartman et al. \(2010\)](#) or in the [Rebull et al. \(2016\)](#) samples in the common region on the sky are variables of other types. This upper limit is obviously largely overestimated given that no BY Dra catalogue can be deemed complete to date, as also

testified by the fact that there are sources in the [Hartman et al. \(2010\)](#) sample that are not contained in the [Rebull et al. \(2016\)](#) sample, and vice versa.

At the other extreme, we can make an educated guess regarding the contamination of close grazing binaries that would be incorrectly classified as BY Dra because of the similarity of the light curves. If the orbital period is long, then the grazing eclipse occurs only in a small part of the orbit, and it does not look like rotation modulation. We assume an upper limit period of 10 days for this contaminating effect. The fraction of stars with such short-period companions is only about 2–5% ([Raghavan et al. 2010](#)), of which only 2–5% are grazing binaries (at most), which means 0.04–0.25 % of the stars, while we did not yet take into account that for most binaries, the secondaries are substantially smaller. Even though we cannot consider the rotation modulation sample a random subset of *G* and later-type stars (which is the assumption in above estimate), it seems unlikely that this type of contamination would exceed 1%. The same conclusion is reached from our internal validation against known and newly identified (but not published) eclipsing binaries, based on which, we estimate an upper limit of 0.5% grazing contamination. From the estimated grazing binary contamination and non-occurrence in the reference catalogue discussed above, we estimate a contamination level ranging from <1% to $\approx 50\%$, as listed in Table 3. More details on the SVD solar-like and SOS rotational modulation packages and results can be found in [Lanzafame et al. \(2018\)](#).

4.5. Sources with multiple SOS types

There are 80 sources in DR2 with an entry in more than one SOS table. These cases are all overlaps between short-timescale objects and RR Lyrae stars (72), Cepheids (5), and rotation modulation objects (3), all of which are justified by their overlapping type definition. No overlap is found with long-period variables.

4.6. Sources with different classification and SOS type

Differences are found between classification types (as obtained by supervised classification in the geq2 path) and the types that are eventually derived in SOS results (either from the geq12 or geq20 paths). In particular, one source was classified as DSCT_SXPHE, but also appears in SOS rotational modulation. Similarly, one source was classified as MIRA_SR, was not confirmed by SOS LPV, but appears in SOS rotational modulation. Of the sources classified as a Cepheid (any type), 6 appear in more than one SOS table. One is a confirmed SOS Cep and also appears in SOS short timescale, while 3 appear as SOS rotational modulation and 2 as SOS short timescale. Of the sources classified as an RR Lyrae star (any type), 70 appear in the SOS short timescale table, one was instead confirmed as a Cepheid by SOS (shown in the top panel of Fig. A.2), and many were confirmed as RR Lyrae star by SOS. Overall, a reassignment of the type from RRL into Cep by SOS occurred in 618 cases, but only in 77 cases from Cepheid to RR Lyrae star by SOS. We have decided to retain the original classification type despite the possible reassignment by the SOS Cep&RRL module or in cases where they were confirmed as another type by SOS modules.

5. Conclusions

In this second *Gaia* data release, a sample of 550 737 variable sources with their three-band time series and analysis results

has been released to the astronomical community, showcasing samples of high-amplitude variable candidates distributed over (a large portion of) the whole sky. It demonstrates the immense potential of *Gaia* data to provide an unbiased all-sky photometric survey together with astrometric and spectroscopic data. Although the variability analyses in DR2 did only mildly rely on the astrometric data and on none of the spectroscopic data, this dependence will become much heavier in future data releases.

Several caveats have been outlined in this paper:

- Most provided stellar samples are rather incomplete as a result of the current state of data calibration (e.g., even well-known variables might be missing) or the limited scanning law coverage for the minimum number of selected observations,
- known incomplete flagging of outliers in the time series,
- existence of uncalibrated systematics or spurious calibration error signals in some of the time series,
- the probabilistic and automated nature of this work implies certain completeness and contamination rates (Tables 2 and 3), hence even well-known literature sources could be misidentified.

Overall, we estimate¹⁰ to have identified the following numbers of new variables: a few hundred Cepheids, a few tens of thousand RR Lyrae stars, about one hundred thousand stars with rotation modulation, several tens of thousand long-period variables, a few thousand δ Scuti and SX Phoenixis stars, and a few thousand short-timescale variables. In total, this is about half of the variable sources released in this DR2. Compared to the DR1 example, we increased the sample size about 100 times, which is expected to result in a large number of novel studies, regardless of whether they are combined with existing data sets.

The next variability release will be based on data that are better calibrated yet again and on a longer time baseline. It will also include BP/RP spectrophotometry and radial velocities from RVS, which will allow us to improve the quality and quantity of the *Gaia* DR3 release.

Acknowledgements. We dedicate this paper to our dear friend and colleague Jan Cuypers (1956–2017), whose passing has left a great void in our variability group. Thank you, Jan, for your friendship and lifelong contribution to science. We would like to thank the referee, Andrzej Udalski, for various suggestions and providing so far unpublished OGLE-IV estimates that improved the contents of this paper. We thank Xavier Luri, Antonella Vallenari, and Annie Robin for valuable feedback during the preparation of this paper. We also acknowledge Mark Taylor for creating and implementing new requested features in the astronomy-oriented data handling and visualization software TOPCAT (Taylor 2005). This work has made use of data from the ESA space mission *Gaia*, processed by the *Gaia* Data Processing and Analysis Consortium (DPAC). Funding for the DPAC has been provided by national institutions, some of which participate in the *Gaia* Multilateral Agreement, which include, for Switzerland, the Swiss State Secretariat for Education, Research and Innovation through the ESA PRODEX program, the “Mesures d’accompagnement”, the “Activités Nationales Complémentaires”, the Swiss National Science Foundation, and the Early Postdoc. Mobility fellowship; Belgium, the BELgian federal Science Policy Office (BELSPO) through PRODEX grants; for Italy, Istituto Nazionale di Astrofisica (INAF) and the Agenzia Spaziale Italiana (ASI) through grants I/037/08/0, I/058/10/0, 2014-025-R.0, and 2014-025-R.1.2015 to INAF (PI M.G. Lattanzi); for France, the Centre National d’Etudes Spatiales (CNES). This research has received additional funding from the Israeli Centers for Research Excellence (I-CORE, grant No. 1829/12); the Hungarian Academy of Sciences through the Lendület Programme LP2014-17, and the János Bolyai Research Scholarship (L. Molnár and E. Plachy); the Hungarian National Research, Development, and Innovation Office through grants OTKA/NKFIH K-113117, K-115709, PD-116175, and PD-121203; the ÚNKP-17-3 program of the Ministry of Human Capacities of Hungary (Á.L.J.); the ESA Contract

