A&A 674, A20 (2023)
https://doi.org/10.1051/0004-6361/202244178

Astronomy & Astrophysics
Special issue

Gaia Data Release 3

Gaia Data Release 3

Rotational modulation and patterns of colour variation in solar-like variables

E. Distefano[1], A. C. Lanzafame[1,2], E. Brugaletta[1], B. Holl[3], A. F. Lanza[1], S. Messina[1], I. Pagano[1], M. Audard[3], G. Jevardat de Fombelle[3], I. Lecoeur-Taibi[3], N. Mowlavi[3], K. Nienartowicz[3,5], L. Rimoldini[3], D. W. Evans[4], M. Riello[4], P. García-Lario[6], P. Gavras[6], and L. Eyer[3]

[1] INAF – Osservatorio Astrofisico di Catania, Via S. Sofia 78, 95123 Catania, Italy
e-mail: elisa.distefano@inaf.it
[2] University of Catania, Astrophysics Section, Dept. of Physics and Astronomy, Via S. Sofia 78, 95123 Catania, Italy
[3] Department of Astronomy, University of Geneva, Chemin Pegasi 51, 1290 Versoix, Switzerland
[4] Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
[5] Sednai Sàrl, Geneva, Switzerland
[6] RHEA for European Space Agency (ESA), Camino bajo del Castillo s/n, Urbanizacion Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain

Received 3 June 2022 / Accepted 5 November 2022

ABSTRACT

Context. The *Gaia* third Data Release (GDR3) presents a catalogue of 474 026 stars (detected by processing a sample of about 30 million late-type stars) with variability induced by magnetic activity. About 430 000 of these stars are newly discovered variables. For each star, the catalogue provides a list of about 70 parameters among which the most important are the stellar rotation period P, the photometric amplitude A of the rotational modulation signal, and the Pearson correlation coefficient r_0 between magnitude and colour variation.

Aims. In the present paper we highlight some features of the *Gaia* photometric time series used to obtain the catalogue and we present the main attributes of the catalogue.

Methods. The Specific Objects Study (SOS) pipeline, developed to characterise magnetically active stars with *Gaia* data, was described in the paper accompanying the *Gaia* second rata release (DR2). Here we describe the changes made to the pipeline and a new method developed to analyse *Gaia* time series and to reveal spurious signals induced by instrumental effects or by the peculiar nature of the investigated stellar source. Such a method is based on the measurement of the per-transit-corrected-excess-factor ($c*$) for each time-series transit, where $c*$ is a parameter that allows us to check the consistency between G, G_{BP}, and G_{RP} fluxes in a given transit.

Results. The period–amplitude diagram obtained with the DR3 data confirms the DR2 findings that is, the existence of a family of low-amplitude fast rotators never seen by previous surveys. The GDR3 data permit, for the first time, the analysis of patterns in magnitude–colour variation for thousands of magnetically active stars. The measured r_0 values are tightly correlated with the star positions in the period–amplitude diagram.

Conclusions. The relationship between the P, A, and r_0 parameters inferred for thousands of stars are potentially very useful for improving our understanding of stellar magnetic fields and ameliorating theoretical models, especially in the fast rotation regime. The method developed to reveal the spurious signals can be applied to each of the released *Gaia* photometric time series and can be exploited by anyone interested in working directly with *Gaia* time series.

Key words. stars: activity – starspots – stars: solar-type – stars: rotation – surveys – techniques: photometric

1. Introduction

The expression solar-like variables is commonly used to designate a wide class of objects (including late-type dwarfs, T-Tauri, and stars in the RS CVn binary systems) whose variability is induced by the presence and evolution of magnetically active regions (MARs). These regions are complexes of dark spots and bright faculae located in regions with an enhanced magnetic field and are unevenly distributed over the stellar photosphere. In solar-like stars, the evolution of MARs is responsible for different variability phenomena whose timescales can range between a few minutes, as in the case of flare events, to years, as in the case of the 11 yr cycle observed in the Sun (see e.g., Messina et al. 2004). Flares are outbursts induced by the magnetic reconnection, that is, a rearrangement of the magnetic field topology occurring over a short timescale (see e.g., Shibata 1999; Benz & Güdel 2010, for a theoretical view of this mechanism). A typical flare event exhibits a rapid (few minutes) increase in the stellar flux followed by a slower decay phase whose duration can range between 20 min and 6 h (see e.g., Walkowicz et al. 2011; Davenport 2016; Doyle et al. 2019). At the other end of the scale, periodic or quasi-periodic variations in stellar flux can be induced by variations in the total area and geometrical distribution of MARs. Usually, these variations take place on timescales of several years and are referred to as activity cycles. On an intermediate timescale, the typical signature of the light curves of solar-like stars is the so-called rotational modulation signal, which is a quasi-periodic flux variation induced by the stellar rotation that modulates the visibility of MARs over the stellar disk. Analysis of the rotational modulation signal is

a powerful tool for studying the physical properties of a given star and for constraining theoretical models on stellar evolution and magnetic fields. Firstly, estimating the period of the signal is equivalent to measuring the stellar rotation period P. Secondly, the amplitude A of the signal can be used to constrain the hemispheric asymmetry of MAR distribution, the MAR filling factor, and the contrast between MARs and photosphere, which are the main factors driving the rotational modulation variability of the stellar flux. Thirdly, the temporal variations in the signal amplitude can be used to infer the timescales over which the stellar magnetic field changes topology. Finally, the estimate of P in a large sample of stars with known ages allows us to investigate how the stellar rotation period depends on the stellar age and to probe and improve the gyrochronology theories developed to model such a dependency (Skumanich 1972; Barnes 2003; Lanzafame & Spada 2015).

In addition to the period P and the amplitude A of the rotational modulation signal, another important diagnostic that can provide precious information on the stellar magnetic field is given by the Pearson correlation coefficient r_0 between magnitude and colour variations. Indeed, this index can provide information on the contrast in temperature between the MARs and the stellar photosphere and on their geometrical distribution (see e.g., Messina 2008; Iwanek et al. 2019). The multiband nature of the *Gaia* survey allowed us to measure r_0 and to study its correlation with P and A for an unprecedented number of stars.

In the present paper, we illustrate the results obtained by searching for rotational modulation signals in the photometric time series of the third *Gaia* data release (GDR3). The analysis of GDR3 data allowed us to detect rotational modulation in a sample of 474 026 stars, among which about 430 000 are, as far as we know, newly discovered variables. The results of our analysis are stored in the GDR3 `vari_rotational_modulation` catalogue (hereafter `gdr3_rotmod`), where we report, for each star, a list of 66 parameters characterising their rotation and magnetic activity. A full description of these parameters can be found in the GDR3 documentation (Rimoldini et al. 2022). Here, we focus on the P, A, and r_0 distributions and on their reciprocal relationships.

In the second section of the paper, we describe the pipeline used to detect and characterise the stellar magnetic activity. In Sect. 3, we present the main features of the catalogue and in Sect. 4 we draw conclusions.

In the three appendices, we analyse how the technical features of *Gaia*, such as the time-sampling and the processing of photometric data, affect the completeness of the `gdr3_rotmod` catalogue. These appendices are recommended to readers that are interested in gaining a deeper understanding of *Gaia* data and in working with the released photometric time series. In Appendix A, we provide a detailed illustration of a method developed to detect spurious signals in the *Gaia* photometric time series. In Appendix B, we assess the completeness and the contamination of the `gdr3_rotmod` catalogue. Finally, in Appendix C, we present a comparison between the results for rotational modulation variables obtained with the second *Gaia* data release (GDR2) and GDR3. This analysis shows that, though DR3 provides a richer sample of rotational modulation variables, about 60% of the variables detected in GDR2 could not be recovered in GDR3. The lack of detection of rotational modulation for these sources in GDR3 is due to several issues related to changes in the variability pipeline and in the strategy adopted to calibrate the photometric data. However, this does not imply that the rotational modulation data published in GDR2 are invalid.

2. The method

The pipeline used to search for the rotational modulation signal in the GDR3 time series is a new version of that used for GDR2 and described in Lanzafame et al. (2018). The main differences between the two versions of the code are that the quality criteria used to select the final sample of good candidates have been tightened and that the amplitude of the rotational modulation signal is now computed in all three *Gaia* photometric bands. The data processing can be outlined in six main steps:

- source selection;
- time-series segmentation;
- data cleaning;
- analysis of brightness–colour correlation;
- search for the rotational modulation signal;
- estimate of the stellar rotation period.

The pipeline was developed inside the framework of the Coordination Unit 7 (CU7) of the *Gaia* DPAC (Data Processing and Analysis Consortium). The CU7 softwares are designed to analyse the photometric time series collected by *Gaia* in order to classify and characterise the variability phenomena exhibited by the targeted stars. The raw time series collected by *Gaia* are affected by poor-quality measurements related to instrumental or calibration issues like those described in Riello et al. (2021). These 'bad' measurements can prevent the detection or can affect the characterisation of variability phenomena. In order to mitigate the effects of these measurements, the CU7 team designed a chain of operators to clean the raw time series and to reject all the outliers that are due to instrumental effects. The final output of this chain is given by the `ExtremeErrorCleaningMagnitudeDependent` time series collected in the *Gaia* pass-bands G, $G_{\rm BP}$, and $G_{\rm RP}$ (see Riello et al. (2021) and references therein for details on the *Gaia* photometric system). All the CU7 packages use these cleaned time series as input. These time series span an interval of 34 months. A full description of the chain of operators can be found in Eyer et al. (2023).

2.1. Data selection

The sources analysed in the present work are extracted from the `geq5` catalogue described in Holl et al. (2023), which is the catalogue consisting of all the sources for which at least five non-null and positive field-of-view (FoV) flux observations were available. The `geq5` catalogue lists about 1800 million sources. In order to save computing time, we ran our pipeline only on a subset of about 30 million sources. This subset was obtained by adopting the selection procedure described in Lanzafame et al. (2018) and Rimoldini et al. (2022). Briefly, a star is selected if it fulfills the following criteria:

1. it has a positive parallax;
2. the relative error on parallax is less than 20%;
3. its location in the M_G vs. $G_{\rm BP}-G_{\rm RP}$ falls in the region delimited by the black lines[1] (Fig. 1);
4. the sampling of its G photometric time series allows the extraction of at least one subseries satisfying the condition

$$L \leq 120\,\mathrm{d} \text{ and } N_{\rm P} \geq 12, \tag{1}$$

where L is the time interval spanned by the subseries and $N_{\rm P}$ is the number of photometric measurements inside the subseries.

[1] The equations defining the region used for candidate selection are provided in Rimoldini et al. (2022).

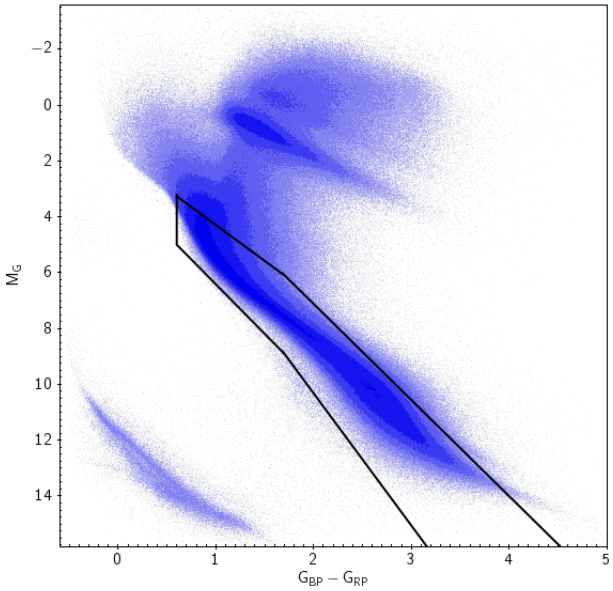

Fig. 1. DR3 G vs. $G_{BP} - G_{RP}$ diagram. The black continuous lines mark the region used to perform a first selection of solar-like variables.

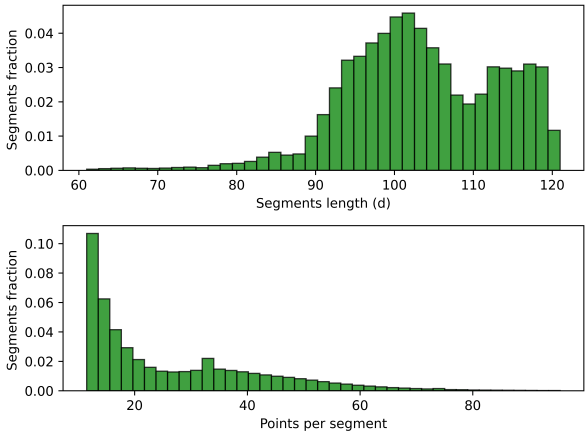

Fig. 2. Distributions of the segment properties. Top panel: distribution of the segment length L. Most of the processed segments have a length $L = 100$ d. Bottom panel: distribution of the number of points per segment N_P.

The first, second, and third criteria are applied in order to limit the analysis to stars whose location in the M_G vs. $G_{BP} - G_{RP}$ is consistent with dwarfs of spectral types later than F5 or with T-Tauri objects. This preliminary filtering is very rough and the region used for selection is expected to include several types of contaminants. Most of these contaminants will be removed by the subsequent steps of the pipeline. The final percentage of contaminants in the released sample of variables is evaluated in Appendix B.2. The fourth criterion is used to select the stars whose photometric time series have a sampling that is suitable for the detection of the rotational modulation signal, as discussed in the following section.

2.2. Segmentation

The amplitude and the shape of the rotational modulation signal change over time because of the intrinsic evolution of MARs. The timescale over which the signal is stable and coherent is closely related to the typical lifetime of MARs τ_{MAR}. Ideally, the search for the rotational modulation signal should be performed in subseries whose length does not exceed τ_{MAR}.

The estimate of this timescale has been the subject of several works and is related to different parameters, such as stellar age and spectral type. The analysis of the total solar irradiance (TSI) and solar spectral irradiance (SSI) time series showed that the typical lifetime of the spots on the Sun τ_{spot} is about 9 days, whereas the typical lifetime of faculae τ_{fac} is about 60 days, which roughly corresponds to two solar rotations (Lanza et al. 2004). In the same work, the authors show that, in some cases, the rotational modulation signal in the Sun can be coherent up to 150 d because spots and faculae tend to occur at preferential longitudes giving rise to ensembles of MARs whose typical lifetime τ_{MAR} is about 200–250 d. However, the authors point out that the rotational modulation signal of the Sun can be detected through period search algorithms only in 150 d intervals close to the minima of the 11 yr activity cycle. In these intervals, the Sun activity is dominated by faculae and therefore the rotational modulation signal is less affected by the luminosity dips induced by spots. These findings were recently confirmed by Shapiro et al. (2017). The MARs lifetime for cool stars other than the Sun has been estimated for a limited sample of stars by Donahue et al. (1997a,b), Hussain (2002), and Messina & Guinan (2003). Donahue et al. (1997a,b) found that in old stars characterised by a low magnetic activity level, τ_{MAR} tends to be shorter than the stellar rotation period, preventing the detection of the rotational modulation signals, whereas in young and magnetically active stars, τ_{MAR} values are of the order of several stellar rotations and range between 50 d and 1 year. Similar values are also reported by Hussain (2002) and by Messina & Guinan (2003). On the basis of these works, Distefano et al. (2016) demonstrated that segmenting long-term photometric time series in 50 d intervals can be a good strategy to correctly retrieve the rotational modulation signal in magnetically active stars. Unfortunately, because of the *Gaia* time-series sampling, a 50 d interval may not have enough observations to perform a meaningful period search. A compromise between the subseries length and the number of points per subseries must therefore be found. Moreover, the sampling changes with the star coordinates according to the *Gaia* scanning law (see Eyer & Mignard 2005; Distefano et al. 2012, for further details). In order to take into account the features of the *Gaia* sampling and the intrinsic evolution of MARs, the G, G_{BP}, and G_{RP} photometric time series of the selected target stars are divided into subseries satisfying condition 4 reported in Sect. 2.1, through the adaptive segmentation algorithm described in Lanzafame et al. (2018). This algorithm is adaptive because the segmentation strategy changes with the sky regions in which the analysed stars are located.

The top and bottom panels of Fig. 2 report the distribution of the subseries length and of the number of points per subseries, respectively. Most of the segments are characterised by a 100 d length and include 12 points, but there is also a small fraction of segments shorter than 20 d and a small fraction of segments with more than 100 points.

The distribution of N_P and L changes with the sky location of the targets stars. In Figs. 3 and 4, we report the median values of N_P and L as a function of the ecliptic coordinates. The stars located at ecliptic latitudes close to $\beta \pm 45°$ are scanned more frequently by the satellite and the corresponding subseries have a higher number of points N_P.

In Fig. 5, we illustrate a typical case in which the time-series segmentation is crucial in retrieving the stellar rotation

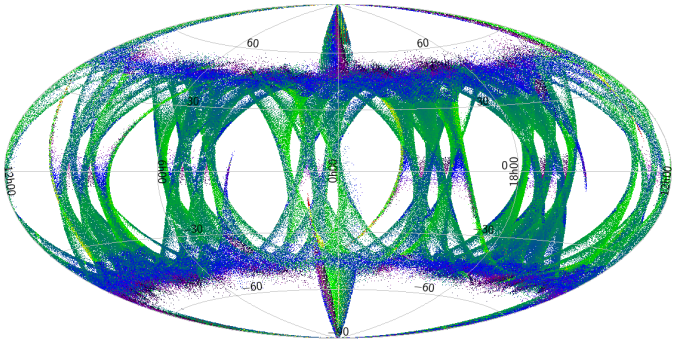

Fig. 3. Distribution map (in ecliptic coordinates) of the average segment length $\langle L \rangle$.

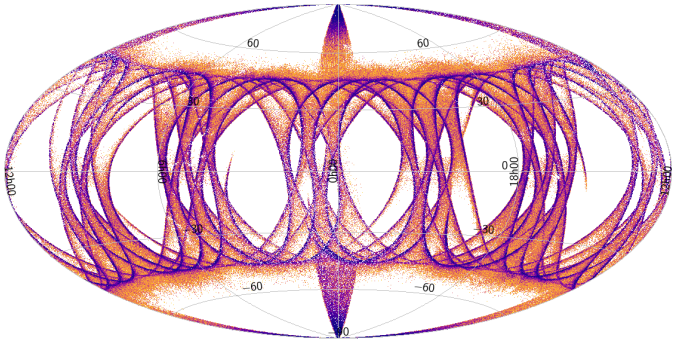

Fig. 4. Distribution map (in ecliptic coordinates) of the average number of points per segment $\langle N_P \rangle$.

period. In the top-left panel of the plot, we display the whole G time series of the source `Gaia DR3 6503897945888972544`. The grey-shaded regions indicate two of the subseries generated by the segmentation algorithm (we note that the other segments extracted by the pipeline are not marked in order to make the plot clearer). In both segments, a signal with period $P = 0.867$ d was detected. In the top-right and bottom-left panels, we report the folded G subseries corresponding to the first and second segment, respectively. In the bottom-right panel, we report the whole time series folded according to the same period $P = 0.867$ d. While in the two subseries the modulation signal is coherent and stable, in the whole time series it loses coherence because MARs evolution induces a shift in phase between the two segments. Although the signal cannot be detected in the whole time series, the segmentation algorithm allows us to retrieve the stellar rotation period. A similar case is reported in Fig. 6. We note that MARs evolution not always destroys the signal coherence. In Fig. 7 we illustrate the case of the star `Gaia DR3 5059101630762735360`. In this star, the amplitude of the signal increases in time and changes from 0.07 mag in the first segment to 0.13 mag in the second one. Nevertheless, the signal is still detectable throughout the time series. Finally, there are stars for which the signal is stable and coherent throughout the time series, as in `Gaia DR3 2925077104599544960` (see Fig. 8).

We highlight the fact that, recently, Basri et al. (2022) conducted an extensive study based on a sample of about 60 000 stars for which both *Kepler* photometric time series and *Gaia* stellar parameters were available. Basri et al. (2022) found that the $\tau_{\rm spot}$ distribution ranges between 10 and 350 d and is peaked at around 140 d for young stars rotating faster than the Sun, whereas it ranges between 10 and 250 d and is peaked at 80 d for stars older than or as old as the Sun. Unfortunately, this work was not available at the time when DR3 data were processed (i.e., December 2020). In the future *Gaia* releases, it could be very helpful to redefine the segmentation strategy.

2.3. Data cleaning

Once the time series have been segmented, each subseries is cleaned from spurious data. A first cleaning procedure is performed by rejecting all transits for which one or two of the G, $G_{\rm BP}$, and $G_{\rm RP}$ measurements are missing. This cleaning is performed in order to ensure that the magnetic activity indexes computed in the different photometric bands are inferred from the same set of transits. We note that, in this work, the term transit indicates the crossing of the entire focal plane by a given source. During a given transit *Gaia* collects a set of four measurements with the different instruments on board the satellite (see Gaia Collaboration 2016, for further details on *Gaia* instrumentation). Once transits with missing measurements are removed, the pipeline searches for outliers due to possible flare events according to the procedure described in Lanzafame et al. (2018) and in Sect. 10.14.3 of Rimoldini et al. (2022). These outliers are flagged as candidate flare events and are removed from the subseries because they could prevent the detection of the rotational modulation signal.

