A&A, 680, A36 (2023)
https://doi.org/10.1051/0004-6361/202347287

Astronomy & Astrophysics

Gaia Focused Product Release: Radial velocity time series of long-period variables

Gaia Collaboration: M. Trabucchi[1,2,⋆], N. Mowlavi[2], T. Lebzelter[3], I. Lecoeur-Taibi[4], M. Audard[2,4], L. Eyer[2], P. García-Lario[5], P. Gavras[6], B. Holl[2,4], G. Jevardat de Fombelle[2], K. Nienartowicz[7,4], L. Rimoldini[4], P. Sartoretti[8], R. Blomme[9], Y. Frémat[9], O. Marchal[10], Y. Damerdji[11,12], A. G. A. Brown[13], A. Guerrier[14], P. Panuzzo[8], D. Katz[8], G. M. Seabroke[15], K. Benson[15], R. Haigron[8], M. Smith[15], A. Lobel[9], A. Vallenari[16], T. Prusti[17], J. H. J. de Bruijne[17], F. Arenou[8], C. Babusiaux[18], A. Barbier[14], M. Biermann[19], O. L. Creevey[20], C. Ducourant[21], D. W. Evans[22], R. Guerra[5], A. Hutton[23], C. Jordi[24,25,26], S. A. Klioner[27], U. Lammers[5], L. Lindegren[28], X. Luri[24,25,26], F. Mignard[20], S. Randich[29], R. Smiljanic[30], P. Tanga[20], N. A. Walton[22], C. A. L. Bailer-Jones[31], U. Bastian[19], M. Cropper[15], R. Drimmel[32], M. G. Lattanzi[32,33], C. Soubiran[21], F. van Leeuwen[22], J. Bakker[5], J. Castañeda[34,24,26], F. De Angeli[22], C. Fabricius[26,24,25], M. Fouesneau[31], L. Galluccio[20], E. Masana[26,24,25], R. Messineo[35], C. Nicolas[14], F. Pailler[14], F. Riclet[14], W. Roux[14], R. Sordo[16], F. Thévenin[20], G. Gracia-Abril[36,19], J. Portell[24,25,26], D. Teyssier[37], M. Altmann[19,38], J. Berthier[39], P. W. Burgess[22], D. Busonero[32], G. Busso[22], H. Cánovas[37], B. Carry[20], N. Cheek[40], G. Clementini[41], M. Davidson[42], P. de Teodoro[5], L. Delchambre[11], A. Dell'Oro[29], E. Fraile Garcia[6], D. Garabato[43], N. Garralda Torres[44], N. C. Hambly[42], D. L. Harrison[22,45], D. Hatzidimitriou[46], J. Hernández[5], S. T. Hodgkin[22], S. Jamal[31], S. Jordan[19], A. Krone-Martins[47,48], A. C. Lanzafame[49,50], W. Löffler[19], A. Lorca[23], P. M. Marrese[51,52], A. Moitinho[48], K. Muinonen[53,54], M. Nuñez Campos[23], I. Oreshina-Slezak[20], P. Osborne[22], E. Pancino[29,52], T. Pauwels[9], A. Recio-Blanco[20], M. Riello[22], A. C. Robin[55], T. Roegiers[56], L. M. Sarro[57], M. Schultheis[20], C. Siopis[58], A. Sozzetti[32], E. Utrilla[23], M. van Leeuwen[22], K. Weingrill[59], U. Abbas[32], P. Ábrahám[60,61], A. Abreu