No. 4000121377/17/NL/CBi; and the Polish National Science Center through ETIUDA grant No. 2016/20/T/ST9/00170; Charles University in Prague through PRIMUS/SCI/17 award; and from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 670519: MAMSIE).

References

- Arenou, F., Luri, X., Babusiaux, C., et al. 2018, *A&A*, **616**, A17 (*Gaia* 2 SI)
- Clementini, G., Ripepi, V., Leccia, S., et al. 2016, *A&A*, **595**, A133 (*Gaia* SI)
- Clementini, G., Ripepi, V., Molinaro, R., et al. 2018, *A&A*, submitted [arXiv:1805.02079] (*Gaia* 2 SI)
- Distefano, E., Lanzafame, A. C., Lanza, A. F., et al. 2012, *MNRAS*, **421**, 2774
- Drake, A. J., Graham, M. J., Djorgovski, S. G., et al. 2014, *ApJS*, **213**, 9
- ESA 1997, *ESA SP, The HIPPARCOS and Tycho Catalogues. Astrometric and Photometric Star Catalogues Derived from the ESA HIPPARCOS Space Astrometry Mission*, 1200
- Evans, D. W., Riello, M., De Angeli, F., et al. 2018, *A&A*, **616**, A4 (*Gaia* 2 SI)
- Eyer, L., & Genton, M. G. 1999, *A&AS*, **136**, 421
- Eyer, L., Mowlavi, N., Evans, D. W., et al. 2017, *A&A*, submitted [arXiv:1702.03295] (*Gaia* SI)
- Gaia* Collaboration (Brown, A. G. A., et al.) 2016, *A&A*, **595**, A2 (*Gaia* SI)
- Gaia* Collaboration (Brown, A. G. A., et al.) 2018a, *A&A*, **616**, A1 (*Gaia* 2 SI)
- Gaia* Collaboration (Eyer, L., et al.) 2018b *A&A*, submitted [arXiv:1804.09382] (*Gaia* 2 SI)
- Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, *ApJ*, **622**, 759
- Hartman, J. D., Bakos, G. Á., Kovács, G., & Noyes, R. W. 2010, *MNRAS*, **408**, 475
- Hernitschek, N., Schlafly, E. F., Sesar, B., et al. 2016, *ApJ*, **817**, 73
- Ivezić, Ž., Goldston, J., Finlator, K., et al. 2000, *AJ*, **120**, 963
- Ivezić, Ž., Smith, J. A., Miknaitis, G., et al. 2007, *AJ*, **134**, 973
- Lanza, A. F., Rodonò, M., & Pagano, I. 2004, *A&A*, **425**, 707
- Lanza, A. F., Messina, S., Pagano, I., & Rodonò, M. 2006, *Astron. Nachr.*, **327**, 21
- Lanzafame, A. C., Distefano, E., Messina, S., et al. 2018, *A&A*, **616**, A16 (*Gaia* 2 SI)
- Lindgren, L., Hernández, J., Bombrun, A., et al. 2018, *A&A*, **616**, A2 (*Gaia* 2 SI)
- Molnár, L., Plachy, E., Juhász, Á. L., & Rimoldini, L. 2018, *A&A*, submitted [arXiv:1805.11395] (*Gaia* 2 SI)
- Mowlavi, N., Lecoœur-Taïbi, I., Lebzelter, T., et al. 2018, *A&A*, in press, DOI: [10.1051/0004-6361/201833366](https://doi.org/10.1051/0004-6361/201833366) (*Gaia* 2 SI)
- Palaversa, L., Ivezić, Ž., Eyer, L., et al. 2013, *AJ*, **146**, 101
- Raghavan, D., McAlister, H. A., Henry, T. J., et al. 2010, *ApJS*, **190**, 1
- Rebull, L. M., Stauffer, J. R., Bouvier, J., et al. 2016, *AJ*, **152**, 113
- Riello, M., De Angeli, F., Evans, D. W., et al. 2018, *A&A*, **616**, A3 (*Gaia* 2 SI)
- Roelens, M., Eyer, L., Mowlavi, N., et al. 2017, *MNRAS*, **472**, 3230
- Roelens, M., Eyer, L., Mowlavi, N., et al. 2018, *A&A*, in press, DOI: [10.1051/0004-6361/201833357](https://doi.org/10.1051/0004-6361/201833357) (*Gaia* 2 SI)
- Soszyński, I., Udalski, A., Szymański, M. K., et al. 2009, *Acta Astron.*, **59**, 239
- Soszyński, I., Udalski, A., Szymański, M. K., et al. 2011, *Acta Astron.*, **61**, 217
- Soszyński, I., Udalski, A., Poleski, R., et al. 2012, *Acta Astron.*, **62**, 219
- Soszyński, I., Smolec, R., Dziembowski, W. A., et al. 2016, *MNRAS*, **463**, 1332
- Taylor, M. B. 2005, in *Astronomical Data Analysis Software and Systems XIV*, eds. P. Shopbell, M. Britton, & R. Ebert, *ASP Conf. Ser.*, **347**, 29
- Udalski, A. 2017, *Eur. Phys. J. Web Conf.*, **152**, 01002
- Wevers, T., Jonker, P. G., Hodgkin, S. T., et al. 2018, *MNRAS*, **473**, 3854
- Zechmeister, M., & Kürster, M. 2009, *A&A*, **496**, 577

¹⁰ From occurrence in a great variety of cross-match catalogues (most of which are documented in the DR2 documentation), and taking into account that our results have a varying degree of contamination.