2.4. Analysis of the correlation between brightness and colour variations

The cleaned time series are analysed to investigate the correlation between the brightness and colour variations in the selected sample of stars. For each star and for each segment, the pipeline performs a robust linear regression between the G and $G_{\rm BP}-G_{\rm RP}$ measurements collected in the segment and estimates the slope s and the intercept i of the straight line best-fitting the data (see Lanzafame et al. 2018, for details on the algorithm employed to compute the regression). For each segment, the Pearson correlation coefficient $r_0(G, (G_{\rm BP} - G_{\rm RP}))$ and its associated p value are also computed. The closer $|r_0|$ is to 1, the higher the strength of the linear correlation between the colour and the magnitude variation. The p value gives the probability that the measured r_0 is obtained by chance. Hence, the lower p is, the higher the

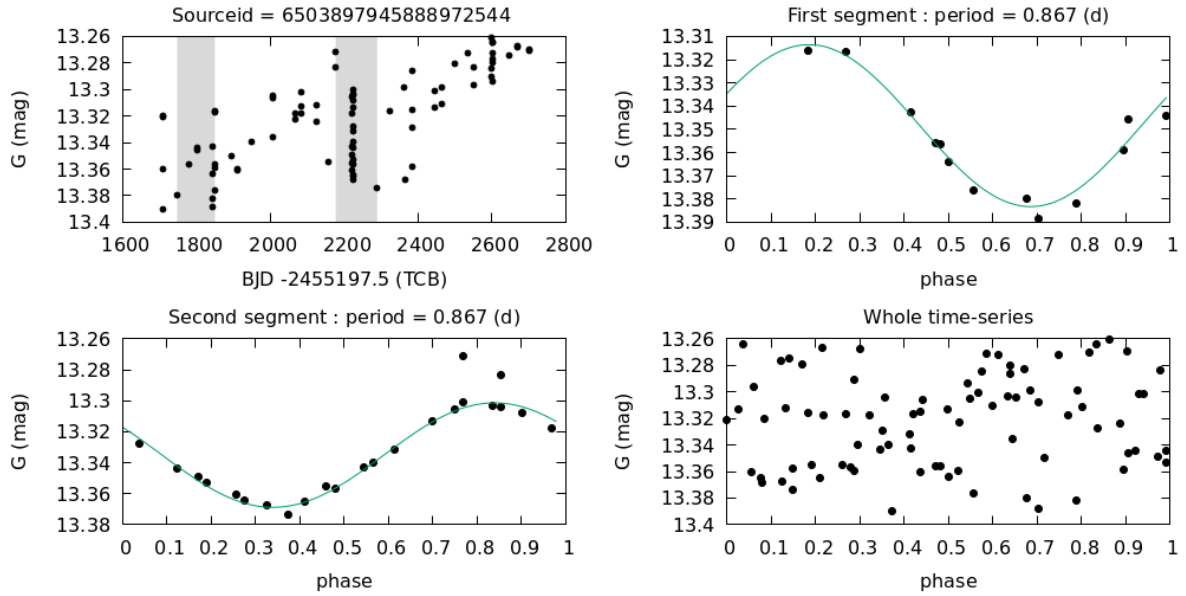

Fig. 5. Time series segmentation for the star `Gaia DR3 6503897945888972544`. Top left panel: full G time series. The grey shadows mark two of the segments extracted by the segmentation algorithm. The period search algorithm retrieved a period $P = 0.867$ d in both segments. Top right panel: first subseries folded according to the period $P = 0.867$. Bottom left panel: second subseries folded according to $P = 0.867$ d. Bottom right panel: full time series folded according to $P = 0.867$ d. The rotational modulation signal loses coherence across the full time series because of the intrinsic evolution of MARs and can be detected only in the shorter subseries.

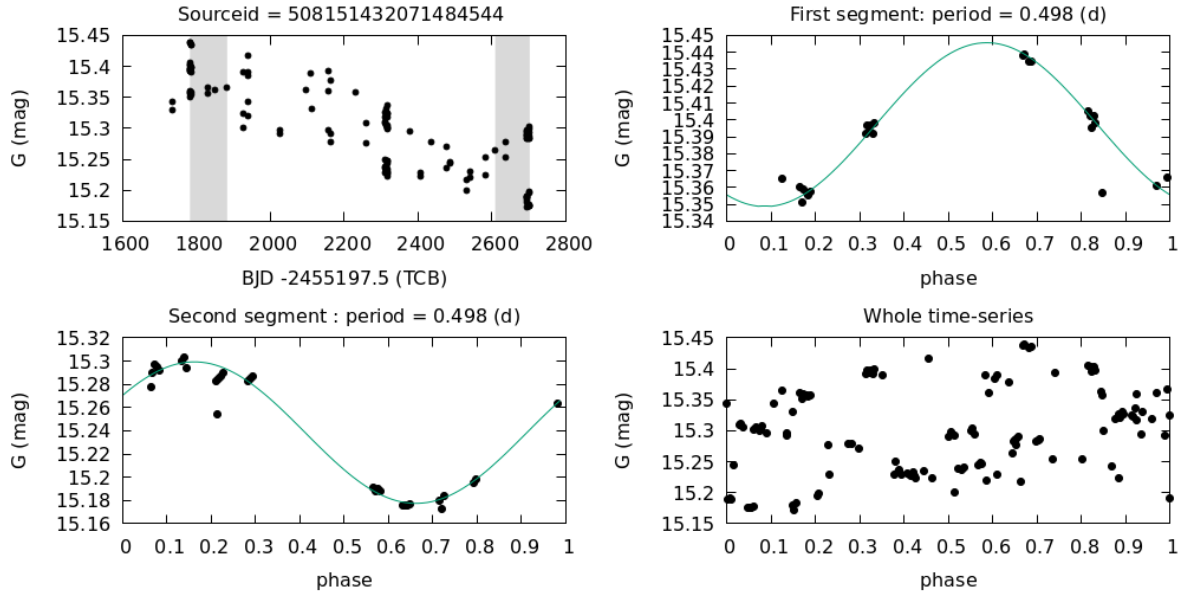

Fig. 6. Same as Fig. 5 but for the star `Gaia DR3 508151432071484544`. In this case, the rotational modulation signal again loses coherence across the full time series.

statistical significance of the correlation. For a given star, the s, i, r_0, and p parameters are stored in the following respective arrays of the `gdr3_rotmod` :

- `segments_color_mag_slope`,
- `segments_color_mag_intercept`,
- `segments_correlation_coefficient`,
- `segments_correlation_significance`.

We note that these parameters were estimated and reported also for the segments in which a meaningful rotational modulation signal is not detected. The analysis of these parameters shows that there are roughly three distinct families of stars that, according to the definition given by Messina (2008), we call:

- `reddening-color-magnitude-correlated` (hereafter RCMC) stars;
- `blueing-color-magnitude-correlated` (hereafter BCMC) stars;
- `color-magnitude-uncorrelated` (hereafter UMC stars).

In RCMC stars, the colour and the magnitude variation are positively correlated; that is, as their brightness decreases their colour gets redder. In BCMC stars, the colour and the magnitude variations are anti-correlated; that is, as they get fainter they become bluer. Finally, in UMC stars, the colour and magnitude variations are poorly correlated. The transition between the three families is very smooth. In some RCMC stars, the positive correlation is well defined in all segments. In other RCMC stars, a

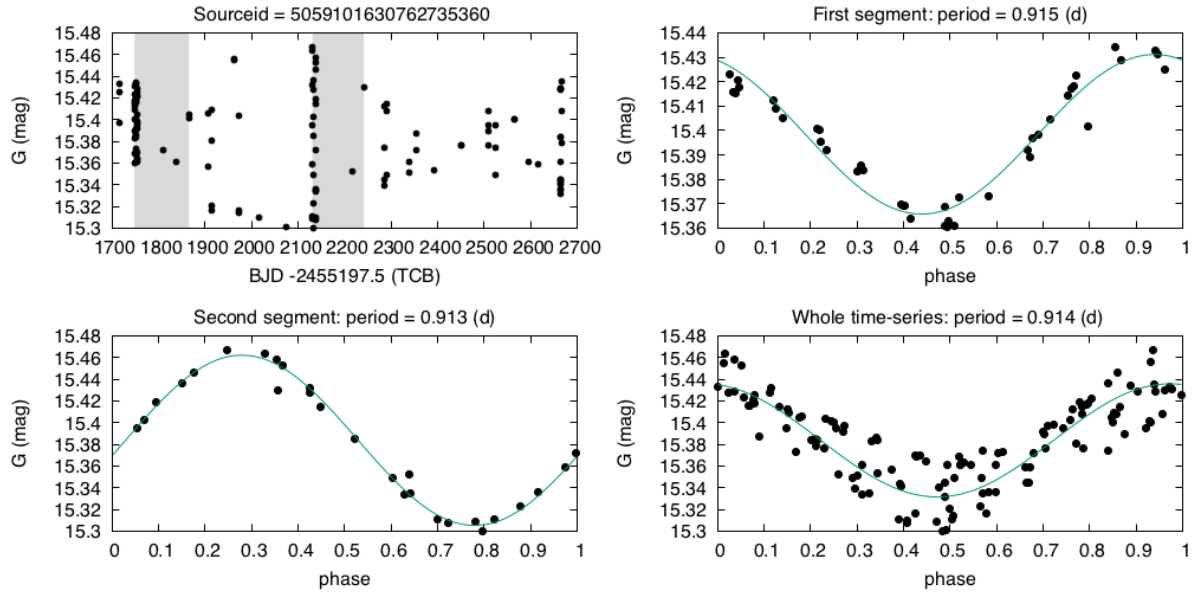

Fig. 7. Same as Fig. 5 but for the stars `Gaia DR3 5059101630762735360`. In this case, the rotational modulation signal is coherent across the full time series, indicating that the MARs are characterised by a long-term stable pattern.

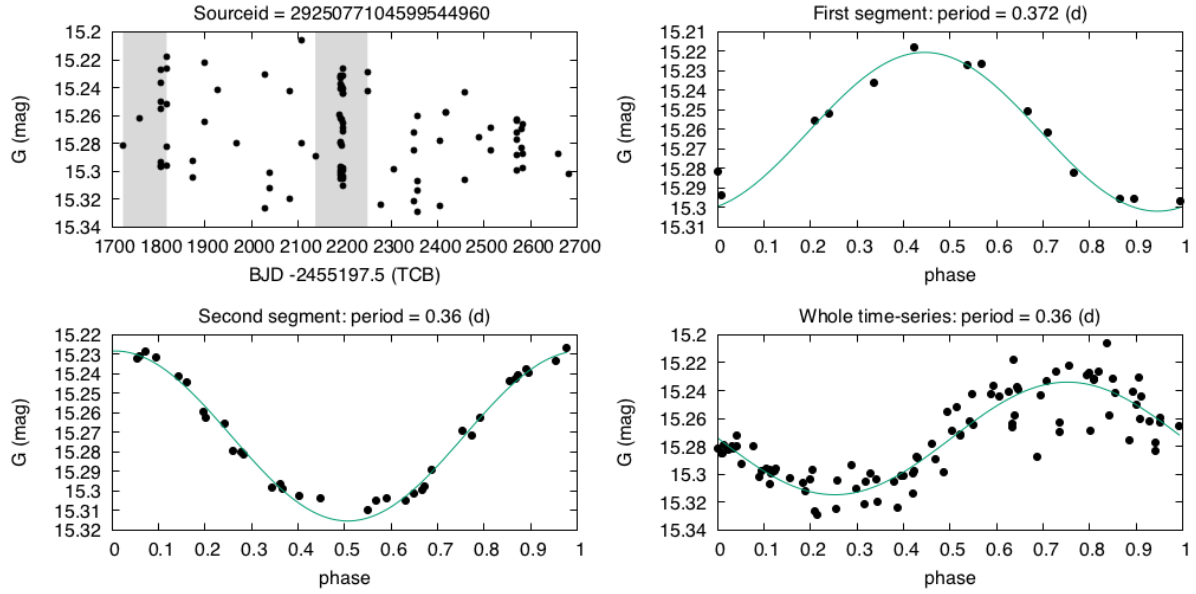

Fig. 8. Same as Fig. 5 but for the star `Gaia DR3 2925077104599544960`. In this case, the rotational modulation signal is stable across the full time series.

significant correlation is observed only in some segments. The same applies to BCMC stars. Finally, there are stars that in some segments are reddening and in others are blueing.

Figure 9 presents the normalised distribution of the median Pearson correlation coefficient MED($r_0(G, (G_{BP} - G_{RP}))$) in a different range of magnitudes: as the stars get fainter, the percentage of uncorrelated stars increases because the photometric noise can mask the correlation between colour and magnitude variation. We note that these distributions are built without taking into account the statistical significance associated with MED($r_0(G, (G_{BP} - G_{RP}))$), otherwise the percentage of uncorrelated stars would be underestimated.

Figure 10 reports the distribution of MED($r_0(G, (G_{BP} - G_{RP}))$) for all the stars for which the associated p value is less than 0.1. The histogram shows that the blueing stars with a high statistical significance are only a small percentage ($\simeq$4%) of the entire `gdr3_rotmod` catalogue, whereas the reddening stars are more numerous and are about 17% of the catalogue.

Figures 11–14 report some examples of the different cases.

Figure 11 illustrates an example of an RCMC star that is given by the source `Gaia DR3 3321251347610387968`. The first three panels display the G vs. $(G_{BP} - G_{RP})$ diagram for three distinct segments, whereas the bottom-right panel reports the same diagram for the whole time series. In this example, the colour and brightness variations are strongly and significantly correlated ($r_0 > 0.9$ and $p < 0.05$) in each segment and in the whole time series.

In other RCMC stars, the positive correlation is seen only in some segments and is lost in the whole time series because of the intrinsic evolution of spots and faculae. An example of this kind

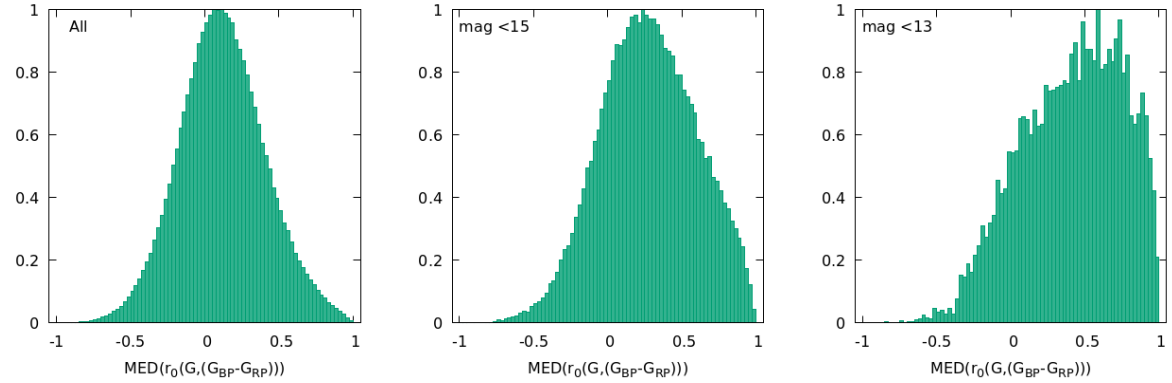

Fig. 9. Normalised distributions of MED($r_0(G, (G_{BP}) - G_{RP})$)). Left panel: distribution for all the stars of the `gdr3_rotmod` catalogue. Central panel: distribution for stars with $G < 15$. Right panel: distribution for stars with $G < 13$.

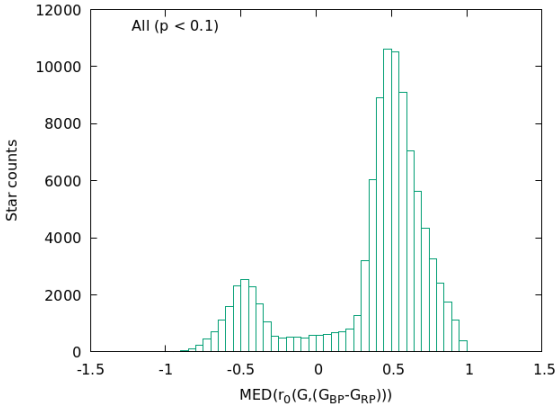

Fig. 10. Distribution of the median Pearson correlation coefficient MED($r_0(G, (G_{BP} - G_{RP}))$) for all the targets in which $p < 0.1$.

of star is given in Fig. 12 where we plot the G vs. $(G_{BP} - G_{RP})$ diagram for the source `Gaia DR3 6629372460509004928`. In this source, a significant correlation is observed only in the first two time-series segments. Such a correlation disappears in the third segment ($r_0 = 0.13$) and in the whole time series ($r_0 = -0.04$).

In Fig. 13, we illustrate an example of a BCMC star that is given by the source `Gaia DR3 5020821858561059456`. In this star, a significant anti-correlation is seen in all the three displayed segments ($r_0 \leq -0.9$ and $p < 0.05$). The anti-correlation is visible also in the whole time series but is slightly attenuated ($r_0 \leq -0.62$ and $p < 0.05$).

Finally, Fig. 14 illustrates an example of a UMC star that is given by the source `Gaia DR3 5823916073592266624`. The values of the r_0, p, s, and i parameters are complex functions of different factors, such as the stellar effective temperature $T_{\rm eff}$, the contrast in temperature $\Delta T_{\rm spot}$ between the spots and the stellar photosphere, the contrast in temperature $\Delta T_{\rm fac}$ between the faculae and the stellar photosphere, the geometrical distribution of spots and faculae, and the photometric noise of the G, G_{BP}, and G_{RP} time series. A full understanding of the r_0, p, s, and i distribution seen in our sample of stars would require theoretical modelling that is beyond the scope of this work. However, in the final section of the paper, we show how r_0, p, s, and i correlate with the other parameters listed in the `gdr3_rotmod` catalogue, we discuss how these relationships can shed light on the difference between RCMC and BCMC stars, and we compare our results with previous studies.

2.5. Search for the rotational modulation signal

The cleaned time-series segments are processed by means of the Lomb-Scargle algorithm as implemented by Zechmeister & Kürster (2009). For each segment, the period P corresponding to the highest peak of the periodgram is selected and its false alarm probability (FAP) is evaluated according to the Baluev (2008) prescriptions. If the FAP value associated with P satisfies the condition

$$\mathrm{FAP} \leq 0.05, \tag{2}$$

then the period is flagged as valid and a sinusoidal model is fitted to the photometric data:

$$m(t) = a_{\rm m} + b_{\rm m} \cos\left(\frac{2\pi t}{P}\right) + c_{\rm m} \sin\left(\frac{2\pi t}{P}\right), \tag{3}$$

where $m \in (G, G_{BP}, G_{RP})$. The FAP threshold given in Eq. (2) is designed to maximise the ratio between true and false positives and is based on the analysis of Süveges et al. (2015). We note that while the fitting procedure is applied to all three sets of photometric data, the period search is performed only on the G subseries because they have a higher signal-to-noise ratio (S/N). After the fitting procedure, the residuals ϵ_i with respect to the fitted sinusoidal signals are computed. The transits with the higher residuals are flagged as candidate flare events. A given transit is flagged as an outlier if it satisfies the condition:

$$|\epsilon_i| \geq \langle|\epsilon|\rangle + 3\sigma_{|\epsilon|}. \tag{4}$$

The amplitude of the rotational modulation signal is related to the non-axisymmetric part of the spot distribution and is often used as an index of the stellar magnetic activity level (see e.g., Rodonò et al. 2000; Ferreira Lopes et al. 2015; Reinhold et al. 2013; Lehtinen et al. 2016). However, the data coming from the *Kepler* mission showed that this index has to be treated with caution (see e.g., Reinhold et al. 2013; McQuillan et al. 2014; Basri & Nguyen 2018). We discuss the use of this index in the following section, where the relation between the amplitude of the rotational modulation and the stellar rotation period is investigated.

The pipeline makes two different estimates of this index and provides a percentile-based index[2]:

$$A_{\rm perc}(m) = m_{95\rm th} - m_{5\rm th}, \tag{5}$$

[2] This index is very similar to the variability range defined by Basri et al. (2010, 2011) for the *Kepler* data. The only difference is that the *Kepler* index is computed in flux units instead of mag units.

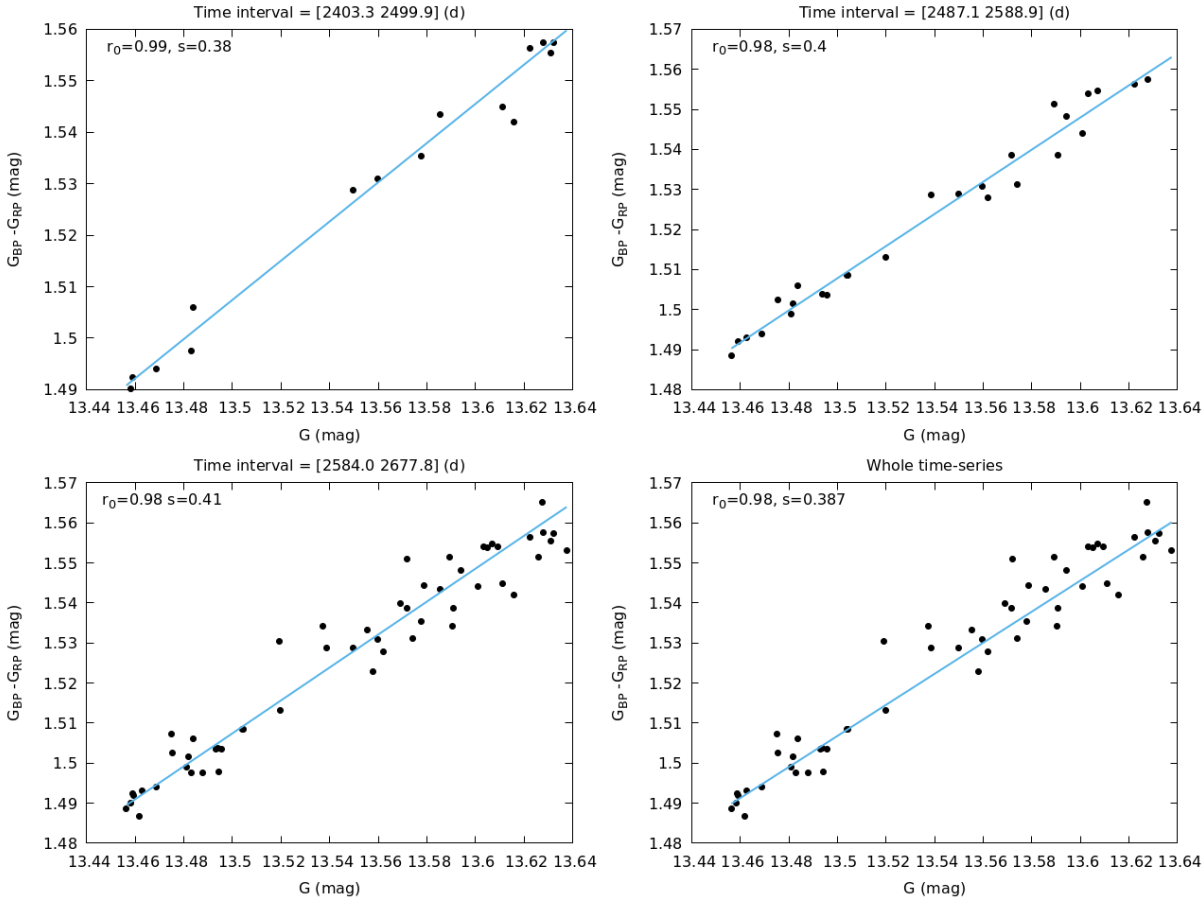

Fig. 11. G vs. $(G_{BP} - G_{RP})$ diagram for the star `Gaia DR3 3321251347610387968`. In the top-left, top-right and bottom-left panels the diagram is displayed for three distinct time series segments, respectively. In the bottom-right the diagram is displayed for the full time series.

and a fit-based index:

$$A_{\rm fit}(m) = 2\sqrt{b_{\rm m}^2 + c_{\rm m}^2}, \tag{6}$$

where $m_{95\rm th}$ and $m_{5\rm th}$ are the 95th and 5th percentile of the magnitude distribution in a given segment, respectively, and $b_{\rm m}$ and $c_{\rm m}$ are the coefficients of Eq. (3) and $m \in (G, G_{\rm BP}, G_{\rm RP})$. Therefore, overall, the pipeline computes six activity indexes for each segment in which a significant period P (FAP $\leq$ 0.05) is detected. The amplitude inferred from the fit can be useful when dealing with faint sources. In fact, in faint sources the activity index computed with Eq. (5) can be overestimated because of the increased photometric noise. The activity indexes in the $G_{\rm BP}$ and $G_{\rm RP}$ bands can be useful for constraining MAR temperatures. Throughout the following analysis, we make use of only the $A_{\rm per}(G)$ index, for brevity. Moreover, we mainly focus our analysis on stars with G magnitude in the range (13,15.5), which is the interval where the *Gaia* photometric sensitivity reaches its maximum, as illustrated in Fig. 15. In this picture, we report the median activity index MED($A_{\rm perc}(G)$) vs. the G magnitude, for each star. The lowest amplitudes detected by the pipeline are of about 0.003 mag and fall in the G range (13,15.5). The best estimate of the stellar rotation period (supplied as `best_rotation_period` in the `gdr3_rotmod` catalogue) is computed by taking the mode of the significant period distribution (see Lanzafame et al. 2018; Rimoldini et al. 2022, for details on the algorithm used to compute the mode).