Aramburu[44], C. Aerts[62,63,31], G. Altavilla[51,52], M. A. Álvarez[43], J. Alves[3], F. Anders[24,25,26], R. I. Anderson[64], T. Antoja[24,25,26], D. Baines[65], S. G. Baker[15], Z. Balog[19,31], C. Barache[38], D. Barbato[2,32], M. Barros[66], M. A. Barstow[67], S. Bartolomé[26,24,25], D. Bashi[68,69], N. Bauchet[8], N. Baudeau[70], U. Becciani[49], L. R. Bedin[16], I. Bellas-Velidis[71], M. Bellazzini[41], W. Beordo[32,33], A. Berihuete[72], M. Bernet[24,25,26], C. Bertolotto[35], S. Bertone[32], L. Bianchi[73], A. Binnenfeld[74], A. Blazere[75], T. Boch[10], A. Bombrun[76], S. Bouquillon[38,77], A. Bragaglia[41], J. Braine[21], L. Bramante[35], E. Breedt[22], A. Bressan[78], N. Brouillet[21], E. Brugaletta[49], B. Bucciarelli[32,33], A. G. Butkevich[32], R. Buzzi[32], E. Caffau[8], R. Cancelliere[79], S. Cannizzo[80], R. Carballo[81], T. Carlucci[38], M. I. Carnerero[32], J. M. Carrasco[26,24,25], J. Carretero[82,83], S. Carton[80], L. Casamiquela[21,8], M. Castellani[51], A. Castro-Ginard[13], V. Cesare[49], P. Charlot[21], L. Chemin[84], V. Chiaramida[35], A. Chiavassa[20], N. Chornay[22,4], R. Collins[42], G. Contursi[20], W. J. Cooper[85,32], T. Cornez[80], M. Crosta[32,86], C. Crowley[76], C. Dafonte[43], M. David[87], P. de Laverny[20], F. De Luise[88], R. De March[35], J. De Ridder[62], R. de Souza[89], A. de Torres[76], E. F. del Peloso[19], M. Delbo[20], A. Delgado[6], T. E. Dharmawardena[31], S. Diakite[90], C. Diener[22], E. Distefano[49], C. Dolding[15], K. Dsilva[58], J. Durán[6], H. Enke[59], P. Esquej[6], C. Fabre[75], M. Fabrizio[51,52], S. Faigler[68], M. Fatović[91], G. Fedorets[53,92], J. Fernández-Hernández[6], P. Fernique[10], F. Figueras[24,25,26], Y. Fournier[59], C. Fouron[70], M. Gai[32], M. Galinier[20], A. Garcia-Gutierrez[26,24,25], M. García-Torres[93], A. Garofalo[41], E. Gerlach[27], R. Geyer[27], P. Giacobbe[32], G. Gilmore[22,94], S. Girona[95], G. Giuffrida[51], R. Gomel[68], A. Gomez[43], J. González-Núñez[96], I. González-Santamaría[43], E. Gosset[11,97], M. Granvik[53,98], V. Gregori Barrera[26,24,25], R. Gutiérrez-Sánchez[37], M. Haywood[8], A. Helmer[80], A. Helmi[99], K. Henares[65], S. L. Hidalgo[100,101], T. Hilger[27], D. Hobbs[28], C. Hottier[8], H. E. Huckle[15], M. Jabłońska[102,103], F. Jansen[104], Ó. Jiménez-Arranz[24,25,26], J. Juaristi Campillo[19], S. Khanna[32,99], G. Kordopatis[20], Á. Kóspál[60,31,61], Z. Kostrzewa-Rutkowska[13], M. Kun[60], S. Lambert[38], A. F. Lanza[49], J.-F. Le Campion[21], Y. Lebreton[105,106], S. Leccia[107], G. Lecoutre[55], S. Liao[108,32,109], L. Liberato[20,110], E. Licata[32], H. E. P. Lindstrøm[32,111,112], T. A. Lister[113], E. Livanou[46], C. Loup[10], L. Mahy[9], R. G. Mann[42], M. Manteiga[114], J. M. Marchant[115], M. Marconi[107], D. Marín Pina[24,25,26], S. Marinoni[51,52], D. J. Marshall[116], J. Martín Lozano[40],