¹ Department of Astronomy, University of Geneva, Ch. des Maillettes 51, 1290 Versoix, Switzerland
e-mail: Berry.Holl@unige.ch

² Department of Astronomy, University of Geneva, Ch. d’Ecogia 16, 1290 Versoix, Switzerland

³ SixSq, Rue du Bois-du-Lan 8, 1217 Geneva, Switzerland

⁴ GÉPI, Observatoire de Paris, Université PSL, CNRS, Place Jules Janssen 5, 92195 Meudon, France

⁵ INAF – Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Gobetti 93/3, 40129 Bologna, Italy

⁶ Instituut voor Sterrenkunde, KU Leuven, Celestijnenlaan 200D, 3001 Leuven, Belgium

⁷ Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

- ⁸ Large Synoptic Survey Telescope, 950 N. Cherry Avenue, Tucson, AZ 85719, USA
- ⁹ Università di Catania, Dipartimento di Fisica e Astronomia, Sezione Astrofisica, Via S. Sofia 78, 95123 Catania, Italy
- ¹⁰ INAF – Osservatorio Astrofisico di Catania, Via S. Sofia 78, 95123 Catania, Italy
- ¹¹ University of Vienna, Department of Astrophysics, Tuerkenschanzstrasse 17, 1180 Vienna, Austria
- ¹² Department of Geosciences, Tel Aviv University, Tel Aviv 6997801, Israel
- ¹³ Dipartimento di Fisica e Astronomia, Università di Bologna, Via Piero Gobetti 93/2, 40129 Bologna, Italy
- ¹⁴ Departamento de Astrofísica, Centro de Astrobiología (INTA-CSIC), PO Box 78, 28691 Villanueva de la Cañada, Spain
- ¹⁵ INAF – Osservatorio Astronomico di Capodimonte, Via Moiarriello 16, 80131 Napoli, Italy
- ¹⁶ School of Physics and Astronomy, Tel Aviv University, Tel Aviv 6997801, Israel
- ¹⁷ Departamento de Inteligencia Artificial, UNED, C/ Juan del Rosal 16, 28040 Madrid, Spain
- ¹⁸ Department of Astrophysics/IMAPP, Radboud University, PO Box 9010, 6500 GL Nijmegen, The Netherlands
- ¹⁹ European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching b. München, Germany
- ²⁰ Laboratoire d'astrophysique de Bordeaux, Université Bordeaux, CNRS, B18N, allée Geoffroy Saint-Hilaire, 33615 Pessac, France
- ²¹ Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA
- ²² Royal Observatory of Belgium, Ringlaan 3, 1180 Brussels, Belgium
- ²³ Konkoly Observatory, Research Centre for Astronomy and Earth Sciences, Hungarian Academy of Sciences, Konkoly Thege Miklós út 15-17, 1121 Budapest, Hungary
- ²⁴ Department of Astronomy, Eötvös Loránd University, Pázmány Péter sétány 1/a, 1117 Budapest, Hungary
- ²⁵ Villanova University, Department of Astrophysics and Planetary Science, 800 Lancaster Avenue, Villanova, PA 19085, USA
- ²⁶ Faculty of Mathematics and Physics, University of Ljubljana, Jadranska ulica 19, 1000 Ljubljana, Slovenia
- ²⁷ Academy of Sciences of the Czech Republic, Fricova 298, 25165 Ondřejov, Czech Republic
- ²⁸ Baja Observatory of University of Szeged, Szegedi út III/70, 6500 Baja, Hungary
- ²⁹ Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Bd de l'Observatoire, CS 34229, 06304 Nice cedex 4, France
- ³⁰ CENTRA FCUL, Campo Grande, Edif. C8, 1749-016 Lisboa, Portugal
- ³¹ EPFL SB MATHAA STAP, MA B1 473 (Bâtiment MA), Station 8, 1015 Lausanne, Switzerland
- ³² Warsaw University Observatory, Al. Ujazdowskie 4, 00-478 Warszawa, Poland
- ³³ Institute of Theoretical Physics, Faculty of Mathematics and Physics, Charles University in Prague, Prague, Czech Republic
- ³⁴ Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany

Table A.2. Statistical parameters for the SOS types published in DR2.