2.6. Spurious-period filtering

As discussed in Lanzafame et al. (2018), the non-uniformity of the *Gaia* sampling could lead to the detection of spurious periods. In these cases, visual inspection of folded light curves reveals that the data are poorly sampled and not well reproduced by the fitted sinusoidal curve. In order to reject these spurious periods, the FAP parameter needs to be complemented with other quality indicators. We defined four quality assurance parameters, namely the phase coverage (PC), the maximum phase gap (MPG), the ratio Q between the $A_{\rm fit}(G)$ and the $A_{\rm perc}(G)$ indexes, and the reduced chi-squared $\tilde{\chi}^2(G)$ associated with the fitting procedure described in Sect. 2.5.

Each time-series segment in which a significant period P is detected is folded according to P and divided into ten equally spaced phase bins. The PC parameter is defined as the fraction of bins containing at least one data point. The MPG parameter is defined as

$$\mathrm{MPG} = \max(\phi_i - \phi_{i+1}), \tag{7}$$

where ϕ_i is the phase of the ith point of the folded light curve. The reduced chi-squared is defined according to the classical formulation:

$$\tilde{\chi}^2(G) = \frac{1}{d}\sum_{i=1}^{N}\frac{(G_i - E_i)^2}{\sigma_i^2}, \tag{8}$$

where N is the number of observations in the segment, $d = N - 3$, G_i the ith data-point of the segment, σ_i the associated

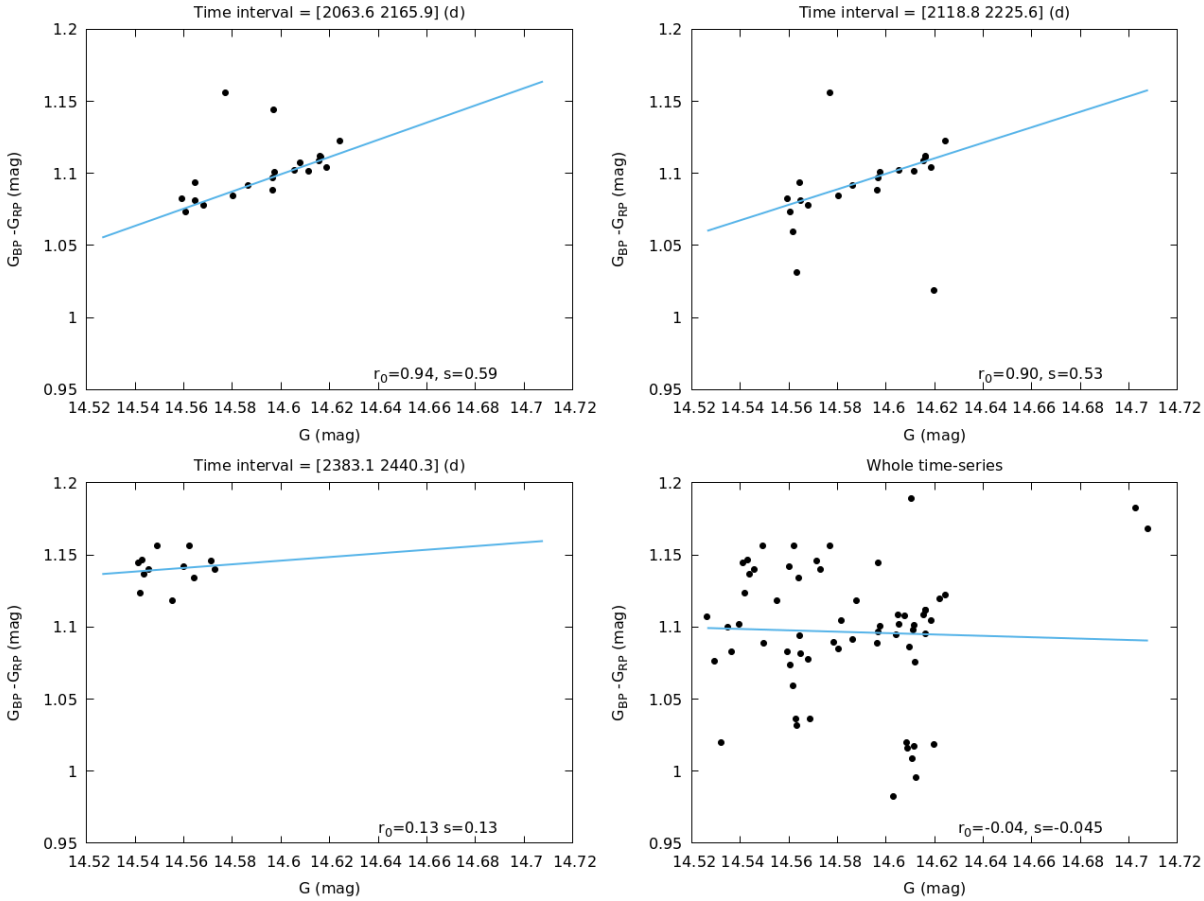

Fig. 12. Same as Fig. 11 for the source `Gaia DR3 6629372460509004928`.

photometric error, and E_i the value predicted by the fitted sinusoidal model.

The Q parameter is defined as:

$$Q = \frac{A_{\rm fit}(G)}{A_{\rm perc}(G)}. \quad (9)$$

The PC and MPG are used to establish the goodness of the sampling, whereas the $\tilde{\chi}^2$ and Q parameters are used to evaluate the goodness of the fit. High PC values and small MPG values indicate a good sampling, whereas $\tilde{\chi}^2$ and Q values close to 1 indicate a good consistency between the data and the fitted sinusoidal function. The Q parameter was introduced because a low chi-squared is not always sufficient to establish the consistency between the sinusoidal model and the data, as discussed below.

In Fig. 16, we show the illustrative case of the source `Gaia DR3 6428392592628180864`. The whole G time series is plotted in the top panel. The top black segments mark the segments into which the time series has been divided. In the first segment (highlighted with the green shadow), the pipeline detected the period $P = 24.613$ d associated with a FAP $= 3 \times 10^{-7}$ whereas in the fourth red shaded segment the detected period is $P_4 = 4.708$ d with a FAP $= 3 \times 10^{-17}$. The central and the bottom panels display the photometric points of the two segments folded according to the detected periods. The sinusoidal curves computed by the fitting procedure are over-plotted on the data. Both periods are associated with FAP values of lower than the threshold used to flag a detection as reliable, but in the first segment the fitted curve is poorly sampled (PC = 0.3) and the photometric points are mainly concentrated in only one phase bin (MPG = 0.88). Moreover, visual inspection of the folded light curves reveals that, in the first segment, though the reduced chi-square is quite close to 1 ($\tilde{\chi}^2 = 4.44$), the fitted sine wave cannot be regarded as a reliable model for the data. The inconsistency between the data and the model, in this case, is reflected by the high Q value. Indeed, the variability amplitude inferred by the data distribution and that inferred from the fitted sine wave are noticeably different from each other with $A_{\rm perc}(G) = 0.054$ mag, $A_{\rm fit}(G) = 1.43$ mag, and $Q = 26.4$.

A given star is flagged and released as a rotational modulation variable if the segments used to infer the `best_rotation_period` parameter satisfy the following requirements:

1. $0.5 \leq Q \leq 1.6$ in all segments;
2. PC ≥ 0.4 and MPG < 0.3 and $\tilde{\chi}^2(G) \leq 32.5$ in at least one of the segments;
3. the segment satisfying requirement (2) needs to be different from the whole time series.

The adopted threshold values occurring in the first requirement are the 5th and 95th percentile of the Q distribution, respectively. The threshold value adopted in the second requirement is the 95th percentile of the $\tilde{\chi}^2(G)$ distribution.

The above criteria are stricter than those adopted in DR2. Indeed, in DR2 only the first and second requirement were adopted and the second requirement did not include the condition on $\tilde{\chi}^2(G)$. The third requirement was added because DR3 time series span a longer time interval than DR2 time series and, as shown in Sect. 2.2, the long-term evolution of MARs could seriously affect the detection of the correct rotation period.

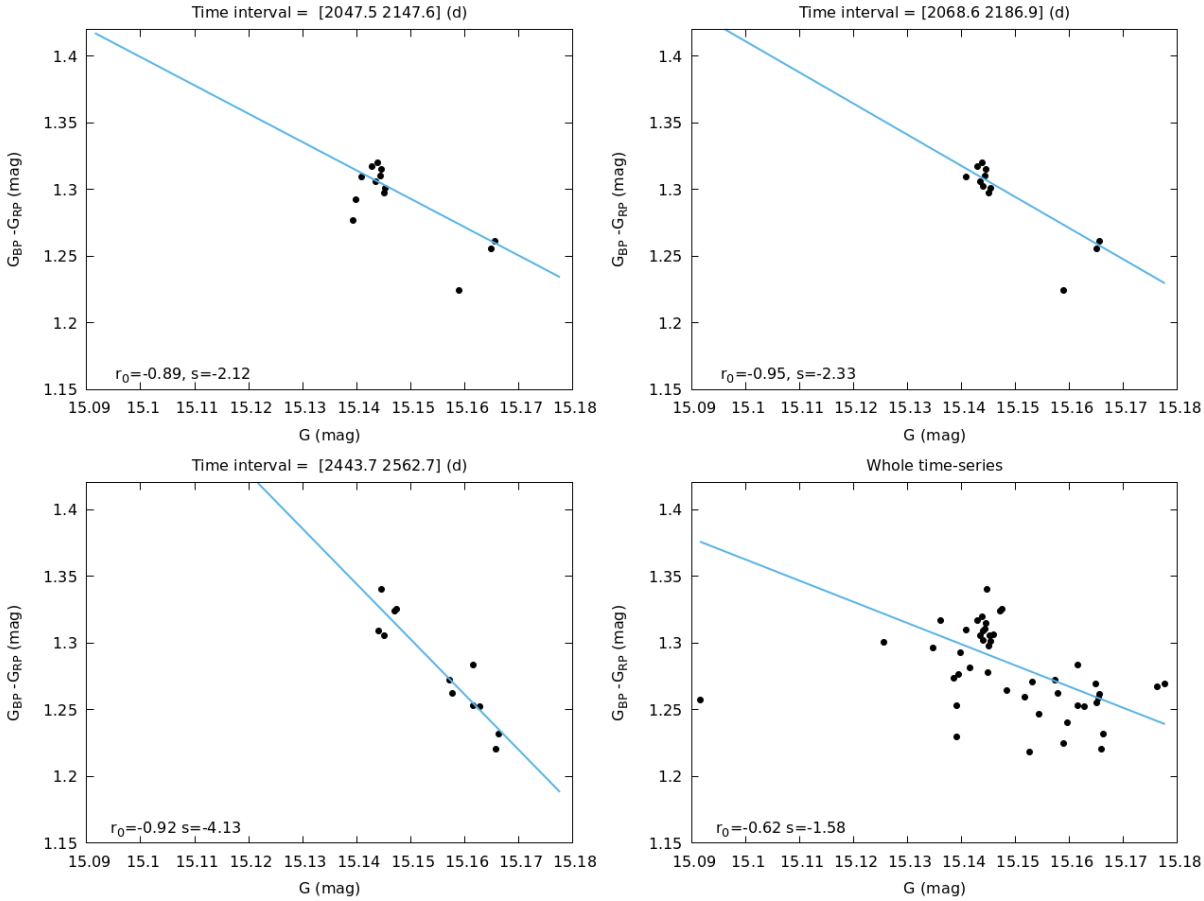

Fig. 13. Same as Fig. 11 for the source `Gaia DR3 5020821858561059456`.

The necessity to enforce the DR2 requirements is evident from the comparison between Figs. 17 and 18, where we plot the `best_rotation_period` distributions obtained by applying the DR2 and the DR3 requirements, respectively. The use of the DR3 requirements filters out most of the spurious periods centred around 0.5, 18, 25, 32, and 49 d. We note that DR3 requirements are still not sufficient to remove all the spurious peaks of the distribution. A further filtering was performed during the post-processing operations as discussed in Appendix A. These post-processing operations allowed us to further clean the sample of the detected variables and to obtain a final catalogue of 474 026 bona fide stars with rotational modulation. The period distribution of this final sample of stars is reported in Fig. 19.

3. Catalogue overview

The `gdr3_rotmod` catalogue reports 474 026 bona fide stars with rotational modulation. For each star, the catalogue provides 66 different parameters whose full description can be found in the Sect. 3.3.15 of the GDR3 documentation (Rimoldini et al. 2022). In the present section, we show the correlation between some of these parameters and discuss the impact of the catalogue on our understanding of the variability phenomena seen in magnetically active stars.

The completeness of the catalogue is limited due to the peculiarity of the *Gaia* scanning law and is about 0.4% down to the limit magnitude $G \simeq 21.5$ and about 4% down to the limit magnitude $G \simeq 15$. These latter are only average values and can significantly change across the sky. In Appendix B.1 and Figs. B.1 and B.3, we show how the detection efficiency (which is an upper limit for the survey completeness) varies with the stellar magnitude and the ecliptic coordinates. The completeness is also a function of the stellar rotation period. In fact, Distefano et al. (2012) demonstrated that the *Gaia* scanning law favours the detection of stars with short rotation periods ($P < 5$ d).

The contamination level of the catalogue has been assessed to be between 6% and 14% and is mainly due to binary systems (see details in Appendix B.2) The rate of correct period detection (estimated by the comparison between *Gaia* and other photometric surveys) has been assessed to be between 70% and 80%. The incorrect detections are partly due to the sparseness of the *Gaia* sampling and partly due to the physical properties of magnetically active stars (see details in Appendix B.3).

Because of the low completeness, the catalogue cannot be regarded as fully representative of the entire population of magnetically active stars. Nevertheless, as we discuss in this section, it is rich in precious information and provides details never seen in previous surveys.

3.1. Stable and unstable stars

In Sect. 2.2, we show two different types of light curve (see Figs. 5–8). In some stars, the rotational modulation signal is very stable, whereas in others, it loses coherence over time. This loss of coherence can be ascribed to different reasons: in some cases, the instability could be due to the intrinsic evolution of MARs, which induces a shift in the phase of the rotational signal; in other cases it could be induced by a combined effect of surface differential rotation (SDR) and MARs latitude migration

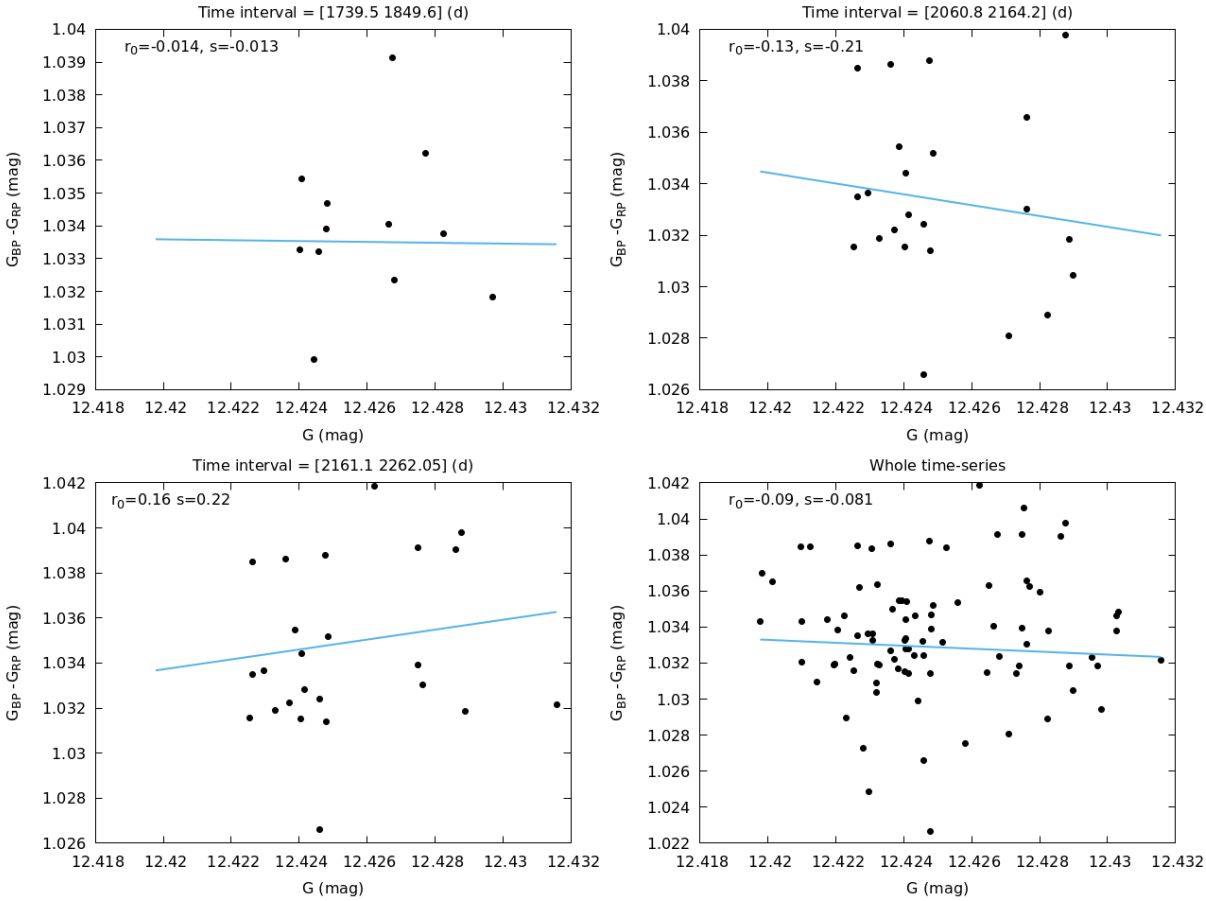

Fig. 14. Same as Fig. 11 for the source `Gaia DR3 5823916073592266624`.

(Distefano et al. 2016) or (in most of cases) could be determined by the appearing and disappearing of MARs at random stellar longitudes (Basri & Shah 2020).

We grouped the whole sample of rotational modulation variables into two distinct sets. We flagged a light curve as stable if it satisfies two conditions: (1) the period search algorithm is able to find a meaningful (FAP < 0.05) period in the whole time series[3]; and (2) the period retrieved in the whole time series differs by less than 5% from the `best_rotation_period` parameter. The stars satisfying these conditions are 38% of the whole sample. The remaining stars were flagged as unstable. In Fig. 20 we show the normalised distributions of periods, amplitudes, and r_0 values for both sets of stars. The stable stars have, on average, shorter periods than the unstable stars (see top left panel). This is in agreement with the findings of Basri et al. (2022), who reported that $\tau_{\mathrm{MARs_{rot}}}$ (i.e., the MARs lifetime expressed in units of the stellar rotation period) tends to be longer for stars with shorter rotation periods. The amplitude distribution (top right panel) is bimodal for stable stars, with two peaks at about 0.015 and 0.005 mag, and unimodal for unstable stars, with a peak at about 0.015 mag. In the bottom left panel, we report the normalised distributions of r_0 values with $p < 0.1$ for both samples of stars. The distribution of stable stars is slightly skewed towards positive values. This trend is more evident if we limit our analysis to stars falling in the magnitude range (13,15.5), where the photometric sensitivity of *Gaia* is higher (see Fig. 15).

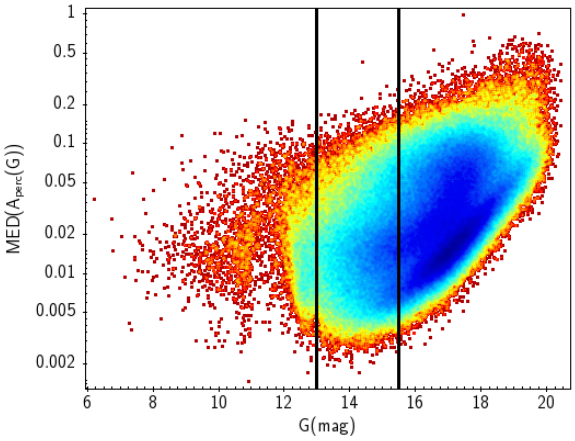

Fig. 15. Median photometric amplitude $A_{\mathrm{perc}}(G)$ vs. G mag.

3.2. Period–amplitude diagram

In the present section, we treat the amplitude A of the rotational modulation signal as a proxy for the stellar magnetic activity level and we investigate its relationship with the stellar rotation period P. The relationship between the stellar magnetic activity and rotation period has been investigated in several papers. According to the traditional picture, based on magnetic activity

[3] The rotation period retrieved in the whole time series is stored in the `gdr3_rotmod` catalogue and is given by the last element of the `segments_rotation_period` array; the associated FAP is given by the last element of `segments_rotation_period_fap`.