G band	Cepheids				RR Lyrae stars				Long-period var.				Rot. modulation				Short timesc.			
	15	28	118	81	13	27	81	69	12	27	69	24	43	88	19	32	69			
Filtered number of FoV transits	514.00	627.70	667.48	430.10	593.78	657.13	427.22	591.78	638.78	481.77	580.21	651.04	475.32	587.09	642.99					
Time duration (d)	1741.28	2018.29	2114.19	1761.88	2022.56	2163.10	1887.55	2036.56	2160.34	1835.54	2027.26	2183.81	1844.59	2026.86	2163.03					
Mean observation time (d)	8.07	15.82	17.98	13.08	17.59	20.18	7.87	13.08	16.93	12.38	16.07	18.86	16.13	16.90	18.47					
Min. magnitude	8.57	16.33	18.56	13.74	18.24	20.95	8.97	14.12	17.81	12.43	16.13	19.04	16.64	17.37	19.76					
Max. magnitude	8.39	16.11	18.29	13.49	17.98	20.61	8.50	13.60	17.28	12.40	16.10	18.95	16.51	17.10	18.87					
Median magnitude	8.40	16.13	18.31	13.51	18.02	20.64	8.48	13.59	17.27	12.40	16.10	18.96	16.49	17.08	18.86					
Magnitude range	0.13	0.49	1.12	0.26	0.65	1.33	0.21	0.53	3.15	0.01	0.05	0.30	0.14	0.44	2.43					
Standard deviation (mag)	0.04	0.16	0.37	0.08	0.19	0.37	0.06	0.16	1.08	0.00	0.01	0.08	0.04	0.12	0.56					
Skewness of magnitudes	-1.56	-0.29	0.93	-2.61	-0.44	0.57	-1.26	0.15	1.69	-2.21	0.05	2.24	-0.91	0.38	2.40					
Kurtosis of magnitudes	-1.84	-1.15	2.19	-1.83	-0.91	9.21	-1.94	-0.87	3.40	-1.48	-0.28	13.04	-1.55	-0.47	6.49					
MAD (mag)	0.04	0.17	0.45	0.06	0.19	0.42	0.02	0.15	1.44	0.00	0.01	0.08	0.03	0.11	0.43					
Abbe	0.17	0.67	1.19	0.50	0.97	1.41	0.06	0.28	0.61	0.14	0.54	1.20	0.71	0.99	1.42					
IQR (mag)	0.07	0.26	0.69	0.10	0.30	0.66	0.04	0.24	2.12	0.00	0.02	0.12	0.05	0.16	0.65					
G_{BP} band ^a																				
Filtered number of FoV transits	12	27	118	2	24	77	9	26	67	23	42	87	18	30	68					
Time duration (d)	493.23	627.60	667.05	182.20	588.20	655.81	421.03	590.19	638.78	461.80	579.89	650.80	427.53	585.24	642.58					
Mean observation time (d)	1741.25	2018.23	2119.45	1755.42	2020.82	2179.33	1885.59	2034.97	2168.89	1834.54	2027.21	2186.35	1837.85	2026.82	2164.76					
Min. magnitude	8.64	15.94	17.88	12.53	17.11	19.70	9.99	15.57	18.88	12.71	16.27	18.80	14.84	17.04	18.47					
Max. magnitude	9.36	16.71	19.12	14.02	18.62	22.37	11.37	17.35	22.62	12.89	16.83	21.59	16.62	17.89	20.30					
Mean magnitude	9.06	16.40	18.54	13.71	18.17	20.67	10.85	16.59	20.46	12.85	16.72	20.25	16.40	17.57	19.25					
Median magnitude	9.04	16.43	18.56	13.74	18.21	20.64	10.80	16.58	20.45	12.85	16.73	20.28	16.41	17.58	19.26					
Magnitude range	0.21	0.76	3.03	0.24	1.20	5.75	0.36	1.17	6.25	0.03	0.35	4.52	0.20	0.76	4.07					
Standard deviation (mag)	0.06	0.22	0.58	0.08	0.33	1.30	0.10	0.32	1.94	0.01	0.06	0.78	0.05	0.16	0.82					
Skewness of magnitudes	-4.18	-0.39	1.15	-4.19	-0.58	1.63	-3.18	0.01	1.61	-6.14	-1.74	2.55	-5.00	-0.60	2.16					
Kurtosis of magnitudes	-1.77	-0.78	21.82	-1.81	0.16	44.52	-1.91	-0.66	14.30	-1.04	6.72	42.92	-1.33	2.34	28.09					
MAD (mag)	0.05	0.21	0.54	0.06	0.27	0.91	0.05	0.29	2.51	0.01	0.03	0.44	0.04	0.11	0.53					
Abbe	0.20	0.72	1.25	0.47	0.99	1.47	0.08	0.35	1.06	0.31	0.94	1.28	0.52	1.02	1.42					
IQR (mag)	0.08	0.32	0.83	0.09	0.41	1.44	0.08	0.45	3.70	0.01	0.04	0.61	0.05	0.16	0.77					
G_{RP} band ^a																				
Filtered number of FoV transits	12	27	117	2	25	77	10	26	67	23	42	87	18	30	68					
Time duration (d)	493.23	627.60	666.98	184.18	588.70	655.97	425.29	590.45	638.78	474.04	580.14	650.80	428.93	585.42	642.58					
Mean observation time (d)	1741.25	2017.51	2119.45	1756.01	2019.99	2178.16	1887.12	2034.64	2166.20	1834.20	2026.69	2185.00	1838.43	2025.96	2164.88					
Min. magnitude	7.37	15.06	17.21	11.24	16.05	18.85	6.51	11.55	15.17	11.46	14.93	17.08	11.54	15.75	17.30					
Max. magnitude	7.92	15.76	18.08	13.27	17.45	20.97	7.55	12.58	16.14	11.84	15.38	17.89	15.57	16.43	19.07					
Mean magnitude	7.61	15.54	17.70	13.05	17.14	19.72	7.10	12.08	15.64	11.80	15.32	17.63	15.42	16.21	18.11					
Median magnitude	7.60	15.57	17.75	13.05	17.18	19.76	7.06	12.06	15.64	11.80	15.33	17.68	15.43	16.22	18.13					
Magnitude range	0.15	0.57	5.26	0.12	0.91	7.62	0.17	0.51	3.08	0.02	0.21	3.62	0.12	0.59	4.83					
Standard deviation (mag)	0.05	0.15	0.94	0.04	0.24	1.46	0.05	0.15	1.01	0.01	0.04	0.58	0.03	0.13	0.90					
Skewness of magnitudes	-5.86	-0.48	1.02	-5.41	-0.72	1.68	-3.75	0.13	1.73	-6.68	-2.83	2.76	-5.48	-0.74	2.02					
Kurtosis of magnitudes	-1.77	-0.41	45.26	-1.77	1.15	62.16	-1.95	-0.76	17.23	-1.11	11.49	49.34	-1.57	2.68	33.68					
MAD (mag)	0.04	0.14	0.37	0.03	0.19	0.55	0.02	0.14	1.32	0.00	0.02	0.11	0.02	0.09	0.41					
Abbe	0.21	0.78	1.27	0.48	1.00	1.46	0.07	0.32	1.01	0.31	0.95	1.24	0.52	1.04	1.44					
IQR (mag)	0.06	0.21	0.57	0.04	0.27	0.85	0.04	0.21	1.96	0.01	0.02	0.17	0.03	0.13	0.61					

Notes. The 1%, 50% (median), and 99% percentile are provided for each type. ^(a)For RR Lyrae stars, the percentiles for G_{BP} and G_{RP} were computed for a number of FoV transits > 1 , although there are stars with 0 or 1 transit remaining in the band after our cleaning. For the skewness and kurtosis, we included only RR Lyrae stars with > 2 and > 3 transits, respectively.

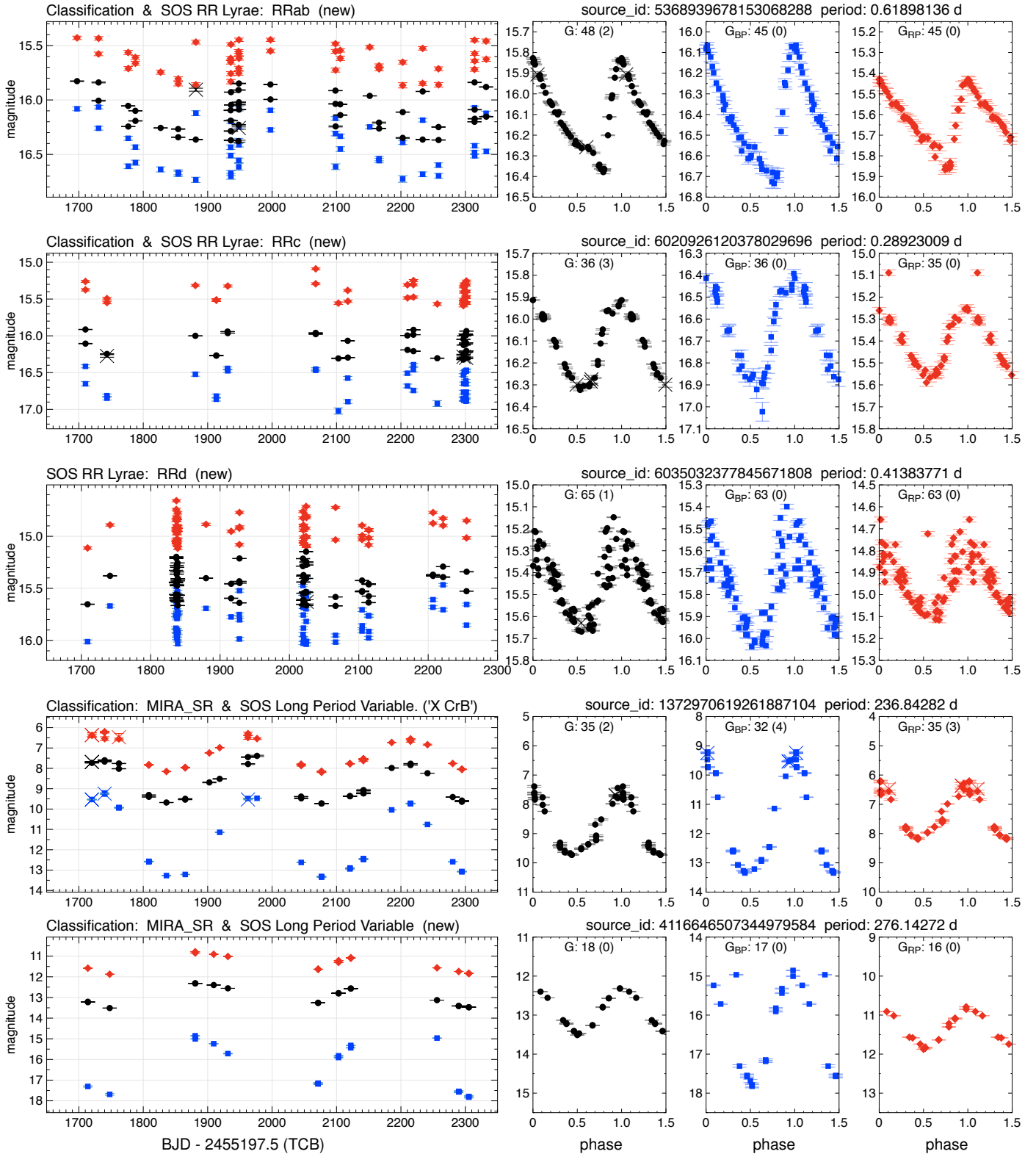


Fig. A.1. Example light curves of RR Lyrae stars and LPV in the data. For each band we show the valid FoV transits, as well as variability-rejected FoV transits (number in parentheses, data plotted with crosses). In most cases, two FoV transits occur within the 6 h spin period, causing indistinguishable overlapping points in the *left panels*. Some outliers are not flagged in the exported data (see e.g., the G_{RP} light curve of the RRc and RRd example); these transits were removed in the more strict outlier rejection used in the SOS Cep and RRL module, as mentioned in Sect. 4.1.