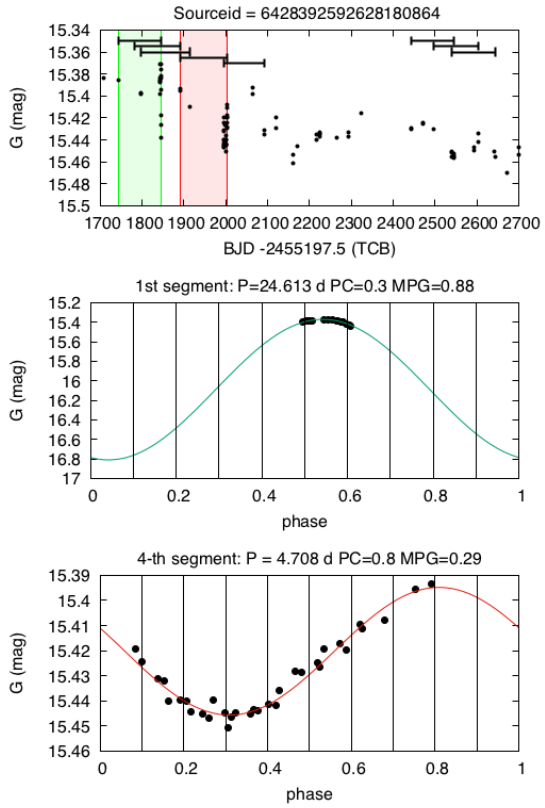

Fig. 16. Illustration of the quality requirements applied to flag a period as reliable. Top panel: full G time series for the star `Gaia DR3 6428392592628180864`. The dark segments enclose the subseries extracted by the segmentation algorithm. The period search algorithm detected the periods $P = 24.613$ d and $P = 4.708$ d in first and fourth segments, respectively. Middle panel: first subseries folded according to the period $P = 24.613$ d. Though the FAP associated with the period is below the rejection threshold, the period is discarded because of the low PC and high MPG values. Bottom panel: fourth subseries folded according to $P = 4.708$ d. In this case, the detected period is flagged as valid because the subseries exhibits an excellent phase coverage (PC = 0.8) and a very small phase gap (MPG = 0.29).

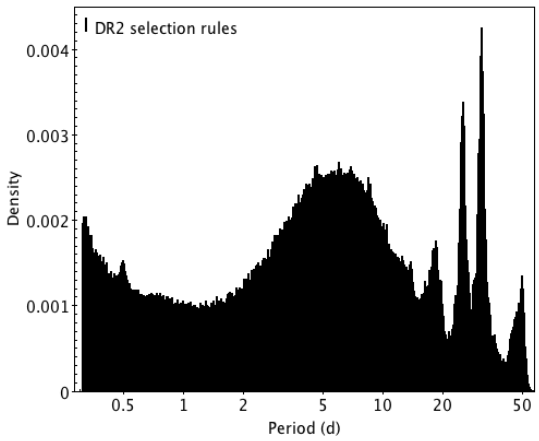

Fig. 17. Distribution of the rotation periods detected by the pipeline. The sharp peaks at 0.5, 18, 25, 32, and 49 d are spurious periods due to sampling and instrumental issues.

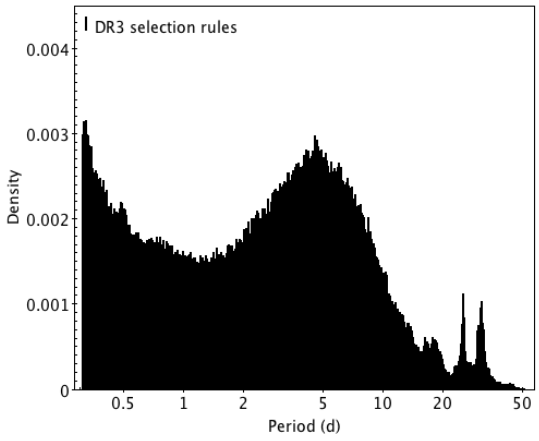

Fig. 18. Distribution of the rotation periods after the cleaning procedure described in Sect. 2.3. The peaks of spurious periods are remarkably flattened but are still persistent in the distribution.

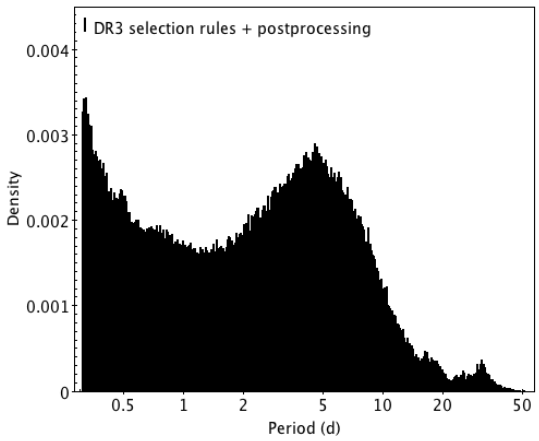

Fig. 19. Distribution of the rotation periods after the post-processing operations described in Appendix A.

indicators like the coronal X-ray emission or the intensity of chromospheric emission lines, the level of magnetic activity increases towards shorter rotation periods and, depending on the stellar mass, saturates below a critical rotation period $P_{\rm crit}$ (see e.g., Pizzolato et al. 2003; Reiners et al. 2014; Fang et al. 2018).

Different authors point out that the amplitude of the rotational modulation signal has a similar dependence on the stellar rotation period than other magnetic activity indicators and that therefore this index can be used as a stellar activity proxy (see e.g., Messina et al. 2003; Hartman et al. 2011; Arkhypov 2018). However, these works are based on limited samples of stars. In recent years, the *Kepler* mission has allowed us to investigate the relationship between this index and the stellar rotation period for several thousands of stars (see e.g., Reinhold et al. 2013; McQuillan et al. 2014; Basri & Nguyen 2018). The *Kepler* data confirm that, despite a certain scatter, the amplitude A tends to increase towards shorter rotation periods and then saturates below a limiting rotation period.

The *Gaia* DR2 data allow a very detailed investigation of the saturated regime of the P–A diagram and reveal, for the first time, the existence of a family of fast rotating ($P \leq 2$ d) stars with small amplitudes. These features have been confirmed by the DR3 data. Fig. 21 displays the P–A diagram for the stars of the `gdr3_rotmod` catalogue with a G magnitude falling in the interval (13,15.5). We restricted our analysis to this interval because at brighter and fainter magnitudes the photometric uncertainty could prevent the detection of the low-amplitude signals and consequently bias the amplitude distribution (see Fig. B.2).

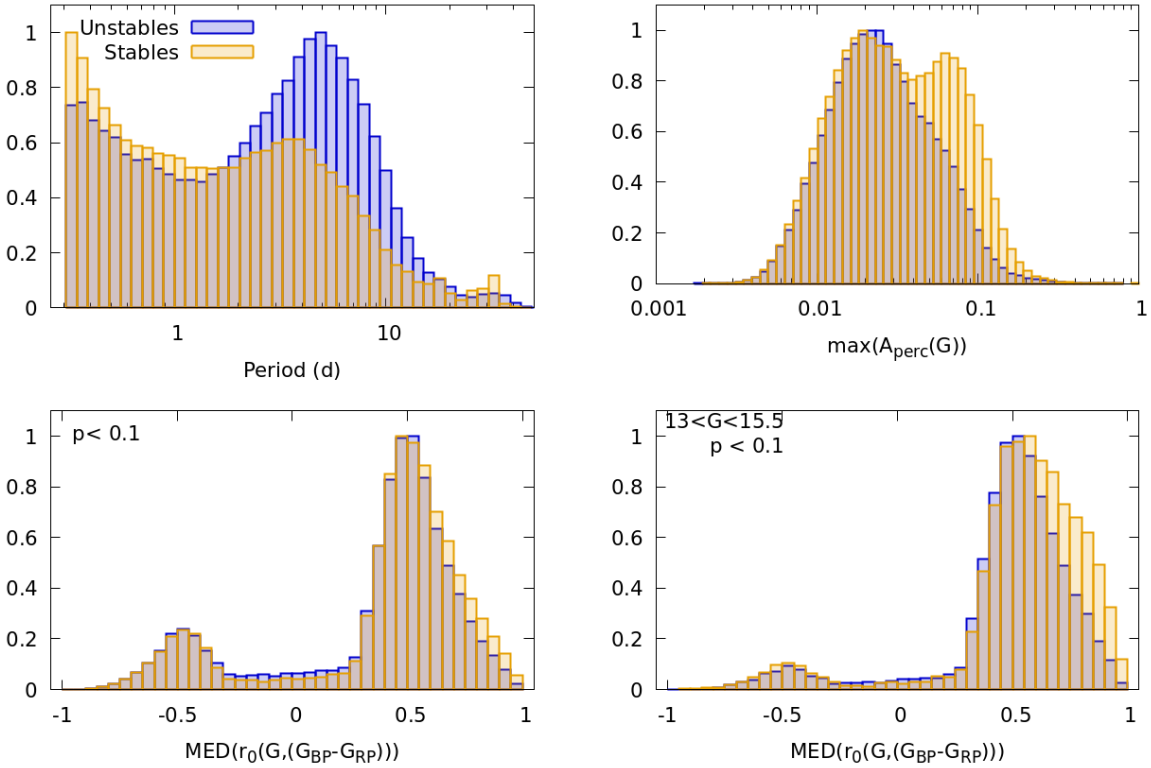

Fig. 20. Normalised distributions of different physical parameters for stable (violet) and unstable (yellow) stars. Top-left panel: normalised distributions of rotation periods. Top-right panel: normalised distribution of variability amplitude. Top-left panel: normalised distribution of MED($r_0(G,(G_{BP}-G_{RP}))$) for all the stars with $p < 0.1$. Top-right panel: normalised distribution of MED($r_0(G,(G_{BP}-G_{RP}))$) for all the stars with $p < 0.1$ and G falling in the interval (13:15.5).

As outlined in Sect. 2.5, for a given star, the catalogue reports different values of the $A_{\rm perc}(G)$ index (one for each of the segments extracted from the whole time series). The amplitude used to obtain Fig. 21 is given by the maximum value of these activity indexes, that is, by max($A_{\rm perc(G)}$)[4]. The rotation period adopted to build the plot is given by the `best_rotation_period` parameter described in Sect. 2.5. The points of the P–A diagram are coloured according to the target density. The blue regions are those with a higher density whereas the yellow ones are those with a lower density.

The picture clearly shows three different clusters of rotating stars that, according to Lanzafame et al. (2019), we call high-amplitude rotators (HARs), low-amplitude-fast-rotators (LAFRs), and low-amplitude slow rotators (LASRs). More specifically, the HAR branch is the region approximately satisfying the condition

$$\max(A_{\rm perc}) > 0.04\,\mathrm{mag}; \tag{10}$$

the LAFR branch is the area satisfying the condition

$$P \leq 2\,\mathrm{d} \text{ and } \max(A_{\rm perc}) < 0.015\,\mathrm{mag}; \tag{11}$$

and the LASR branch is the area satisfying the condition

$$P > 2\,\mathrm{d} \text{ and } \max(A_{\rm perc}) < 0.04\,\mathrm{mag}. \tag{12}$$

The LAFR and the HAR regions are, in particular, well separated by a sparsely populated area that in Lanzafame et al. (2019) is referred to as the gap region and that has been interpreted as evidence of a rapid transition between magnetic configurations characteristic of HARs and configurations characteristic of LAFRs.

As mentioned above, the regime of fast rotation is poorly explored by *Kepler*. The samples of stars analysed for instance by Reinhold et al. (2013) and McQuillan et al. (2014) contain only a few hundred stars with periods shorter than 1 d, whereas the *Gaia* variables with period shorter than 1 d number approximately 150 000[5]. A comparison between the *Kepler* and *Gaia* results performed by Lanzafame et al. (2019) on the other hand shows that the two sets of data are complementary and that the LASR family seen by *Gaia* is the 'tip' of the unsaturated regime sampled by *Kepler* (see Fig. 5 of Lanzafame et al. 2019, for details). The LAFR branch in *Kepler* data is not visible and has been highlighted for the first time by *Gaia*.

Recently, Santos et al. (2021) found a few fast rotating stars with a low-amplitude signal among the *Kepler* targets, but pointed out that these stars are all above the Kraft break (Kraft 1967), which is stars characterised by an effective temperature of $T_{\rm eff} > 6200$ K. This is not the case for the LAFR stars listed in the `gdr3_rotmod` catalogue. Indeed, Fig. 22 shows the $T_{\rm eff}$ distribution for all the LAFR stars. Almost the whole LAFR sample lies under the Kraft break. We note that the $T_{\rm eff}$ values used to build the distribution are the DR3 values inferred by the $G_{\rm BP}$ and $G_{\rm RP}$ mean spectra according to the procedure described in Andrae et al. (2023). The different intensity and topology of surface magnetic fields in the three classes of rotating stars is still

[4] This value is stored in the ttgdr3_rotmod catalogue for each star and is given by the index `maximum_activity_index`.

[5] We remind that, though the average detection efficiency of our pipeline is very low, it dramatically increases towards shorter rotation periods as demonstrated in Distefano et al. (2012).

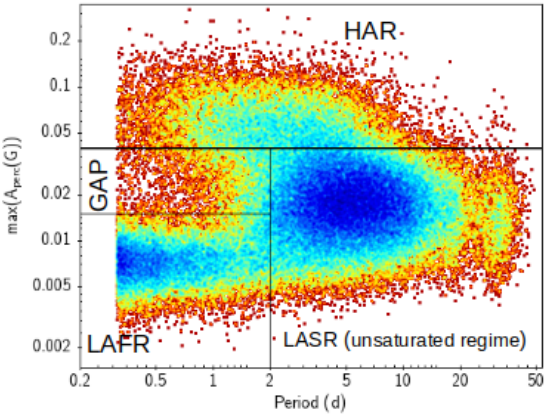

Fig. 21. P–A density diagram for the `gdr3_rotmod` stars with G mag falling in the (13,15.5) interval.

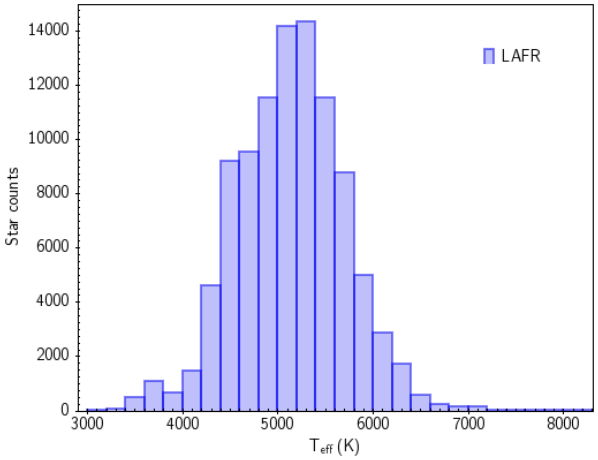

Fig. 22. $T_{\rm eff}$ distribution for the LAFR stars.

unknown, and a deep analysis of *Gaia* DR3 data can provide precious insights into their properties.

Figures 23 and 24 present the density P–A diagram for the stable and unstable stars, respectively. In both pictures, we used stars with G mag falling in the (13, 15.5) range for the reasons discussed at the beginning of this section. The comparison between the two pictures shows that the stable stars are mainly concentrated in the HAR branch, whereas the unstable stars are mainly concentrated in the LAFR and in the LASR branches. In Table 1, we report the percentage of stable and unstable stars in each branch. The highest percentage of stable stars (47%) is found in the HAR branch. This percentage decreases to 34% in the LAFR branch and drops to 28% in the LASR branch.

3.3. The $P-A$–r_0 diagram

Another important insight into the difference between the three branches is given by the analysis of the $r_0(G, G_{\rm BP} - G_{\rm RP})$ distribution. Figure 25 reports the P–A diagrams with data points colour coded according to the median value of the Pearson correlation coefficients obtained in the different segments, that is, MED($r_0(G, G_{\rm BP} - G_{\rm RP})$). These diagrams were obtained by selecting stars with magnitude G falling in the interval (13,15.5). Visual inspection shows that all HAR stars are also RCMC stars, that is, they exhibit a strong positive correlation between magnitude and colour variations. This correlation persists but is attenuated in the LASR branch. Finally, in the LAFR branch, the colour and brightness variations are poorly correlated.

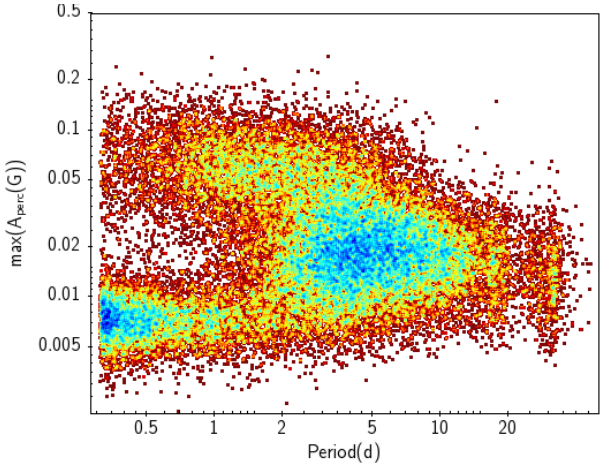

Fig. 23. P–A density diagram obtained for the stable stars with G mag values included in the G range (13,15.5).

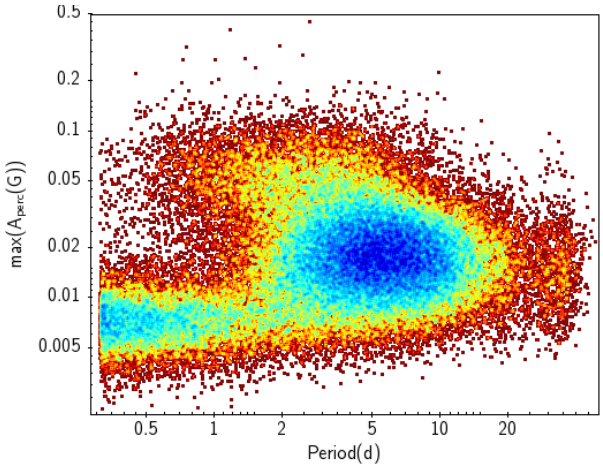

Fig. 24. P–A density diagram obtained for the unstable stars with G mag values included in the G range (13,15.5).

Table 1. Percentage of stable and unstable stars in the different regions of the P–A diagram.

Branch	Stable	Unstable
HAR	47%	53%
GAP	39%	61%
LAFR	34%	66%
LASR	28%	72%

The different nature of the three regimes of rotating stars is also evident in Fig. 26 where the points of the P–A diagram are coloured according to the Pearson Coefficient $r_{0_{\rm TS}}$ value measured in the whole time series. In HAR stars, the correlation between magnitude and colour variation tends to be strong not only in the single segments but also in the whole time series, indicating that in these stars the configuration of MARs is stable for the time interval covered by the *Gaia* DR3 time series. The correlation tends to fade in the LASR branch and is close to zero in the LAFR region.

The $A - P{-}r_0$ diagrams displayed in Figs. 25 and 26 were obtained without taking into account the statistical significance associated with r_0; otherwise the uncorrelated stars would have been erased from the plot. However, these diagrams do not show the position of the small percentage of blueing stars present in the `gdr3_rotmod` catalogue (see Fig. 10).

Figures 27 and 28 show the same $A - P{-}r_0$ diagrams for the sample of stars in which the p value associated with the median Pearson coefficient is less than 0.1. In Fig. 27, the points are colour coded according to median Pearson coefficient, whereas in Fig. 28 the points are colour coded according to the Pearson coefficient measured for the whole time series. Visual inspection of the two figures reveals that blueing stars are mainly located in the LAFR branch and that the anti-correlation is observed only in the segments but not in the whole time series where the Pearson coefficient is close to zero. The pictures clearly show that while the reddening stars located in the HAR region of the diagram are very stable in time (at least in the time spanned by the DR3 time series), the blueing stars are quite unstable.

As stressed in Sect. 2.4, the values of $r_0(G, (G_{\rm BP} - G_{\rm RP}))$ and $r_0(G_{\rm BP}, G_{\rm RP})$ are complex functions of the geometrical distribution of MARs and of the contrast between their temperature and the stellar photosphere temperature. A full understanding of the $A - P{-}r_0$ diagrams would require a theoretical modelling that will be the topic of a future paper (Distefano et al. 2023). However, some useful information for their interpretation might be provided by the works of Messina (2008) and Iwanek et al. (2019), where patterns of brightness–colour variations are analysed in a sample of 14 magnetically active close binary stars and about 12 000 stars in the OGLE (Optical Gravitational Lensing Experiment Udalski et al. 2015) Galactic Bulge fields, respectively. The two works take a slightly different approach: Messina (2008) splits the long-term time series into intervals where the rotation modulation signal is stable, whereas Iwanek et al. (2019) measure the correlation coefficient in the whole OGLE time series, which spans up to 13 years of observations. Hence, the r_0 value computed by Messina (2008) is comparable to our MED(r_0) index and the r_0 computed by Iwanek et al. (2019) is more similar to our $r_{0_{\rm TS}}$ value.

Messina (2008) interprets reddening variables as stars dominated by cool spots and uncorrelated variables as stars dominated by faculae. In particular, he ascribes the lack of correlation between magnitude and colour to the fact that faculae are spatially or temporally uncorrelated with spots. This happens in our Sun for example, where the typical lifetime of faculae is longer than that of spots (see e.g., Lanza et al. 2004). Such an interpretation and the location of reddening and uncorrelated stars in the $P - A{-}r_0$ diagram are coherent with the works of Montet et al. (2017) and Basri (2018), according to which spot-dominated stars have a higher variability amplitude than that of faculae-dominated stars. In particular, Basri (2018) points out that faculae domination can also occur in fast-rotating stars; this this is really surprising because theoretical models predict that fast rotators should be spot-dominated (Shapiro et al. 2014).

As mentioned above, Iwanek et al. (2019) focus their analysis on the long-term correlation between brightness and colour variations. These authors interpret reddening variables as stars dominated by long-lived spots and uncorrelated variables as stars characterised by a high level of magnetic activity and by a rapid change of MARs configuration. This is coherent with the bottom-right panel of Fig. 20, according to which stable stars tend to have a higher correlation coefficient than unstable stars.

Finally, the meaning of blueing variables is quite puzzling. Messina (2008) interprets these stars as binary systems characterised by a magnetically active component and by an inactive earlier type stellar companion, whereas Iwanek et al. (2019) state that these variables could be peculiar stars covered by chemical spots characterised by an overabundance of heavy elements and by variable line-blanketing effect. However, both explanations are unsatisfactory. Indeed, Fig. 29 reports the location of the blueing variable in the $(M_G)_0$ vs. $(G_{\rm BP} - G_{\rm RP})_0$ diagram observed by *Gaia*, where $(M_G)_0$ and $(G_{\rm BP} - G_{\rm RP})_0$ are the absolute magnitude in the G band and the colour $(G_{BP} - G_{RP})$ corrected for interstellar reddening, respectively. If blueing stars were binary systems, they would lie in the binary sequence of the diagram, but this is only true for a few of them. Finally, if blueing variables were peculiar stars, their spectral type should be earlier than $F8$ (Sikora et al. 2019) and their effective temperatures $T_{\rm eff}$ higher than ≃6300 K, but this is the case for just a very small fraction of them (see left panel of Fig. 30).