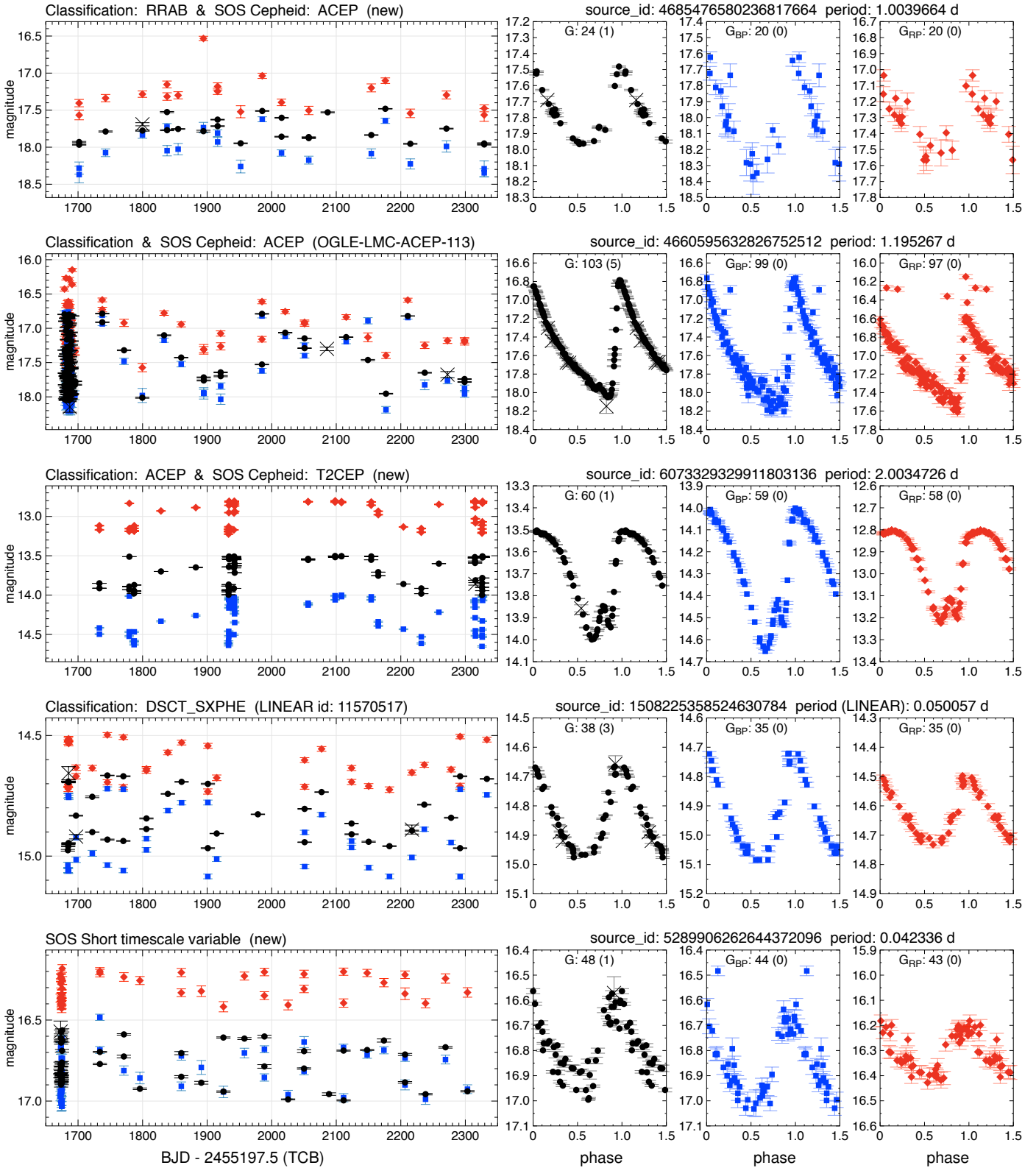


Fig. A.2. Example light curves of the Cepheid, δ Scuti star, and short-timescale variables. For each band we show the valid FoV transits, as well as variability-rejected FoV transits (number in parentheses, data plotted with crosses). In most cases, two FoV transits occur within the 6 h spin period, causing indistinguishable overlapping points in the *left panels*. Some outliers are not flagged in the exported data (see e.g., the G_{BP} and G_{RP} light curve of the ACEP example); these transits were removed in the more strict outlier rejection used in the SOS Cep and RRL module, as mentioned in Sect. 4.1. No period is published for the classification results, therefore the period for the DSCT_SXPHE example is taken from the LINEAR survey (Palaversa et al. 2013).

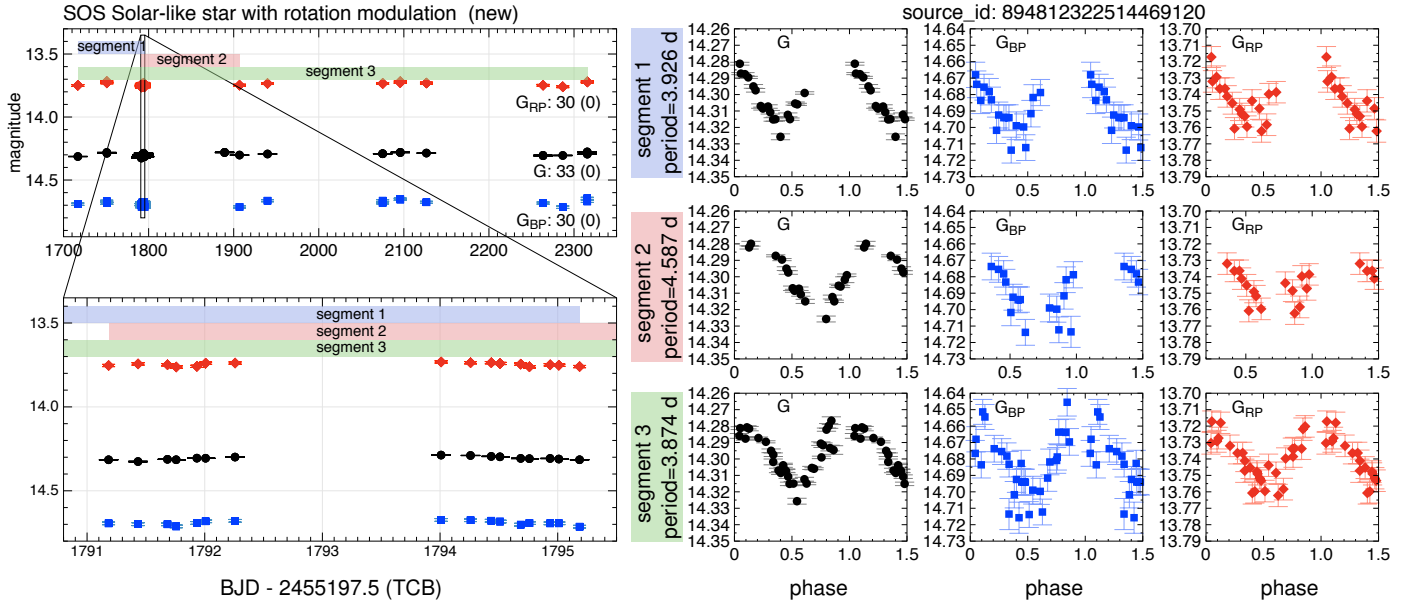


Fig. A.3. Example light curves of a solar-like star with rotation modulation. Period search is performed over multiple segments, as shown in the *right panels*; see Sect. 4.4.4 for more details. The *bottom left panel* expands a high-cadence time range of the time series. The *top left panel* shows for each band the valid FoV transits and variability-rejected FoV transits (number in parentheses, here 0).