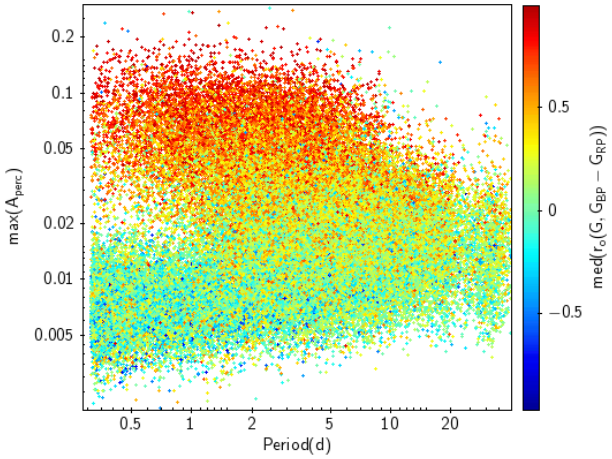

Fig. 25. $P{-}A$ diagram obtained for the stars with G mag values included in the G range (13.5,15). The colours of the points are coded according to MED($r_0((G_{\rm BP} - G_{\rm RP}), G)$), i.e., the median value of the Pearson correlation coefficients measured in the different segments.

Grankin et al. (2007) also found an anti-correlation between colour and magnitude variations in a group of five Classical T-Tauri stars. These authors speculate that this pattern could be ascribed to the combined effect of the stellar occultation by the circumstellar disk and the inhomogeneous structure of the disk itself (see Bouvier et al. 1999 for a phenomenological description and Bertout 2000 for a theoretical model of this variability phenomenon, respectively). However, the variability amplitude associated with this kind of event is usually of the order of a few tenths of a magnitude (see e.g., light curves in Bertout 2000; Stauffer et al. 2016) whereas the blueing variables reported here have smaller amplitudes (see second panel of Fig. 30). Moreover, the period distribution of the blueing variables exhibits a considerable fraction of short rotation periods (see second panel of Fig. 30) whereas the period distribution of stars surrounded by a circumstellar disk is usually skewed towards longer values (see, e.g., Fig. 11 of Davies et al. 2014 and Fig. 8 of Rebull et al. 2018).

4. Conclusion

In the present paper, we describe a method for analysing the *Gaia* photometric time series and for identifying possible spurious signals induced by the instrument or by the physical properties of the investigated source. This method, which is based on the statistical parameter $r_{\rm exf}$ described in Appendix A, can

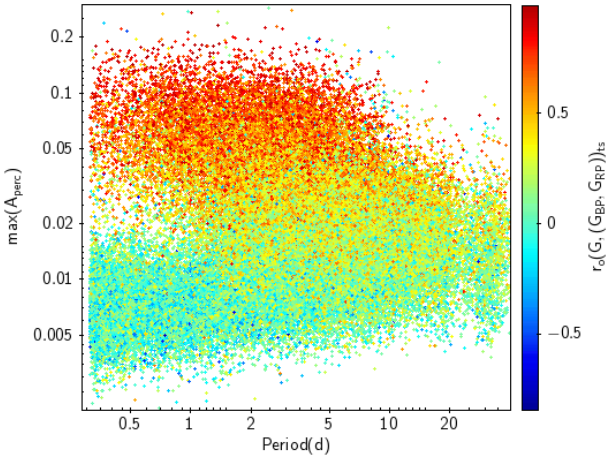

Fig. 26. P–A diagram obtained for the stars with G mag values included in the G range (13.5,15). The colours of the points are coded according to $(r_0((G_{BP}-G_{RP}),G)_{TS})$, i.e., the value of the Pearson correlation coefficient measured for the whole time series.

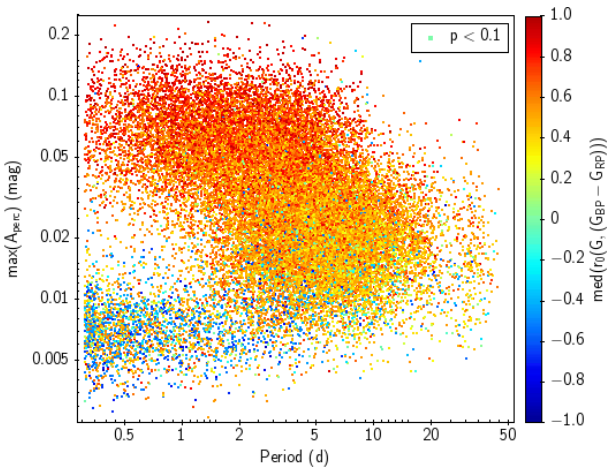

Fig. 27. Same as Fig. 25 but for stars with $p < 0.1$.

be applied to all the *Gaia* DR3 photometric time series and can be useful to researchers interested in specific sources or photometric data. The values of r_{exf} will be published in the *Gaia* DR3 auxiliary table `vari_spurious_signals` with the field name `g_spearman_corr_exf` and will be available for all the released *Gaia* DR3 time series.

The `gdr3_rotmod` catalogue reports 474 026 magnetically active stars, among which about 430 000 stars are, to the best our knowledge, newly discovered variables. As discussed in Appendix B, the completeness of this all-sky catalogue is about 0.4% down to the limiting magnitude $G = 21.5$ and about 4% down to the limiting magnitude $G = 15$. Such a low completeness is partly due to the *Gaia* survey properties (i.e. the time-series sampling and the technical issues discussed in Appendix B) and partly due to the intrinsic nature of solar-like variables. Indeed, stars active or less active than the Sun, in which τ_{MARs} is shorter than a stellar rotation period, are barely detectable using period search algorithms. For all these reasons, the `gdr3_rotmod` catalogue cannot be considered as fully representative of the entire solar-like star population. Nevertheless, this catalogue allows us to retrieve new and precious information on this class of variable stars, especially in the fast-rotation regime, which is poorly explored by previous surveys. Indeed,

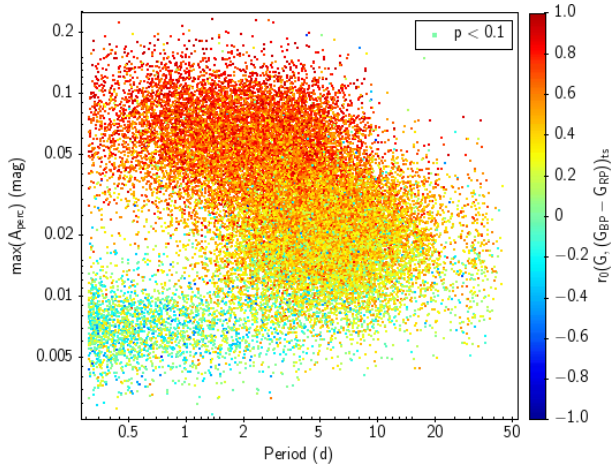

Fig. 28. Same as Fig. 26 but for stars with $p < 0.1$.

about 150 000 stars of the catalogue have rotation periods shorter than 1 day. For each star of the catalogue, we provide a list of about 70 parameters whose detailed description can be found in Rimoldini et al. (2022).

In the present paper, we focus on the stability of the stellar light curves and on the relationship between the period P, the amplitude A of the rotational modulation signal, and the Pearson correlation coefficient r_0 between the brightness and colour variations of the stellar sources. The analysis of the P–A diagram confirms the DR2 findings, namely the existence of three different branches. The LASR and HAR branches correspond to the tip of the unsaturated regime and to the saturated regime, respectively. The LAFR branch is a new family of stars never revealed by previous surveys. The stars located in the HAR branch tend to have a rotational modulation signal that is more stable over time than that revealed in the LAFR and in the LASR branches.

Our analysis of the $P - A - r_0$ diagrams shows that, in HAR stars, there is a strong correlation between colour and brightness variations and that such a correlation is stable along the full 34-months *Gaia* time series. In LAFR stars, instead, the colour and magnitude variations tend to be uncorrelated.

According to the meaning attributed to r_0 by previous works, we can state that HAR stars are dominated by long-lived dark spots, whereas LAFR stars are dominated by bright faculae and are characterised by rapidly variable magnetic fields. The LASR stars are characterised by a moderate correlation between brightness and colour variations. This implies that dark spots are still the main cause of their variability, but that faculae spatially or temporally uncorrelated with the spots tend to attenuate the correlation between magnitude and colour variations.

Finally, there is a small fraction of stars in which the magnitude and colour variation are anti-correlated. These stars are mainly concentrated in the LAFR and LASR regions of the P–A diagram. As discussed Sect. 3.3, the interpretation of these variables is quite puzzling. According to the literature, these stars could be binary systems characterised by an active component and an inactive blue companion or stars covered by chemical spots. However, we show that their luminosity and effective temperature are incompatible with both explanations. A third explanation could be that the variability of these stars is driven by the variable extinction induced by an inhomogeneous circumstellar disk. However, this hypothesis, which is discussed in the previous section, also presents some inconsistencies and should be tested with the aid of circum-stellar disk indicators.

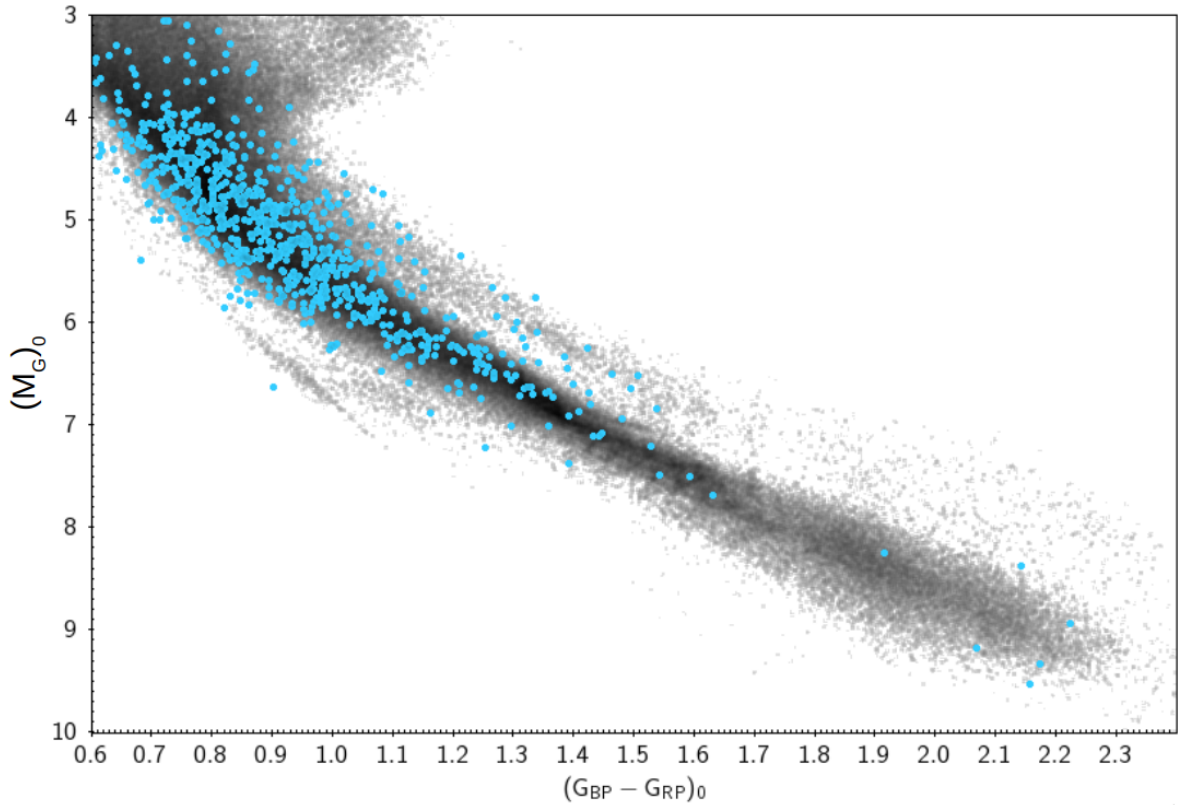

Fig. 29. m_G vs.$(G_{BP} - G_{RP})$ diagram observed by *Gaia*. The blue bullets mark the location of the blueing variables. The region parallel to the main sequence is the binary sequence.

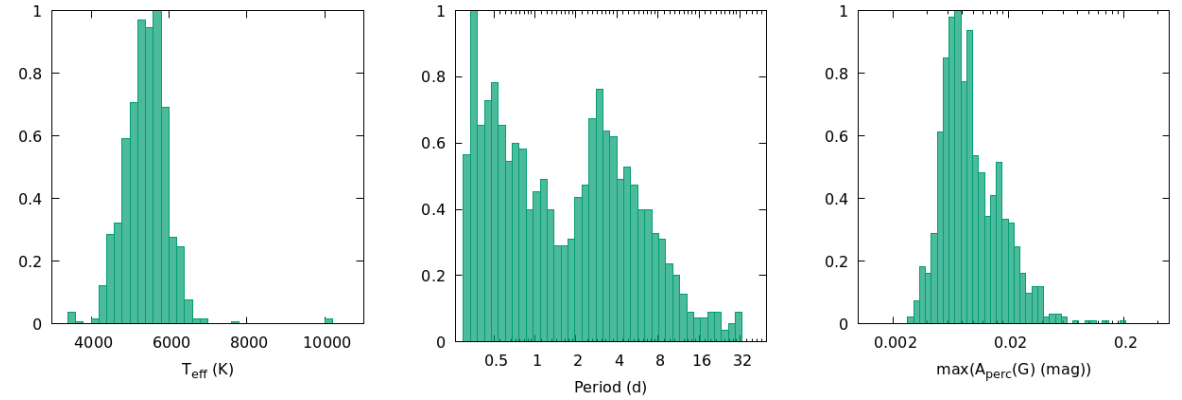

Fig. 30. Distributions of physical parameters for blueing stars. Left panel: $T_{\rm eff}$ distribution. Central panel: rotation period distribution. Right panel: amplitude variability distribution.

Despite the limitations due to the sparse sampling and to the instrumental issues discussed in Appendices A and C, *Gaia* DR3 provides the largest ever catalogue of rotational variables, permitting studies of the patterns of colour–brightness variations from a statistical point of view. Information coming from three photometric bands sheds light on the physical nature of the different branches seen in the period–amplitude diagram. This picture will be enriched in the future *Gaia* releases, where all of the issues discussed in Appendices A and C will be mitigated and where a flaring analysis will be added.

Acknowledgements. The authors are very grateful to the referee Gibor Basri for the constructive comments and suggestion which remarkably helped to improve the quality of the paper. This work presents results from the European Space Agency (ESA) space mission *Gaia*. *Gaia* data are being processed by the Institutions participating in the *Gaia* MultiLateral Agreement (MLA). The *Gaia* mission website is https://www.cosmos.esa.int/gaia. The *Gaia* archive website is https://archives.esac.esa.int/gaia. The *Gaia* mission and data processing have financially been supported by: the Agenzia Spaziale Italiana (ASI) through contracts I/037/08/0, I/058/10/0, 2014-025-R.0, 2014-025-R.1.2015, and 2018-24-HH.0 to the Italian Istituto Nazionale di Astrofisica (INAF), contract 2014-049-R.0/1/2 to INAF for the Space Science Data Centre (SSDC, formerly known as the ASI Science Data Center, ASDC), contracts I/008/10/0, 2013/030/I.0, 2013-030-I.0.1-2015, and 2016-17-I.0 to the Aerospace Logistics Technology Engineering Company (ALTEC S.p.A.), INAF, and the Italian Ministry of Education, University, and Research (Ministero dell'Istruzione, dell'Università e della Ricerca) through the Premiale project 'MIning The Cosmos Big Data and Innovative Italian Technology for Frontier Astrophysics and Cosmology' (MITiC); the Swiss State Secretariat for Education, Research and Innovation through the "Activités Nationales Complémentaires". The *Gaia* project and data processing have made use of: the Set of Identifications, Measurements, and Bibliography for Astronomical Data (SIMBAD, Wenger et al. 2000), the 'Aladin sky atlas' (Bonnarel et al. 2000;

Boch & Fernique 2014), and the VizieR catalogue access tool (Ochsenbein et al. 2000), all operated at the Centre de Données astronomiques de Strasbourg (CDS); the software products TOPCAT (http://www.starlink.ac.uk/topcat/), STIL (http://www.starlink.ac.uk/stil), and STILTS (http://www.starlink.ac.uk/stilts) (Taylor 2005, 2006); Matplotlib (Hunter 2007); Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration 2018).

References

Andrae, R., Fouesneau, M., Sordo, R., et al. 2023, A&A, 674, A27 (*Gaia* DR3 SI)
Arkhypov, O. V., Khodachenko, e. L., Lammer, H., , et al. 2018, MNRAS, 476, 1224
Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, AJ, 156, 123
Baluev, R. V. 2008, MNRAS, 385, 1279
Barnes, S. A. 2003, ApJ, 586, 464
Basri, G. 2018, ApJ, 865, 142
Basri, G., & Nguyen, H. T. 2018, ApJ, 863, 190
Basri, G., & Shah, R. 2020, ApJ, 901, 14
Basri, G., Walkowicz, L. M., Batalha, N., et al. 2010, ApJ, 713, L155
Basri, G., Walkowicz, L. M., Batalha, N., et al. 2011, AJ, 141, 20
Basri, G., Streichenberger, T., McWard, C., et al. 2022, ApJ, 924, 31
Benz, A. O., & Güdel, M. 2010, ARA&A, 48, 241
Bertout, C. 2000, A&A, 363, 984
Boch, T., & Fernique, P. 2014, in Astronomical Data Analysis Software and Systems XXIII, eds. N. Manset, & P. Forshay, ASP Conf. Ser., 485, 277
Bonnarel, F., Fernique, P., Bienaymé, O., et al. 2000, A&AS, 143, 33
Bouvier, J., Chelli, A., Allain, S., et al. 1999, A&A, 349, 619
Castañeda, J., Hobbs, D., Fabricius, C., et al. 2021, Gaia EDR3 documentation Chapter 3: Pre-processing, Gaia EDR3 documentation, European Space Agency; Gaia Data Processing and Analysis Consortium. Online at https://gea.esac.esa.int/archive/documentation/GEDR3/Data_processing/chap_cu3pre/
Chen, X., Wang, S., Deng, L., et al. 2020, ApJS, 249, 18
Davenport, J. R. A. 2016, ApJ, 829, 23
Davies, C. L., Gregory, S. G., & Greaves, J. S. 2014, MNRAS, 444, 1157
Distefano, E., Lanzafame, A. C., Lanza, A. F., et al. 2012, MNRAS, 421, 2774
Distefano, E., Lanzafame, A. C., Lanza, A. F., Messina, S., & Spada, F. 2016, A&A, 591, A43
Distefano, E., Lanzafame, A. C., Lanza, A. F., Messina, S., & Spada, F. 2017, A&A, 606, A58
Distefano, E., Lanzafame, A. C, Brugaletta, E., et al. 2023, A&A, 674, A20 (*Gaia* DR3 SI)
Donahue, R. A., Dobson, A. K., & Baliunas, S. L. 1997a, Sol. Phys., 171, 191
Donahue, R. A., Dobson, A. K., & Baliunas, S. L. 1997b, Sol. Phys., 171, 211
Doyle, L., Ramsay, G., Doyle, J. G., & Wu, K. 2019, MNRAS, 489, 437
Eyer, L., & Mignard, F. 2005, MNRAS, 361, 1136
Eyer, L., Audard, M., Holl, B., et al. 2023, A&A, 674, A13 (*Gaia* DR3 SI)
Fang, X.-S., Zhao, G., Zhao, J.-K., & Bharat Kumar, Y. 2018, MNRAS, 476, 908
Ferreira Lopes, C. E., Leão, I. C., de Freitas, D. B., et al. 2015, A&A, 583, A134
Gaia Collaboration (Brown, A. G. A., et al.) 2016, A&A, 595, A2
Gaia Collaboration (Brown, A. G. A., et al.) 2021, A&A, 649, A1
Graham, M. J., Kulkarni, S. R., Bellm, E. C., et al. 2019, PASP, 131, 078001
Grankin, K. N., Melnikov, S. Y., Bouvier, J., Herbst, W., & Shevchenko, V. S. 2007, A&A, 461, 183
Hartman, J. D., Bakos, G. Á., Noyes, R. W., et al. 2011, AJ, 141, 166
Holl, B., Sozzetti, A., Sahlmann, J., et al. 2023, A&A, 674, A10 (*Gaia* DR3 SI)
Hunter, J. D. 2007, Comput. Sci. Eng., 9, 90
Hussain, G. A. J. 2002, Astron. Nachr., 323, 349
Iwanek, P., Soszyński, I., Skowron, J., et al. 2019, ApJ, 879, 114
Jayasinghe, T., Stanek, K. Z., Kochanek, C. S., et al. 2019, MNRAS, 486, 1907
Kraft, R. P. 1967, ApJ, 150, 551
Lanza, A. F., Rodonò, M., & Pagano, I. 2004, A&A, 425, 707
Lanzafame, A. C., & Spada, F. 2015, A&A, 584, A30
Lanzafame, A. C., Distefano, E., Messina, S., et al. 2018, A&A, 616, A16
Lanzafame, A. C., Distefano, E., Barnes, S. A., & Spada, F. 2019, ApJ, 877, 157
Lehtinen, J., Jetsu, L., Hackman, T., Kajatkari, P., & Henry, G. W. 2016, A&A, 588, A38
Lindegren, L., Klioner, S. A., Hernández, J., et al. 2021, A&A, 649, A2
Masci, F. J., Laher, R. R., Rusholme, B., et al. 2019, PASP, 131, 018003
McQuillan, A., Mazeh, T., & Aigrain, S. 2014, ApJS, 211, 24
Messina, S. 2008, A&A, 480, 495
Messina, S., & Guinan, E. F. 2003, A&A, 409, 1017
Messina, S., Pizzolato, N., Guinan, E. F., & Rodonò, M. 2003, A&A, 410, 671
Messina, S., Rodonò, M., & Cutispoto, G. 2004, Astron. Nachr., 325, 660
Montet, B. T., Tovar, G., & Foreman-Mackey, D. 2017, ApJ, 851, 116
Ochsenbein, F., Bauer, P., & Marcout, J. 2000, A&AS, 143, 23
Oláh, K., Kővári, Z., Petrovay, K., et al. 2016, A&A, 590, A133
Pizzolato, N., Maggio, A., Micela, G., Sciortino, S., & Ventura, P. 2003, A&A, 397, 147
Rebull, L. M., Stauffer, J. R., Cody, A. M., et al. 2018, AJ, 155, 196
Reiners, A., Schüssler, M., & Passegger, V. M. 2014, ApJ, 794, 144
Reinhold, T., Reiners, A., & Basri, G. 2013, A&A, 560, A4
Riello, M., De Angeli, F., Evans, D. W., et al. 2021, A&A, 649, A3
Rimoldini, L., Eyer, L., Audard, M., et al. 2022, Gaia DR3 documentation Chapter 10: Variability, Gaia DR3 documentation, https://gea.esac.esa.int/archive/documentation/GDR3/Data_analysis/chap_cu7var/
Rodonò, M., Messina, S., Lanza, A. F., Cutispoto, G., & Teriaca, L. 2000, A&A, 358, 624
Rowell, N., Davidson, M., Lindegren, L., et al. 2021, A&A, 649, A11
Santos, A. R. G., Breton, S. N., Mathur, S., & García, R. A. 2021, ApJS, 255, 17
Shapiro, A. I., Solanki, S. K., Krivova, N. A., et al. 2014, A&A, 569, A38
Shapiro, A. I., Solanki, S. K., Krivova, N. A., et al. 2017, Nat. Astron., 1, 612
Shappee, B. J., Prieto, J. L., Grupe, D., et al. 2014, ApJ, 788, 48
Shibata, K. 1999, Ap&SS, 264, 129
Sikora, J., Wade, G. A., Power, J., & Neiner, C. 2019, MNRAS, 483, 2300
Skumanich, A. 1972, ApJ, 171, 565
Stauffer, J., Cody, A. M., Rebull, L., et al. 2016, AJ, 151, 60
Süveges, M., Guy, L. P., Eyer, L., et al. 2015, MNRAS, 450, 2052
Taylor, M. B. 2005, in Astronomical Data Analysis Software and Systems XIV, eds. P. Shopbell, M. Britton, & R. Ebert , ASP Conf. Ser., 347, 29
Taylor, M. B. 2006, in Astronomical Data Analysis Software and Systems XV, eds. C. Gabriel, C. Arviset, D. Ponz, & S. Enrique , ASP Conf. Ser., 351, 666
Udalski, A., Szymański, M. K., & Szymański, G. 2015, Acta Astron., 65, 1
Walkowicz, L. M., Basri, G., Batalha, N., et al. 2011, AJ, 141, 50
Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, A&AS, 143, 9
Zechmeister, M., & Kürster, M. 2009, A&A, 496, 577

Appendix A: Post-processing filtering

Although the pipeline described in Sect. 2 employs several quality-assurance criteria, it is not able to completely remove the spurious periods occurring in the `best_rotation_period` distribution (see discussion in Sec. 2.6 and Fig. 18). In order to understand the origin of these peaks in the `best_rotation_period` distribution, we performed a deep investigation of *Gaia* time series and identified two new criteria to filter out the sources affected by the spurious signals. These criteria are based on the analysis of the per-transit corrected excess flux time series and on the image parameter determination (IPD) harmonic model time series and are discussed in the following subsections.