Appendix B: Examples of *Gaia* archive queries

This section describes *Gaia* archive queries in the ADQL format that return the various selections presented in this paper. These queries can be made online in the *Gaia* archive¹¹:

B.1. Total counts and table merging

Figure 3 total number of unique sources with variability results (bottom left unique source_id count):

```
select count(*) from gaiadr2.gaia_source where phot_variable_flag='VARIABLE'
```

Retrieve various astrometric and general statistical variability fields for all published variable sources in *Gaia* DR2:

```
select gs.source_id, ra, dec, parallax, parallax_error, pmra, pmdec, epoch_photometry_url,
num_selected_g_fov, mean_mag_g_fov, range_mag_g_fov, std_dev_mag_g_fov, time_duration_g_fov,
num_selected_bp, mean_mag_bp, range_mag_bp, std_dev_mag_bp, time_duration_bp,
num_selected_rp, mean_mag_rp, range_mag_rp, std_dev_mag_rp, time_duration_rp
from gaiadr2.gaia_source as gs join gaiadr2.vari_time_series_statistics as varistats using(source_id)
where phot_variable_flag='VARIABLE'
```

B.2. Creating a variable summary table

It is very desirable to have an overview of the occurrence of a given source_id in any of the vari_* tables. The following query generates such a table in which each row contains a unique variable source source_id, and each column represents one of the vari-tables. A column contains integer 1 when it contains the source_id and NULL when not. After logging into the *Gaia* archive a user table can be created from the query result (suggested name: gaiadr2_vari_summary), after which it appears in the user tables and can be queried like any other DR2 table.

```
select
gs.source_id
,vrrl.incl as vari_rrlyrae -- 1 if RR Lyrae SOS table entry
,vcep.incl as vari_cepheid -- 1 if Cepheid SOS table entry
,vrm.incl as vari_rotation_modulation -- 1 if rotation modulation SOS table entry
,vsts.incl as vari_short_timescale -- 1 if short timescale SOS table entry
,vlpv.incl as vari_long_period_variable -- 1 if long period variable SOS table entry
,vcr_rrl.incl as vari_classifier_result_rrl -- 1 if classif. entry of any RR Lyrae type
,vcr_cep.incl as vari_classifier_result_cep -- 1 if classif. entry of any Cepheid type
,vcr_dscsxp.incl as vari_classifier_result_dsct_sxphe -- 1 if classif. entry of Delta Scuti
,vcr_mirasr.incl as vari_classifier_result_mira_sr -- 1 if classif. entry of Mira or semi-regular type
,1 as vari_time_series_statistics -- 1 if time series statistics entry (1 for all in table)
```

¹¹ <http://gea.esac.esa.int/archive/>

```
,1 as gaia_source_variable_flag -- 1 if gaia_source phot_variable_flag='VARIABLE' (1 for all in table)
from gaiadr2.gaia_source as gs
LEFT OUTER join (select source_id, 1 as incl from gaiadr2.vari_rrlyrae) as vrrl using(source_id)
LEFT OUTER join (select source_id, 1 as incl from gaiadr2.vari_cepheid) as vcep using(source_id)
LEFT OUTER join (select source_id, 1 as incl from gaiadr2.vari_rotation_modulation) as vrm using(source_id)
LEFT OUTER join (select source_id, 1 as incl from gaiadr2.vari_short_timescale) as vsts using(source_id)
LEFT OUTER join (select source_id, 1 as incl from gaiadr2.vari_long_period_variable) as vlpv using(source_id)
LEFT OUTER join (select source_id, 1 as incl from gaiadr2.vari_classifier_result where
    best_class_name='RRAB' or best_class_name='RRC' or best_class_name='RRD'
    or best_class_name='ARRD') as vcr_rrl using(source_id)
LEFT OUTER join (select source_id, 1 as incl from gaiadr2.vari_classifier_result where
    best_class_name='CEP' or best_class_name='ACEP' or best_class_name='T2CEP')
    as vcr_cep using(source_id)
LEFT OUTER join (select source_id, 1 as incl from gaiadr2.vari_classifier_result where
    best_class_name='DSCT_SXPHE') as vcr_dscsxp using(source_id)
LEFT OUTER join (select source_id, 1 as incl from gaiadr2.vari_classifier_result where
    best_class_name='MIRA_SR') as vcr_mirasr using(source_id)
where gs.phot_variable_flag='VARIABLE'
order by gs.source_id
```

B.3. HEALPix grouped data

To reproduce sky-distribution data sets using HEALPix grouping as shown in Figs. 6 and 5, the convenient `gaia_healpix_index(norder, source_id)` function in the archive that extracts the HEALPix pixel id that is encoded in the *Gaia* `source_id` can be used. Using `norder=6` creates pixels of about 0.84 deg^2 . A program like TOPCAT (Taylor 2005) can read and visualise such data set with ease.

Figure 6 top left panel “Classif.: RR Lyrae stars (RRAB, RRC, RRD, ARRD)” number of sources per square degree, and also the min, average, max, and standard deviation of the `vari_classifier_result.best_class_score` value per HEALPix pixel:

```
select
    gaia_healpix_index(6, source_id) AS healpix_6,
    count(*) / 0.83929 as sources_per_sq_deg,
    min(best_class_score) AS min_best_class_score,
    avg(best_class_score) AS avg_best_class_score,
    max(best_class_score) AS max_best_class_score,
    stddev(best_class_score) as stddec_int_average_g
FROM gaiadr2.vari_classifier_result WHERE
    best_class_name='RRAB' or best_class_name='RRC' or best_class_name='RRD' or best_class_name='ARRD'
GROUP BY healpix_6
```

Figure 6 top right panel “SOS: RR Lyrae stars”: number of sources per square degree, and also the min, average, max, and standard deviation of the `vari_rrlyrae.int_average_g` value per HEALPix pixel:

```
select
    gaia_healpix_index(6, source_id) AS healpix_6,
    count(*) / 0.83929 as sources_per_sq_deg,
    min(int_average_g) AS min_int_average_g,
    avg(int_average_g) AS avg_int_average_g,
    max(int_average_g) AS max_int_average_g,
    stddev(int_average_g) as stddec_int_average_g
FROM gaiadr2.vari_rrlyrae
GROUP BY healpix_6
```

Figure 5 all three panels (“number of transits in time series of the G , G_{BP} , and G_{RP} photometric bands”) average, and also the min, max, and standard deviation of the values per HEALPix pixel:

```
select
    gaia_healpix_index(6, source_id) AS healpix_6,
    count(*) / 0.83929 as sources_per_sq_deg,
    -- G-band FoV
    avg(num_selected_g_fov) AS avg_num_selected_g_fov,
    min(num_selected_g_fov) AS min_num_selected_g_fov,
    max(num_selected_g_fov) AS max_num_selected_g_fov,
    stddev(num_selected_bp) as stddec_num_selected_g_fov,
    -- BP-band
    avg(num_selected_bp) AS avg_num_selected_bp,
    min(num_selected_bp) AS min_num_selected_bp,
    max(num_selected_bp) AS max_num_selected_bp,
    stddev(num_selected_bp) as stddec_num_selected_bp,
    -- RP-band
    avg(num_selected_rp) AS avg_num_selected_rp,
    min(num_selected_rp) AS min_num_selected_rp,
```

```

max(num_selected_rp) AS max_num_selected_rp,
stddev(num_selected_rp) as stddec_num_selected_rp
FROM gaiadr2.vari_time_series_statistics
GROUP BY healpix_6

```

B.4. Detailed counting

This section shows how the counts presented in Figs. 3 and 7 can be reproduced from the published DR2 tables. Alternatively, a summary table can first be created as described in Appendix B.2 and the relevant selections on the columns involved (not shown) can be made.

Figures 3 and 7 total number of sources with classification: all / δ Scuti and SX Phoenicis stars / RR Lyrae stars / Cepheids / LPV (Mira and semi-regular):

```

select count(*) from gaiadr2.vari_classifier_result
select count(*) from gaiadr2.vari_classifier_result where best_class_name='DSCT_SXPHE'
select count(*) from gaiadr2.vari_classifier_result where best_class_name='RRAB'
or best_class_name='RRC' or best_class_name='RRD' or best_class_name='ARRD'
select count(*) from gaiadr2.vari_classifier_result where best_class_name='CEP'
or best_class_name='ACEP' or best_class_name='T2CEP'
select count(*) from gaiadr2.vari_classifier_result where best_class_name='MIRA_SR'

```

Figures 3 and 7 total number of sources with SOS: RR Lyrae star / Cepheid / LPV / short timescale / rotation modulation results:

```

select count(*) from gaiadr2.vari_rrlyrae
select count(*) from gaiadr2.vari_cepheid
select count(*) from gaiadr2.vari_long_period_variable
select count(*) from gaiadr2.vari_short_timescale
select count(*) from gaiadr2.vari_rotation_modulation

```

Figure 3 number of sources in common between classification and SOS for RR Lyrae stars / Cepheids / LPV in *Gaia* DR2, these are the numbers in the overlap blocks of the Venn diagrams:

```

select count(*) from gaiadr2.vari_classifier_result as clasr1
INNER join gaiadr2.vari_rrlyrae as sosr1 on clasr1.source_id=sosr1.source_id
where best_class_name='RRAB' or best_class_name='RRC' or best_class_name='RRD'
or best_class_name='ARRD'
select count(*) from gaiadr2.vari_classifier_result as clascep
INNER join gaiadr2.vari_cepheid as soscep on clascep.source_id=soscep.source_id
where best_class_name='CEP' or best_class_name='ACEP' or best_class_name='T2CEP'
select count(*) from gaiadr2.vari_classifier_result as claslpv
INNER join gaiadr2.vari_long_period_variable as soslpv
on claslpv.source_id=soslpv.source_id
where best_class_name='MIRA_SR'

```

Figure 3 number of sources that are only found in classification and not in SOS for RR Lyrae stars / Cepheids / LPV in *Gaia* DR2, these are the numbers in the left blocks of the Venn diagrams:

```

select count(*) from gaiadr2.vari_classifier_result as clasr1
LEFT OUTER join gaiadr2.vari_rrlyrae as sosr1 on clasr1.source_id=sosr1.source_id
where (best_class_name='RRAB' or best_class_name='RRC' or best_class_name='RRD'
or best_class_name='ARRD') and sosr1.source_id is NULL
select count(*) from gaiadr2.vari_classifier_result as clascep
LEFT OUTER join gaiadr2.vari_cepheid as soscep on clascep.source_id=soscep.source_id
where (best_class_name='CEP' or best_class_name='ACEP' or best_class_name='T2CEP')
and soscep.source_id is NULL
select count(*) from gaiadr2.vari_classifier_result as claslpv
LEFT OUTER join gaiadr2.vari_long_period_variable as soslpv
on claslpv.source_id=soslpv.source_id
where best_class_name='MIRA_SR' and soslpv.source_id is NULL

```

Figure 3 number of sources that are only found in SOS and not in classification for RR Lyrae stars / Cepheids / LPV in *Gaia* DR2, these are the numbers in the right blocks of the Venn diagrams:

```

select count(*) from gaiadr2.vari_classifier_result as clasr1
RIGHT OUTER join gaiadr2.vari_rrlyrae as sosr1 on clasr1.source_id=sosr1.source_id
where (best_class_name!='RRAB' and best_class_name!='RRC' and best_class_name!='RRD'
and best_class_name!='ARRD') or clasr1.source_id is NULL
select count(*) from gaiadr2.vari_classifier_result as clascep
RIGHT OUTER join gaiadr2.vari_cepheid as soscep on clascep.source_id=soscep.source_id
where (best_class_name!='CEP' and best_class_name!='ACEP' and best_class_name!='T2CEP')
or clascep.source_id is NULL
select count(*) from gaiadr2.vari_classifier_result as claslpv
RIGHT OUTER join gaiadr2.vari_long_period_variable as soslpv
on claslpv.source_id=soslpv.source_id
where best_class_name!='MIRA_SR' or claslpv.source_id is NULL

```