A.1. The per-transit corrected excess flux

The corrected excess flux C^* is a parameter defined by Riello et al. (2021) in order to measure the consistency between the G, $G_{\rm BP}$, and $G_{\rm RP}$ cumulated photometry for a given source. Firstly, Riello et al. (2021) defines the excess flux factor C as

$$C = \frac{I_{G_{\rm BP}} + I_{G_{\rm RP}}}{I_G}, \tag{A.1}$$

where $I_{G_{\rm BP}}$, $I_{G_{\rm RP}}$, and $I_{\rm G}$ are the cumulative fluxes in the $G_{\rm BP}$, $G_{\rm RP}$, and G bands, respectively. Given the instrumental response and the profiles of the *Gaia* passbands, C should be slightly larger than unity for a single and isolated main sequence star. In reality, C has a colour dependency fitted by a polynomial function f($G_{\rm BP} - G_{\rm RP}$) whose coefficients are provided in Table 2 of Riello et al. (2021). The corrected excess factor C^* for a given source is then defined as the difference between the measured and the expected C:

$$C^* = C - f(I_{G_{\rm BP}} - I_{G_{\rm RP}}). \tag{A.2}$$

A corrected excess factor significantly different from zero may be an indicator of calibration issues (i.e. blending or background overestimation) or of a peculiar source (i.e. an extended source or a star with strong emission lines in the region where the G and the $G_{\rm RP}$ bands have different sensitivity).

We defined the per-transit corrected excess flux c^* by applying Eq. A.2 to the per-transit photometry. In this way, we obtained a c^* time series for each source that allows us to investigate whether or not the consistency between the three photometric fluxes exhibits some time dependence that could affect the data and explain the detection of the spurious periods.

In order to identify and discriminate between the problematic sources, we focussed our analysis on the correlation between G and c^*. Indeed for isolated, single, main sequence sources, no significant correlation is expected. For each source, we computed the Spearman correlation coefficient $r_{\rm exf} = r(\mathrm{G}, c^*)$ between the G measurements and the c^* values in order to test whether or not the two quantities are connected by a monotonic relationship. In Fig. A.1, we display the distribution of $r_{\rm exf}$ for all the stars detected by the pipeline. Although the peak is around zero, the distribution is moderately skewed with a tail towards positive values.

We investigated the relationship between the detected periods and the $r_{\rm exf}$ values and found that most of the sources with spurious periods are located in the tail of the $r_{\rm exf}$ distribution. This can be seen easily in Fig. A.2, which reports the *period*$-r_{\rm exf}$ density diagram. Indeed, the sources with spurious periods are mainly concentrated in the region of the diagram corresponding to $r_{\rm exf} > 0.5$.

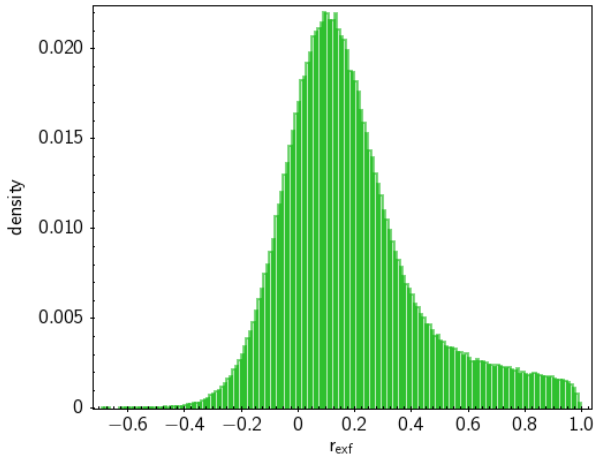

Fig. A.1. Distribution histogram of the Spearman correlation coefficient $r(G, c^*)$ between the G magnitude and the corrected excess factor c^*.

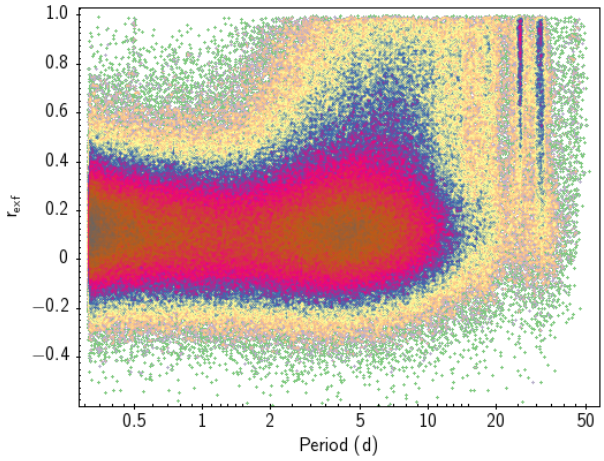

Fig. A.2. Density distribution of the correlation coefficient $r_{\rm exf} = r(G, c^*)$ vs. the detected rotation period. The spurious periods around 25 and 32 d are characterised by a high correlation coefficient.

We visually inspected G, $G_{\rm BP}$, and $G_{\rm RP}$ photometric time series and $G_{\rm BP}$ and $G_{\rm RP}$ epoch spectra for tens of stars with $r_{\rm exf} > 0.5$ and found that the high correlation index can be due to three different factors:

1. the star is a partially resolved binary;
2. the star has some calibration issue due to the transition from a window-gate configuration to another configuration; or
3. the star has strong emission lines whose intensity is correlated with G.

Figure A.3 shows an example of the first kind of source. In the top left panel, we display c^* vs. G for the star `Gaia DR3 2624130051834565504`. The two quantities are highly correlated with a Spearman coefficient $r(c^*, G) = 0.88$. The data are clustered in two different groups: at the brightest G magnitudes, the c^* parameter is close to 0, whereas at the faintest magnitudes c^* is around 0.3. The visual analysis of the epoch spectra time series reveals that the star is blended with another source that we identified as the star `Gaia DR3 2624130051834223104` located at a distance of 0.6 arcsec. When c^* is close to 0, the two stars are unresolved by both the astrometric and the spectrophotometric instrument: in this case, the two instruments measure the total flux coming from the two stars and I_G is consistent with the sum of $I_{G_{\rm BP}}$ and $I_{G_{\rm RP}}$. When c^* is close to 0.3, the two stars are fully resolved by the astrometric instrument

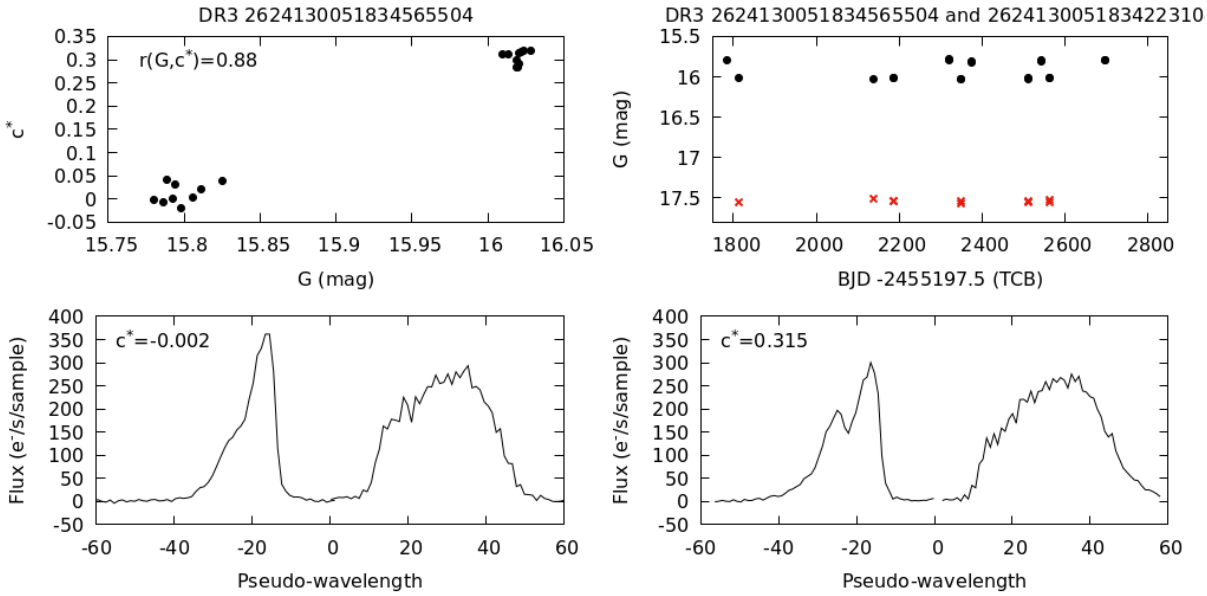

Fig. A.3. Example of a partially resolved binary. Top left panel: c^* vs. G for the star `DR3 2624130051834565504`. The data are distributed around two clumps, one around $c^* \simeq 0$ and the second around $c^* \simeq 0.315$. Top right panel: G magnitude time series of the star `Gaia DR3 2624130051834565504` and of its close neighbour, `Gaia DR3 2624130051834223104`. The two sources form a partially resolved binary; i.e. depending on the observation time and on the scan angle, the two stars can be resolved or unresolved in the astrometric instrument. Bottom left panel: $G_{\rm BP}$ and $G_{\rm RP}$ epoch spectra corresponding to the case in which the two stars are unresolved both by the astrometric and spectrophotometric instruments. In this case, c^* is close to 0 because both instruments collect the fluxes coming from the two stars. Bottom right panel: $G_{\rm BP}$ and $G_{\rm RP}$ epoch spectra corresponding to the case in which the stars are resolved in the astrometric field but blended in the $G_{\rm BP}$ and $G_{\rm RP}$ FoV. In this case, the source exhibits a strong excess factor because the astrometric instrument collects the flux coming from a single source whereas the spectrophotometric instrument collects the flux coming from the star and its neighbour. We note that the $G_{\rm BP}$ and $G_{\rm RP}$ epoch spectra are only shown here for illustration; they will be available to the community in *Gaia* DR4.

and partially resolved by the spectrophotometric instrument: in this second case, the astrometric instrument provides distinct measurements for the two stars, whereas the spectrophotometric instrument attributes all the measured flux only to `Gaia DR3 2624130051834565504`[6] , which is therefore found to have an excess of $I_{G_{\rm BP}}$ and $I_{G_{\rm RP}}$ fluxes. In the bottom panels of Fig. A.3, we display two epoch spectra corresponding to the two different c^* values. When c^* is close to 0.3, the $G_{\rm BP}$ spectrum exhibits two distinct peaks that correspond to the different sources: in this case, the star is partially resolved in the $G_{\rm BP}$ epoch spectrum and unresolved in the $G_{\rm RP}$ epoch spectrum. When c^* is close to 0, the star is unresolved in both epoch spectra. In the top right panel, we show the G time series of the two stars: the black bullets are used to mark the G measurement of `Gaia DR3 2624130051834565504` whereas the red crosses mark the G measurements of `Gaia DR3 2624130051834223104`. When the stars are unresolved, G gets brighter for `Gaia DR3 2624130051834565504` and no measurements are supplied for `Gaia DR3 2624130051834223104`. In such a case, the alternation between measurements in which the sources are fully resolved and measurements in which they are unresolved depends on the scan angle with which the two sources are imaged (see Holl et al. 2023, for further details).

In Fig. A.4, we illustrate an example of the second kind of source. In the top-left panel of the picture we plot G vs. c^* values for the star `Gaia DR3 5367009382403022592`. The two quantities are highly correlated with a Spearman coefficient $r(c^*,\mathrm{G}) = 0.99$. In the bottom panels, we display the epoch spectra corresponding to $c^* = 0.002$ and $c^* = 0.29$. In this star, the spectra do not exhibit significant differences that could explain the high value of the Spearman coefficient. In the top-right panel of the picture, we instead present the time series of I_G and of the sum $I_{G_{\rm BP}} + I_{G_{\rm RP}}$: the sum of $G_{\rm BP}$ and $G_{\rm RP}$ fluxes is almost constant, whereas the I_G flux exhibits time variations of the order of 10%. In this source, the high value of the Spearman coefficient is due to the fact that, as the $I_{G_{\rm BP}} + I_{G_{\rm RP}}$ sum is constant, c^* is directly proportional to the I_G flux. The G variability is probably due to the fact that the star flux has been measured with different gate-window configurations that, in some cases, could have different instrumental responses. Indeed, the star has a magnitude of about 13, which is the value at which the transition between the 2D and the 1D observation window occurs. The signal detected in this star is therefore probably induced by the switching between the different configurations of the photometric instrument.

Finally, in Fig. A.5 and Fig. A.6 we present two examples of the third kind of star. In the top panel of Fig. A.5, we report c^* vs. G for the source `Gaia DR3 3220599441863236608`. The two quantities are moderately correlated with a Spearman coefficient $r(G, c^*) = 0.57$. In the medium and bottom panels of the picture, we display the epoch spectra corresponding to two different transits with $c^* = 0.026$ and $c^* = 0.2$, respectively. In the transit with the higher c^* value, the $G_{\rm RP}$ epoch spectrum exhibits a strong emission line located in a region where the transmissivity of the $G_{\rm RP}$ passband is higher than the G one. This explains the excess of $I_{G_{\rm BP}} + I_{G_{\rm RP}}$ with respect to I_G. A similar situation is illustrated in Fig. A.6 for the star `Gaia DR3 3208310853235527424`. In these two latter stars, the variability detected in the G time series is due to physical reasons and not to instrumental issues.

[6] In DR3, transits affected by blending are simply flagged. In DR4, a de-blending procedure will be put in place.

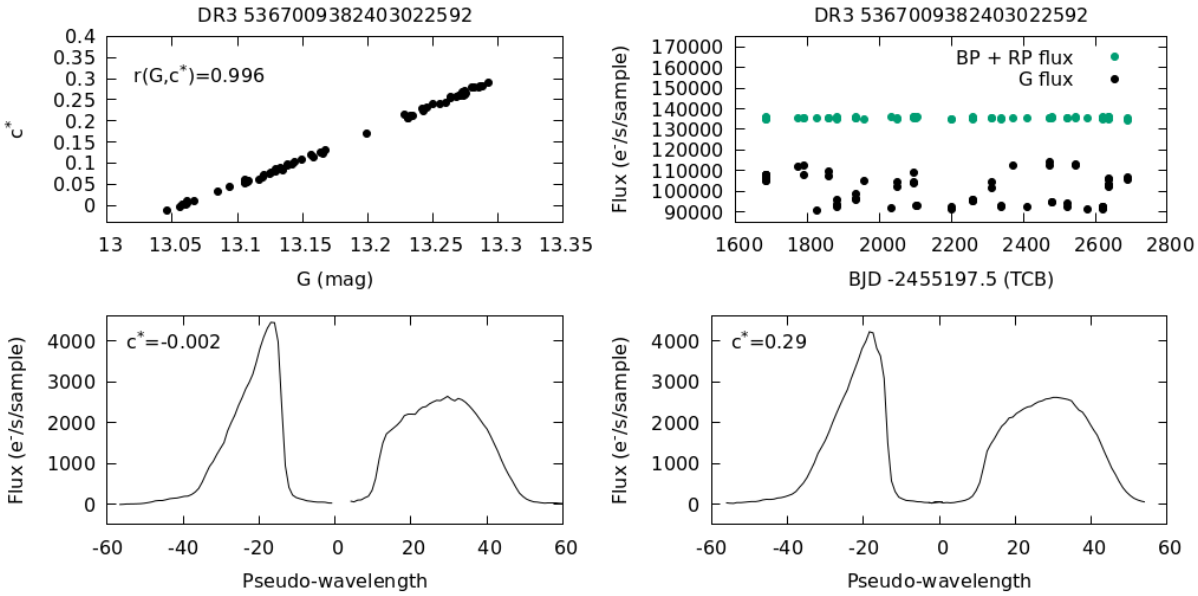

Fig. A.4. Example of a bad calibrated source. Top left panel: c^* vs. G for the star `Gaia DR3 5367009382403022592`. The two quantities are strongly correlated ($r_{\rm exf}$ = 0.996). Top right panel: I_G (black bullets) and $I_{G_{\rm BP}} + I_{G_{\rm RP}}$ (green bullets) time series. While $I_{G_{\rm BP}} + I_{G_{\rm RP}}$ is almost constant vs. time, the I_G time series exhibits temporal variations with a peak-to-peak amplitude of about 20%. This variability is very likely due to instrumental issues, otherwise it should be visible also in the $I_{G_{\rm BP}} + I_{G_{\rm RP}}$ time series. Bottom left panel: $G_{\rm BP}$ and $G_{\rm RP}$ epoch spectra corresponding to c^* = 0.002. Bottom right panel:$G_{\rm BP}$ and $G_{\rm RP}$ epoch spectra corresponding to c^* = 0.29. We note that the $G_{\rm BP}$ and $G_{\rm RP}$ epoch spectra are only shown here for illustration; they will be available to the community in *Gaia* DR4.

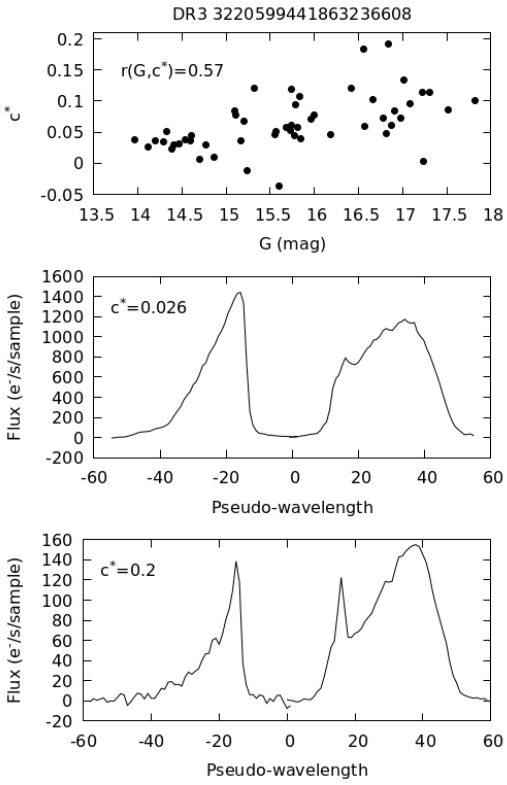

Fig. A.5. Example of a star with a strong emission line. Top panel: c^* vs. G for the star `Gaia DR3 3220599441863236608`. The two quantities are moderately correlated ($r_{\rm exf}$ = 0.57). Middle panel: $G_{\rm BP}$ and $G_{\rm RP}$ epoch spectra corresponding to c^* = 0.026. Bottom panel: $G_{\rm BP}$ and $G_{\rm RP}$ epoch spectra corresponding to c^* = 0.2. In this case the high corrected excess factor is due to a strong emission line falling in a region where the sensitivity of the G and $G_{\rm RP}$ bands are different. We note that the $G_{\rm BP}$ and $G_{\rm RP}$ epoch spectra are only shown here for illustration; they will be available to the community in *Gaia* DR4.

A.2. Spurious periods induced by the scan angle

The procedure of point spread function (PSF) and line spread function (LSF) fitting of the *Gaia* pipeline is optimised for point-like sources (Riello et al. 2021). In the case of extended sources, such as galaxies, or elongated sources, such as partially resolved binaries, the computation of G, $G_{\rm BP}$, and $G_{\rm RP}$ values can be affected by the scan direction ψ of the satellite. For each *Gaia* source, the goodness-of-fit (GoF; the reduced chi-square) of the image parameter determination (IPD; Castañeda et al. 2021, Sect. 3.3.6) is fitted with a harmonic model $hm(\psi)$ (Lindegren et al. 2021, Sect. 5):

$$\ln(GoF) = hm(\psi) = c_0 + c_2\cos(2\psi) + s_2\sin(2\psi), \tag{A.3}$$

whose amplitude is given by

$$\texttt{ipd\_gof\_harmonic\_amplitude} = \sqrt{(c_2^2 + s_2^2)}. \tag{A.4}$$

The higher the amplitude of the harmonic model, the higher the probability that the source is an extended source or partially resolved binary and that G, $G_{\rm BP}$, and $G_{\rm RP}$ magnitudes are affected by ψ. In order to quantify how much the G band signal is affected by the variation in the IPD GoF as a function of scan angle, Holl et al. 2023 computed the Spearman correlation coefficient between the G band signal as a function of scan angle, and the IPD GoF model sampled at the observed scan angles:

$$r_{\rm ipd} = r\left(\{G_i(\psi_i),\ hm(\psi_i) \mid i \in 1, ..., N\}\right), \tag{A.5}$$

with i being the observation index of a source with a total of N observations, and ψ_i being the associated scan angle. In sources with a high $r_{\rm ipd}$ value, the G time-series variation is coherently varying as a function of scan angle and has a phase similar to the IPD GoF model, suggesting it is strongly affected by a scan-angle-dependent signal and therefore could lead to the detection of spurious periods (further details on $r_{\rm ipd}$ can be found in

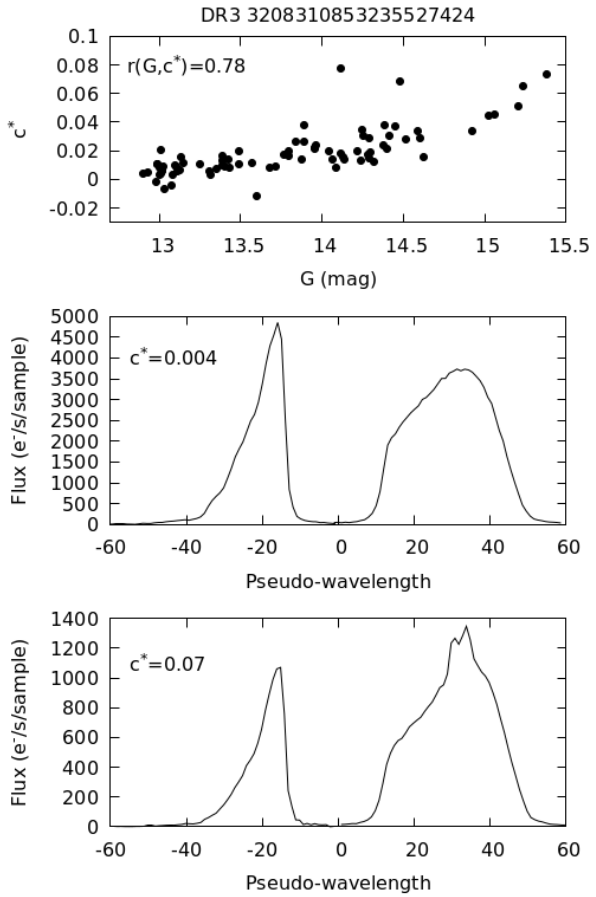

Fig. A.6. Same as Fig. A.5 but for the star `Gaia DR3 3208310853235527424`.

Holl et al. 2023.) In Fig. A.7, we show the density plot of $r_{\rm ipd}$ vs. `best_rotation_period` for the sources detected by our pipeline. Also in this case, the spurious periods at 18, 25, and 32 d tend to be concentrated at higher $r_{\rm ipd}$ values ($r_{\rm ipd} > 0.8$).

A.3. Filtering criterion

In order to avoid reporting periods in our catalogue that are affected by the issues illustrated in the above sections, we decided to discard all sources satisfying the condition:

$$|r_{\rm exf}| > 0.7 \text{ or } |r_{\rm ipd}| > 0.7. \quad \text{(A.6)}$$

We decided to use both indexes because they are complementary as illustrated in Fig. A.8. For a considerable number of stars, the two indexes are correlated and high $r_{\rm exf}$ values correspond to high $r_{\rm ipd}$ values, but there is a subset of stars that

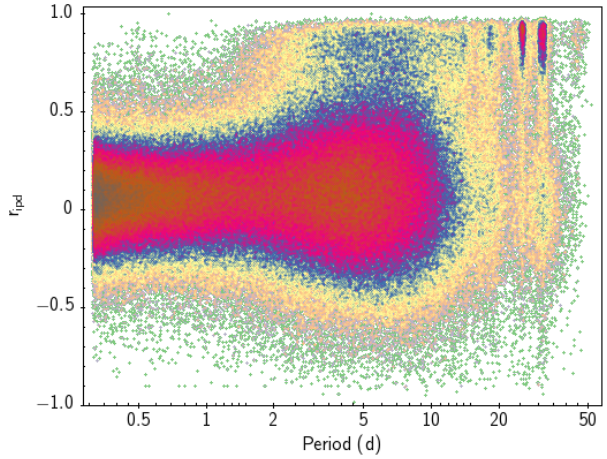

Fig. A.7. Density map of the coefficient $r_{\rm ipd} = r(hm(\psi), G)$ vs. the detected rotation period.

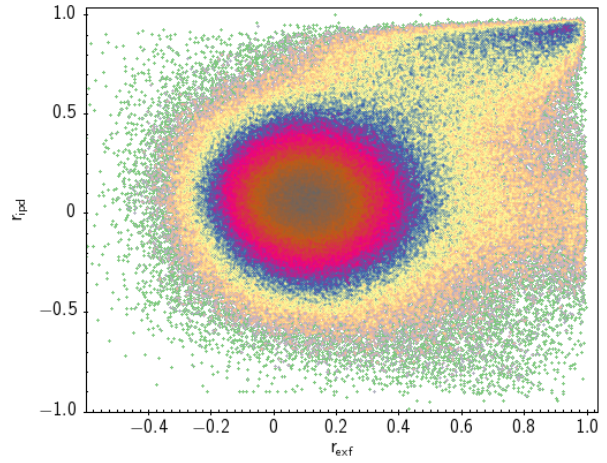

Fig. A.8. Density map of $r_{\rm ipd}$ vs. $r_{\rm exf}$.

have $r_{\rm ipd}$ values close to 0 and high $r_{\rm exf}$ values. This is the case for the star `Gaia DR3 5367009382403022592` reported in Fig. A.4, for example. This star, as discussed in the above section, has been observed in time with different window-gate configurations and suffers from calibration issues that give a $r_{\rm exf}$ index close to 1. The $r_{\rm ipd}$ index is instead very low ($r_{\rm ipd} = 0.2$) because the source is an isolated single star and the PSF fitting is not affected by the scan angle direction. After applying the filter given by Eq. A.6, we obtained a final sample of 474 026 bona fide stars with rotational modulation. The period distribution of these sources is displayed in Fig. 19. Most of the peaks associated with spurious periods are removed or considerably attenuated.

Appendix B: Validation and quality assessment of the data

As mentioned in the above section, the gdr3_rotmod catalogue reports 474 026 sources whose variability is likely induced by stellar magnetic activity. In order to estimate the percentage of newly discovered variables, we cross-matched the gdr3_rotmod with the *Gaia* DR3 catalogue of known variable stars (Panagiotis et al., in prep., hereafter gdr3_xm[]). This latter, compiled on the basis of 152 catalogues, reports the *Gaia* counterparts for a list of about 4.9 million variable sources known from the literature. The cross-match between the gdr3_rotmod and the gdr3_xm catalogues revealed 42306 variables common to the two catalogues. Therefore, about 430000 sources reported in the gdr3_rotmod catalogue are newly discovered variables.

In this section, we assess the quality of the gdr3_rotmod catalogue by estimating the detection efficiency, the contamination, and the percentage of correct rotation periods. Finally, a comparison between the gdr3_rotmod and the gdr2_rotmod catalogues is also carried out and discussed.

B.1. Detection efficiency

In principle, all the stars with magnetic activity are characterised by a rotational modulation signal. An estimate of the survey completeness could therefore be given by the equation:

$$C = \frac{N_{\rm rot}}{N_{\rm sel} \times {\rm CF}}, \tag{B.1}$$

where $N_{\rm rot}$ is the number of stars for which the pipeline detected a rotational modulation signal, $N_{\rm sel}$ is the number of sources selected by the pipeline, that is those falling in the HR region marked in Fig. 1, and CF is a correction factor that should take into account the following issues:

- the selection sample can miss a certain percentage of good candidates because of the 20% threshold in the relative error on parallax and because of the sharp edges of the selection region (see Fig. 1);
- the selection region can host contaminant sources;
- the photometric *Gaia* completeness changes with the stellar magnitude and with the crowding of the scanned region.

Given the above issues, making a precise estimate of CF is very difficult; indeed, it would require 'aprioristic' knowledge of how magnetically active stars are distributed in our Galaxy and detailed knowledge of the *Gaia* photometric completeness. This latter is known only on a large scale (see discussion in Sect. 6 of Gaia Collaboration 2021).

For all these reasons, in the present paper, we limit ourselves to estimating only the detection efficiency of the pipeline, which is given (as a percentage) by:

$$E = 100 \times \frac{N_{\rm rot}}{N_{\rm sel}} \simeq 0.4\%. \tag{B.2}$$

The detection efficiency is a complex function of the stellar magnitude, the amplitude of the rotational modulation, the stellar rotation period, and the ecliptic latitude β. Indeed, as discussed in Distefano et al. (2012), the detection of the stellar rotation period is favoured at certain ecliptic latitudes and, in general, is biased towards shorter rotation periods ($P_{\rm rot} \leq 5d$).

We studied how E depends on ecliptic latitude by binning the $N_{\rm sel}$ sources detected by the pipeline in 1 *deg*-intervals. For each interval, we computed the efficiency $e(\beta)$ as

$$e(\beta) = 100 \times \frac{n_{\rm rot}(\beta)}{n_{\rm sel}(\beta)}, \tag{B.3}$$

where $n_{\rm rot}(\beta)$ is the number of sources detected in the interval $(\beta - 0.5\,{\rm deg}, \beta + 0.5\,{\rm deg})$ and $n_{\rm sel}(\beta)$ the number of sources selected in the same interval. In the same way, we studied how E depends on the G magnitude by binning the data in 0.5*mag*-intervals and computing the efficiency e(G) as:

$$\mathrm{e(G)} = 100 \times \frac{\mathrm{n_{rot}(G)}}{\mathrm{n_{sel}(G)}}. \tag{B.4}$$

In the top panel of Fig. B.1, we plot the trend of the detection efficiency vs. β (cyan bullets): $e(\beta)$ is almost constant ($\simeq 0.03\%$) in the latitude range $(-30\,{\rm deg}, 30\,{\rm deg})$, increases up to 3% at $\beta = \pm 45$ deg and then abruptly decreases and reaches a minimum value of 0.004 % at $\beta = \pm 55$ deg. Finally, it increases again up to 2% at ecliptic poles. This trend is a direct consequence of the Gaia scanning law, which is strongly dependent on ecliptic latitude and determines a better sampling around ecliptic poles and around the latitudes $\beta = \pm 45$ deg (see Figs. 3 and 4 of the present paper and the discussion in Distefano et al. (2012) for further details). The detection efficiency $e(\beta)$ of the gdr2_rotmod catalogue is also plotted (yellow bullets) for comparison.

In the bottom panel of Fig. B.1, we plot how the detection efficiency changes with G : in the G range (13,20.5), the trend of $e(G)$ is very regular and gently decreases with increasing magnitude. In the G range (6,13), the dependency on G is more complex and reflects the trend of the *Gaia* photometric uncertainties in that magnitude range. In Fig.B.2 we display the average value of the G photometric uncertainties for the sources selected by the pipeline. The uncertainty curve exhibits two peaks at $G \simeq 6.5$ *mag* and at $G \simeq 11.5$ *mag*: at the same magnitudes, $e(G)$ has two depth minima. The irregular trend of the *Gaia* photometric uncertainty in the (6,13) mag range is due to the switch between the different window-gate configurations adopted by *Gaia* to acquire the photometric measurements in different magnitude regimes (see Rowell et al. 2021; Riello et al. 2021, for further details). In particular, the decrease in the detection efficiency around $G = 11.5$ mag could be due to the mismatch between the rate at which charges are transferred in the along-scan direction and the rate at which the stellar image drifts across the CCD. This mismatch distorts the *Gaia* PSF, but currently this is not taken into account in the PSF modelling, which leads to an increase in the photometric uncertainty in that magnitude range (Rowell et al. 2021). In *Gaia* DR4, the PSF modelling will take this effect into account and a better photometry is expected.

We note that, while the detection efficiency obtained in DR3 is in general higher than in DR2, at $G \simeq 11.5$ the two releases have about the same efficiency. This is discussed in detail in Appendix C.

In the top panel of Fig. B.3, we show how the detection efficiency depends on β down to the limiting magnitude G=15. In this case, $e(\beta) \simeq 15\%$ at ecliptic poles and about 28 % around $\beta = \pm 45$ deg. If we restrict the analysis to the magnitude range (13:15) (where the average photometric uncertainty is about 0.001 *mag*), we find that $e(\beta)$ is about 20 % at the ecliptic poles and about 33 % around $\beta = \pm 45$ deg.

B.2. Contamination

As discussed in Lanzafame et al. (2018) for the gdr2_rotmod catalogue, the possible contaminants of the gdr3_rotmod could be given by pulsating variables for which parallaxes were not properly estimated or by short-period eclipsing binaries. The sources currently listed in the gdr3_rotmod do not overlap with any of the pulsating variables or eclipsing binaries reported in

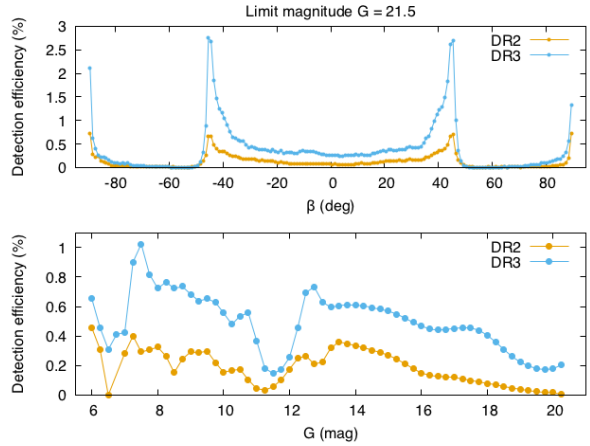

Fig. B.1. Detection efficiency for DR3 and DR2 catalogues. Top panel. Detection efficiency vs. ecliptic latitude for DR3 (cyan line) and DR2 (yellow line). Bottom panel. Detection efficiency vs. G magnitude.

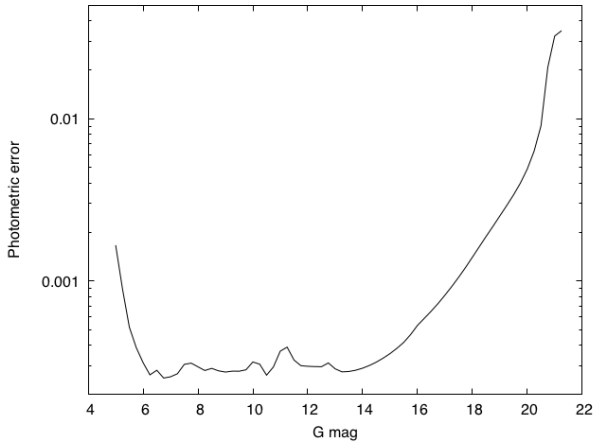

Fig. B.2. Average photometric error vs. the stellar G magnitude.

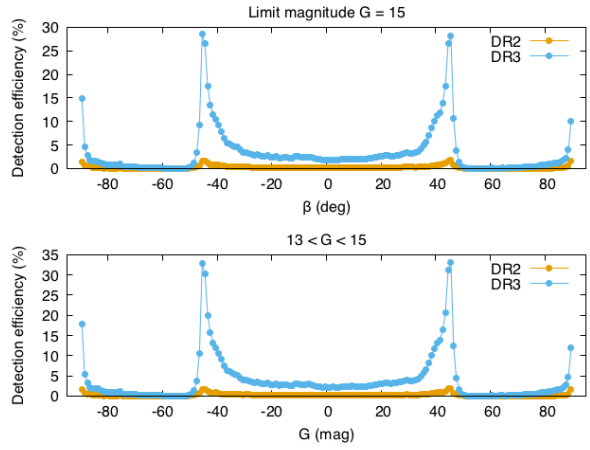

Fig. B.3. Detection efficiency for DR3 and DR2 catalogues in limited magnitude ranges. Top panel: Detection efficiency vs. ecliptic latitude down to the limit magnitude $G = 15$. Bottom panel: Detection efficiency vs. ecliptic latitude in the G range (13,15.5).

the *Gaia* DR3 (Eyer et al. 2023). Moreover all the quality criteria applied to filter out the data and described in Sections 2.6 and A should remove pulsating variables whose light curves significantly deviate from the typical sinusoidal shape of rotational modulation variables (see the discussion in Lanzafame et al. 2018). Therefore, we expect that the contamination level of our catalogue is very low.

An accurate estimate of this contamination will require knowledge of the variability classes for a subset of stars statistically representative of the full catalogue. This is not possible because the currently available surveys of variable stars do not have the same properties as *Gaia* in terms of magnitude range and sky coverage. Moreover, the variable classification performed by these surveys is in turn subject to some uncertainty. Nevertheless, the comparison between the `gdr3_rotmod` catalogue and the available surveys can be useful to estimate at least the order of magnitude of the contamination level.

In order to obtain a rough estimate of the contamination level, we therefore cross-matched the `gdr3_rotmod` with the ZTF (Zwicky Transient Facility) catalogue of periodic variable stars (Chen et al. 2020, hereafter `ztf_var`) and with the ASAS-SN (All Sky Automated Survey for SuperNovae) catalogue of variable stars (herafter `asas-sn_var` Jayasinghe et al. 2019).

B.2.1. Comparison with ZTF classification

The `ztf_var` was obtained by analysing the photometric time series acquired with the ZTF telescope (Masci et al. 2019) in the Sloan g and r photometric bands during a 470 d survey. This survey (Graham et al. 2019) covers the northern sky hemisphere and has a limiting magnitude $r \simeq 20.6$, which is similar to that of the `gdr3_rotmod`. The `ztf_var` catalogue lists 781 602 variable stars that have been classified in 11 different types. About 150 000 of these stars have been classified by the ZTF pipeline as rotational variables. The cross-match between the `gdr3_rotmod` and the `ztf_var` catalogue returned 6513 sources[7] attributed by the pipeline to six different classes. Table B.1 reports the number and the percentage of the sources attributed to each class. Most of the sources (94.6%) were classified as rotational modulation variables according to the `gdr3_rotmod`. Indeed, 91.7% of variables were classified as BY Dra and 2.6% as RS CVn variables. About 5 % of the sources were classified as EW or EA binary systems. A small percentage (0.4 %) was attributed to the semi-regular class and finally only two stars from the sample were classified as pulsating variables.

B.2.2. Comparison with ASAS-SN classification

The `asas-sn_var` catalogue was obtained by analysing the $V - band$ photometric time series collected by the ASAS-SN survey (Shappee et al. 2014), which covers the entire sky down to the limit magnitude $V \simeq 17$. The time series analysed to produce the catalogue were acquired between 2013 and 2018 and are characterised by a 2d or 3d cadence. The catalogue reports 687 695 sources and supplies a classification for 662 627 of them (the remaining are classified as uncertain variables). About 90 000 sources of the catalogue are classified as rotational variables.

The sources common to the `gdr3_rotmod` and `asas-sn_var` catalogues number 1 054. Table B.2 reports the classes attributed by Jayasinghe et al. (2019) to the 1054

[7] The reader may wonder why a large fraction of the rotational variables reported in the `ztf_var` and in the `asas-sn_var` catalogues are missing in the `gdr3_rotmod`. This actually happens because most of these variables fall in sky regions poorly sampled by *Gaia* and therefore do not satisfy the selection criteria to be processed by the pipeline (see Sect. 2.1).

Table B.1. Variability classes attributed by Chen et al. (2020) to the 6 513 stars common to the `ztf_var` and `dr3_rotmod` catalogue.

Class	# sources	%
BY Dra	5 978	91.7 %
RS CVn	170	2.6%
EW	307	4.7%
EA	32	0.5%
SR	28	0.4%
DSCT	1	0.01%
RR	1	0.01%

Table B.2. Variability classes attributed by Jayasinghe et al. (2019) to the 1 054 stars common to the `asas-sn_var` and `dr3_rotmod` catalogue.

Class	# sources	%
ROT	851	80.7 %
EA	118	11.2%
EW	10	0.9%
EB	10	0.9%
RRAB	2	0.2%
RRC	2	0.2%
VAR	34	3.2%
YSO	7	0.66%
UV Cet	20	1.8%

stars[7] and their relative percentages. Of the 1 054 stars, 80.25% were classified as rotational variables, 13 % of the sources were classified as binary systems (most part of which are of EA type), 6 sources were classified as young stellar objects (YSOs), 7 sources as pulsators, 20 sources as UV Cet stars, and finally 22 sources, corresponding to the tag VAR, have an uncertain classification. We note that the UV Cet variables are magnetically active stars whose light curves are mainly characterised by outbursts due to energetic flare events. However, these stars can also exhibit a rotational modulation signal, and so cannot be regarded as contaminants. In the same way, the stars with uncertain classification could be magnetically active stars for which the ASAS-SN pipeline was not able to detect the rotational signature. We note that for these uncertain stars, the `asas-sn_var` catalogue usually reports a period of longer than 100 d, which could be ascribed to short-term magnetic activity cycles like those detected by Distefano et al. (2017). In these cases, the evolution of MARs could have masked the rotational modulation signal and the asas-sn pipeline could have failed in the detection of the rotation period because it does not adopt a segmentation strategy.

Overall, the comparison with the `ztf_var` and `asassn_var` catalogues assesses the contamination level of the `gdr3_rotmod` to be between the 6% and 14 %. The main contaminants of the catalogue are binary systems, whereas pulsators can be considered negligible.

B.3. Rate of correct periods detection

The filtering procedures described in Sects. A and 2.6 leads us to reject about 40% of the sources initially detected by the pipeline. We are therefore confident that most of the sources with spurious periods are removed from the final catalogue. Nevertheless, a small percentage of aliases and artificial periods introduced by the peculiar Gaia sampling or by photometric artefacts can still be present in the released data. An estimate of the rate of correct period detection cannot be easily obtained because catalogues from other surveys are in turn affected by sampling or calibration issues. However, a comparison with other surveys can still be useful for assessing a lower limit for such a rate.

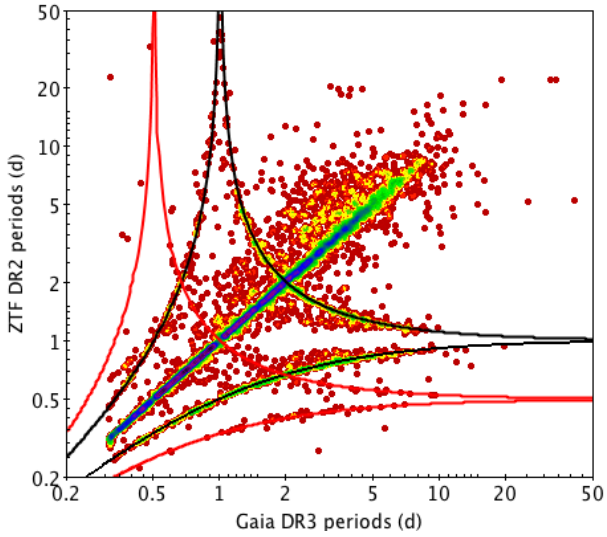

Fig. B.4. Comparison between the periods detected in the ZTF survey and those in Gaia DR3. The points are colour coded according to their density. The red and dark lines mark the loci defined by Eq. B.5 for n=1 and n=2, respectively.

B.3.1. Comparison with ZTF periods

In Fig. B.4, we show a comparison between the `gdr3_rotmod` and the `ztf_var` periods for all the variables common to the two surveys and classified as BY Dra or RS CVn in the ZTF catalogue. Two periods can be considered in agreement if they differ by less than 20%. This tolerance value has been set from the work of Distefano et al. (2017). In fact, this latter study demonstrates that the stellar rotation period detected for a given star can noticeably change from one observational season to another. The authors attribute these variations to the combined effect of Scwabe-like cycles and surface differential rotation. However, Basri & Shah (2020) pointed out that these variations are more likely due to the appearing and disappearing of MARs at random longitudes. We find that the Gaia and ZTF periods are in agreement in 68% of the sample.

In 9% of cases, the Gaia periods are in agreement with either half or twice the period detected by ZTF. These mismatches are very common when dealing with rotational variables because, according to the longitudinal distribution of MARs, the light curves of these stars can exhibit one or two minima per rotation cycle. The long-term monitoring of solar-like variables performed by the *Kepler* mission showed that the light curve of a given star is often characterised by a continuous switch between a single-dip and a double-dip configuration (see Basri & Nguyen (2018) and Basri & Shah (2020) for details). These mismatches therefore reflect an intrinsic property of the rotational variables, and affect the whole rotation period distribution collected by different surveys.

Table B.3. Comparison between `gdr3_rotmod`, `asas-sn_var`, and `HATNet` periods for the five stars encircled in Fig. B.5.

DR3 Id	HATNet Id	P_{asas} (d)	P_{HATNet}(d)	P_{gaia} (d)
417488275919654656	HAT-087-000 8179	426.13	8.88	8.68
391326775424158848	HAT-087-002 4596	544.9	0.75	0.75
436212958941004032	HAT-091-001 0879	494.5	2.47	2.45
219651916081208320	HAT-169-000 7364	285.15	3.41	3.38
1927973852394016384	HAT-205-002 4012	716.4	2.34	2.34

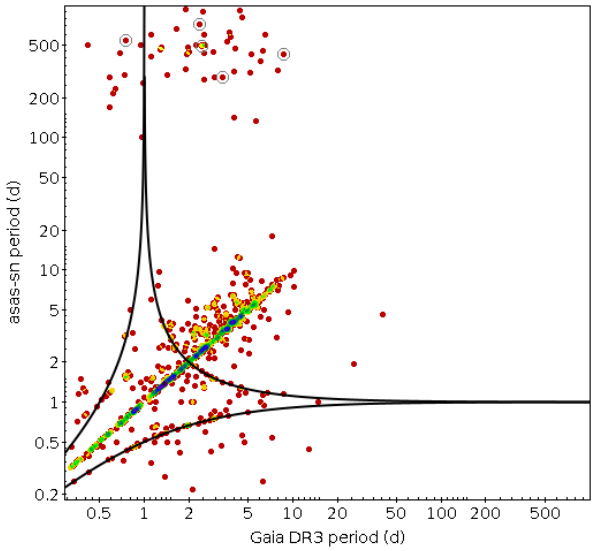

Fig. B.5. Comparison between the periods detected in the asas-sn survey and those in Gaia DR3. The points are colour coded according to their density. The continuous dark line depicts the loci defined by Eq. B.6 for n=1. The sources surrounded with the dark circles are also reported in Hartman et al. (2011) with periods very similar to those reported in Gaia DR3 (see Table B.3).

In 14% of the sample, the ZTF periods and Gaia periods are connected by a relationship of the following type:

$$\frac{1}{P_{ztf}} = \left| \pm \frac{1}{P_{gaia}} \pm \frac{1}{nT_{ztf}} \right|, \tag{B.5}$$

where $T_{ztf} = 1d$ is the typical cadence of the ZTF survey and $n = 1, 2$. In these cases, the periods detected by ZTF are aliases, that is spurious periods generated by the interference between the sampling frequency and the rotational frequency (correctly detected by Gaia). We note that, in principle, *Gaia* periods can also be affected by aliasing, but Eyer & Mignard (2005) showed that the irregularity of the *Gaia* scanning law limits the build up of coherent signatures at regularly spaced frequencies. Finally, in 9% of the sample, there is no evident relationship between the periods detected by Gaia and those detected by ZTF.

B.3.2. Comparison with asas-sn periods.

In Fig. B.5, we show the comparison between the `asas-sn_var` and the `gdr3_rotmod` periods for all the stars common to the two surveys and labelled as rotational variables in the `asas-sn_var` catalogue. The *Gaia* periods agree with the asas-sn periods in 59% of the sample and agree with half or twice the asas-sn periods in 12% of the cases. In about 8% of the cases, the asas-sn periods fall in the locus defined by

$$\frac{1}{P_{asas}} = \left| \pm \frac{1}{P_{gaia}} \pm \frac{1}{nT_{asas}} \right|, \tag{B.6}$$

with n=1 and $T_{asas} = 1d$. Finally, in about 20 % of the sample, the *Gaia* periods do not match at all with the asas-sn periods. In conclusion, the comparison between *Gaia* and the two surveys allows us to assess the rate of correct period detection at between 80% and 91%.

Strictly, these percentages are lower limits for the rate of correct detection because the asas-sn and the ZTF pipelines do not take into account the intrinsic evolution of active regions and do not apply any segmentation strategy. Therefore, the two surveys, in some cases, are not able to pick the rotational modulation signal. This is the case, for instance, for the five stars encircled in Fig. B.5 for which Hartman et al. (2011) reported rotation periods in good agreement with those found in the present work.

Table B.3 reports the periods listed in the `gdr3_rotmod` and in the `asas-sn_var` catalogues and those found Hartman et al. (2011) for the five stars. The longer periods of the order of 200-500 d reported in the `asas-sn_var` are probably due to long-term variations of the stellar magnetic activity that in some cases can produce quasi-periodic signals (see e.g. Oláh et al. 2016; Distefano et al. 2017, and reference therein).

Appendix C: Comparison between DR3 and DR2 results

The detection efficiency of DR3 is about three times that of DR2 (see Fig. B.1). However, many of the rotational modulation variables reported in the `gdr2_rotmod` catalogue are not listed in the `gdr3_rotmod`. The `gdr2_rotmod` catalogue reports 147 774 rotational modulation variables. The sources in common between the two catalogues number only 34 747. Therefore, 113 027 stars listed in DR2 as rotational modulation variables do not appear in the DR3 catalogue.

The missing stars are not to be considered erroneous DR2 detections. Indeed, the contamination level and the rate of correct period detections in this sample of stars are of the same order as those seen in the `gdr3_rotmod` catalogue (see Appendix C for details). These stars do not figure in the `gdr3_rotmod` catalogue simply for technical reasons that are reported and discussed in detail below:

- `selection issues`: 8% of the missing stars have different parallaxes and/or mean magnitudes from those reported in DR2 and no longer satisfy one of the first two criteria reported in Sect. 2.1. In Fig. C.1, we report the location of these rejected stars. Apart from a few hundred stars that now lie significantly far from the main sequence region, most of them (about 90 %) still fall in the HR region used for selection and have been rejected only because the relative error in parallax is greater than 20 %;

– `segmentation issues`: as stated in Sect. 2.1, the photometric time series processed by the CU7 pipeline are obtained by applying a chain of several operators to the G raw time series. This chain of operators, described in detail in Rimoldini et al. (2022), is slightly different from that employed in DR2 and adopts a different strategy to identify and remove possible outliers. In addition, the pipeline used to detect rotational modulation adopts two further filters described in Sect. 2.3 that in DR2 were deactivated. The combined action of CU7 operators and of the pipeline filters reduces the average number of points per segment and in turn the number of segments available for the analysis. This issue affects 38% of the missing sources. These sources were discarded for two different reasons: (1) the period search was not performed because the time-series segments (after filtering) have an insufficient number of points to be processed (we reiterate that the minimum number of points required to perform the period search is 12); and (2) the Gaia DR2 rotation period was correctly detected but only in one segment (a star is released if the same period is detected in at least two segments.
We emphasise that the G measurements removed by the cleaning procedure described in Sect. 2.3 are not necessarily 'bad' points. For instance, the pipeline removes all the G measurements for which the $G_{\rm BP}$ or $G_{\rm RP}$ counterparts are missing. This is done in order to assure that the activity indexes computed in the different bands are inferred from the same set of transits but unfortunately results in the rejection of G measurements suitable for the period search. The pipeline removes also G measurements corresponding to transits in which the $G_{\rm BP} - G_{\rm RP}$ colour significantly deviates from the average $G_{\rm BP} - G_{\rm RP}$ and flags them as candidate flares. A posteriori analysis of the data showed that in the most part of these transits the anomalous colour is due to a bad measurement in the $G_{\rm BP}$ or $G_{\rm RP}$ band and does not correspond to an enhancement of the G flux. In DR4, the pipeline will be refined to improve the cleaning strategy and minimise the loss of valid measurements;
– `quality criteria issues`: 37% of the missing variables were correctly detected by the pipeline but were discarded because they did not match one of the quality criteria described in Sect. 2.3. These sources can be divided into three groups: (1) sources (13% of the total sample) for which the period is correctly detected in all segments but the phase coverage criteria are not satisfied in any of the segments; (2) sources (14% of the total sample) for which the period is correctly detected in all segments but the FAP exceeds the fixed threshold value; and (3) sources (10% of the total sample) for which the period is correctly detected in all the segments but the reduced chi-square $\tilde{\chi}^2$ associated with the model-fitting is above the fixed threshold.
Also in this case, we emphasise that these discarded sources are not necessarily bad. Indeed the cleaning procedure described in the above issue reduces the number of points per segment and makes the match of the phase coverage criteria more difficult than in DR2.
The reduced number of points per segments has also the effect to enhance the FAP value associated to a given period. Indeed the FAP formula employed here (Baluev 2008) is dependent on the number of points used to perform the period search.
Finally, the analysis of the data suggests that the sources with a high $\tilde{\chi}^2$ value are probably stars for which the photometric errors have been under-estimated. This latter problem mainly affects sources brighter than 13 mag. In Fig. C.2, we present the $\tilde{\chi}^2$ values computed by the pipeline according to Eq. 8 vs. the stellar magnitude. The figure shows that the $\tilde{\chi}^2$ values tend to decrease with increasing magnitude. This trend suggests that photometric errors could be underestimated for the brightest sources and overestimated for the faintest sources. Unfortunately, this quality criterion causes the rejection of well-known bright stars such as *AB Dor* and *BD* − 072388;
– `unknown issues`: for 8% of the missing sources we were not able to find a satisfying explanation. Analysis of the data suggests that in some of these stars the rotational modulation signal could have been distorted by the strategy of the photometric calibration employed in DR3. In Fig. C.3 and C.4, we show comparisons of the DR3 and DR2 G time series for two different stars. The two sets of measurements correspond to the same set of transits but assume different values because of the different calibration strategies. In both figures, the two time series are folded according to the rotation periods detected in DR2. In the first star, the DR3 photometric calibration induced a 0.02 mag offset between the two time series but preserved the shape of the rotational modulation signal. In the second case, the shape of the rotational signal is well defined in DR2 time series and strongly distorted in the DR3 one. (We note that the periods detected in the DR2 time series for the two stars are highly reliable because they were detected also by Jayasinghe et al. (2019) in the asas-sn time series.)

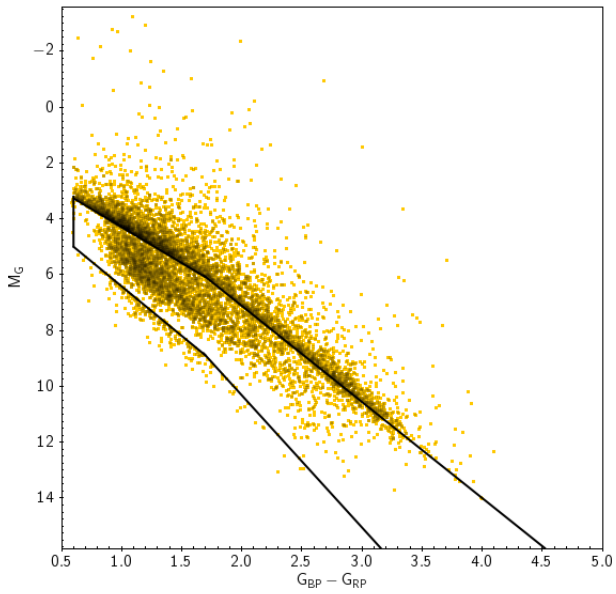

Fig. C.1. Location of the DR2 variables excluded by the DR3 selection criteria in the M_G *vs.* ($G_{\rm BP} - G$RP) diagram.

C.1. Quality assessment of the DR2 excluded sources

We assessed the quality of the `gdr2_rotmod` sources that do not appear in the `gdr3_rotmod` catalogue. Hereafter, this sample of stars is referred to as the `gdr2_excluded` list. We evaluated the contamination level and the rate of correct period detection of this list by employing the same procedure adopted in Sects. B.2 and B.3 for the DR3 variables. We cross-matched the `gdr2_excluded` list with the `ztf_var` and the `asas-sn_var` catalogues. We found 1 692 sources in common

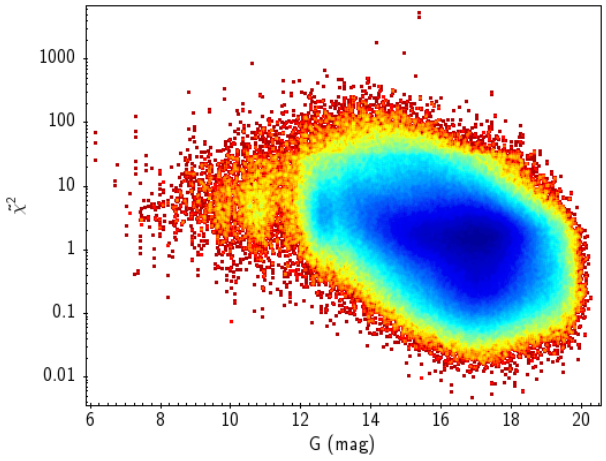

Fig. C.2. Distribution of the $\tilde{\chi}^2$ values obtained by fitting the G photometric time series to the model given by Eq. 3.

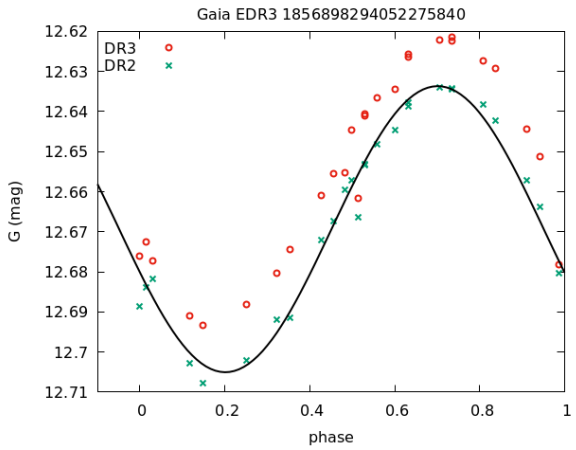

Fig. C.3. Comparison between the DR3 (red circles) and DR2 (green crosses) photometric data for the source `Gaia DR3 1856898294052275840`. The two sets of photometric data correspond to the same set of transits but have different values because of the different calibration strategies applied in DR3 and DR2. In this case, the DR3 data have just a small offset with respect to the DR2 data and the shape of the rotational modulation signal is preserved by the DR3 calibration.

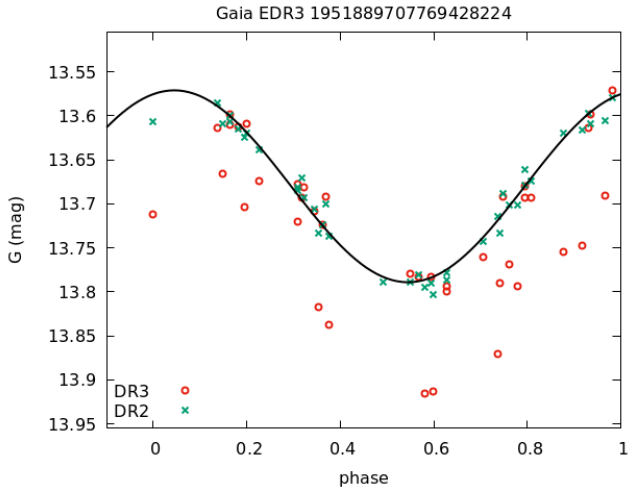

Fig. C.4. Same as Fig. C.3 but for the source *Gaia DR3 1951889707769428224*. In this case, the DR3 calibration distorts the rotational modulation signal detected in the DR2 photometry.

Table C.1. Variability classes attributed by Chen et al. (2020) to the 1 692 stars common to the `ztf_var` and `gdr2_excluded` list.

Class	# sources	%
BY Dra	1 410	83.3 %
RS CVn	26	1.5%
EW	229	13.5%
EA	12	0.7%
SR	10	0.6%
DSCT	4	0.2%
RR	1	0.06%

Table C.2. Variability classes attributed by Jayasinghe et al. (2019) to the 335 stars common to the `asas-sn_var` and the `dr2_rotmod` excluded list.

Class	# sources	%
ROT	256	75.7 %
EA	13	3.85%
EW	49	14.5%
EB	4	1.2%
RRAB	0	0%
RRC	0	0%
VAR	9	2.6%
YSO	0	0%
UV Cet	4	1.2%

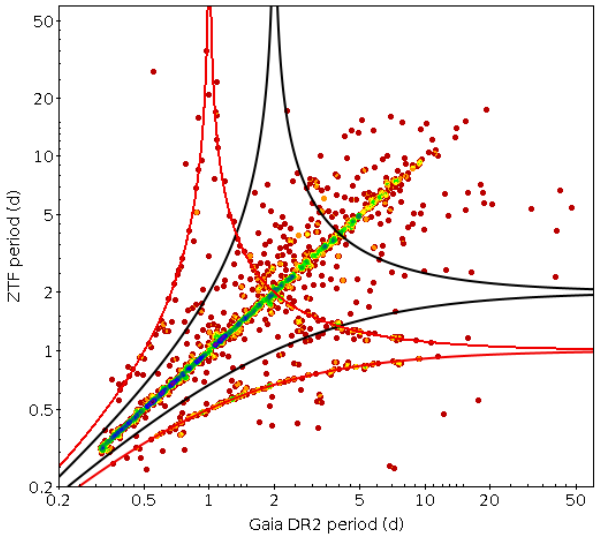

Fig. C.5. Comparison between the periods detected in the ZTF survey and those detected in the *Gaia* DR2 release and excluded in the *Gaia* DR3 release. The points are colour coded according to density.

with the `ztf_var` catalogue and 335 with the `asas-sn_var`. We reported the classes attributed to these stars by the ZTF and the asas-sn classifiers in Tables C.1 and C.2, respectively. For each class, we reported the number and the percentage of the stars assigned to it. The two tables show that the contamination level of the `gdr2_excluded` can be assessed at between 15% and 17% and that, as in the case of the `gdr3_rotmod` catalogue, the main source of contamination is binary systems. We compared the periods of the `gdr2_excluded` list with those reported in the `ztf_var` and in the `asas-sn` catalogues in order to

evaluate the rate of correct period detections. In Fig. C.5 and C.6, we reported a comparison of the `gdr2_excluded` periods with the ZTF and the asas-sn periods, respectively. In the sample of stars in common with ZTF, 65% of the `gdr2_excluded` periods differ by less than 20% from the ZTF periods, 12% of the periods are compatible with half or twice the ZTF periods, and the 17% fall in the loci of aliases defined by Eqs. B.5. In the sample of stars in common with asas-sn, 58% of the `gdr2_excluded` periods differ by less than 20% from the asas-sn periods, 13% of the periods are compatible with half or twice the asas-sn periods, and 4% fall in the loci of aliases defined by Eqs. B.6.

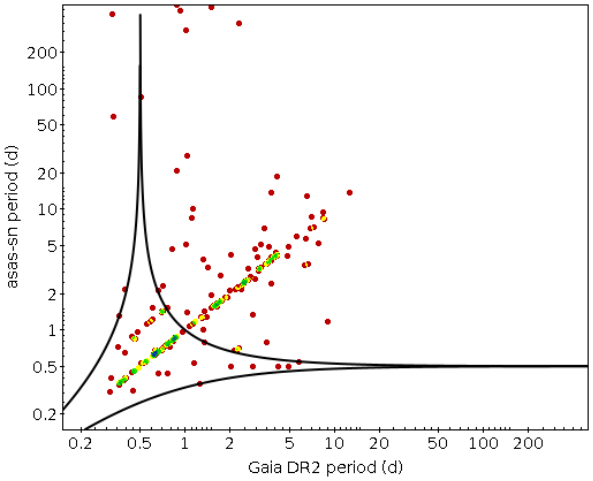

Fig. C.6. Comparison between the periods detected in the asas-sn survey and those detected in the *Gaia* DR2 release and excluded in the *Gaia* DR3 release. The points are color-coded according to density.