⋆ Corresponding author: M. Trabucchi, e-mail: michele.trabucchi@unige.ch

J. M. Martín-Fleitas[23], G. Marton[60], N. Mary[80], A. Masip[26,24,25], D. Massari[41], A. Mastrobuono-Battisti[8], T. Mazeh[68], P. J. McMillan[28], J. Meichsner[27], S. Messina[49], D. Michalik[17], N. R. Millar[22], A. Mints[59], D. Molina[25,24,26], R. Molinaro[107], L. Molnár[60,117,61], G. Monari[10], M. Monguió[24,25,26], P. Montegriffo[41], A. Montero[40], R. Mor[118,25,26], A. Mora[23], R. Morbidelli[32], T. Morel[11], D. Morris[42], D. Munoz[80], T. Muraveva[41], C. P. Murphy[5], I. Musella[107], Z. Nagy[60], S. Nieto[6], L. Noval[80], A. Ogden[22], C. Ordenovic[20], C. Pagani[119], I. Pagano[49], L. Palaversa[91], P. A. Palicio[20], L. Pallas-Quintela[43], A. Panahi[68], C. Panem[14], S. Payne-Wardenaar[19], L. Pegoraro[14], A. Penttilä[53], P. Pesciullesi[6], A. M. Piersimoni[88], M. Pinamonti[32], F.-X. Pineau[10], E. Plachy[60,117,61], G. Plum[8], E. Poggio[20,32], D. Pourbaix[†,58,97], A. Prša[120], L. Pulone[51], E. Racero[40,121], M. Rainer[29,122], C. M. Raiteri[32], P. Ramos[123,24,26], M. Ramos-Lerate[37], M. Ratajczak[102], P. Re Fiorentin[32], S. Regibo[62], C. Reylé[55], V. Ripepi[107], A. Riva[32], H.-W. Rix[31], G. Rixon[22], N. Robichon[8], C. Robin[80], M. Romero-Gómez[24,25,26], N. Rowell[42], F. Royer[8], D. Ruz Mieres[22], K. A. Rybicki[124], G. Sadowski[58], A. Sáez Núñez[26,24,25], A. Sagristà Sellés[19], J. Sahlmann[6], V. Sanchez Gimenez[26,24,25], N. Sanna[29], R. Santoveña[43], M. Sarasso[32], C. Sarrate Riera[34,24,26], E. Sciacca[49], J. C. Segovia[40], D. Ségransan[2], S. Shahaf[124], A. Siebert[10,125], L. Siltala[53], E. Slezak[20], R. L. Smart[32,85], O. N. Snaith[8,126], E. Solano[127], F. Solitro[35], D. Souami[105,128], J. Souchay[38], L. Spina[16], E. Spitoni[20,129], F. Spoto[130], L. A. Squillante[35], I. A. Steele[115], H. Steidelmüller[27], J. Surdej[11], L. Szabados[60], F. Taris[38], M. B. Taylor[131], R. Teixeira[89], K. Tisanić[91], L. Tolomei[35], F. Torra[34,24,26], G. Torralba Elipe[43,132,133], M. Tsantaki[29], A. Ulla[134,135], N. Unger[2], O. Vanel[8], A. Vecchiato[32], D. Vicente[95], S. Voutsinas[42], M. Weiler[26,24,25], Ł. Wyrzykowski[102], H. Zhao[20,136], J. Zorec[137], T. Zwitter[138], L. Balaguer-Nunez[26,24,25], N. Leclerc[8], S. Morgenthaler[139], G. Robert[80], and S. Zucker[74]

(Affiliations can be found after the references)

Received 26 June 2023 / Accepted 11 August 2023

ABSTRACT

Context. The third *Gaia* Data Release (DR3) provided photometric time series of more than 2 million long-period variable (LPV) candidates. Anticipating the publication of full radial-velocity data planned with Data Release 4, this Focused Product Release (FPR) provides radial-velocity time series for a selection of LPV candidates with high-quality observations.

Aims. We describe the production and content of the *Gaia* catalog of LPV radial-velocity time series, and the methods used to compute the variability parameters published as part of the *Gaia* FPR.

Methods. Starting from the DR3 catalog of LPV candidates, we applied several filters to construct a sample of sources with high-quality radial-velocity measurements. We modeled their radial-velocity and photometric time series to derive their periods and amplitudes, and further refined the sample by requiring compatibility between the radial-velocity period and at least one of the G, $G_{\rm BP}$, or $G_{\rm RP}$ photometric periods.

Results. The catalog includes radial-velocity time series and variability parameters for 9614 sources in the magnitude range $6 \lesssim G/\text{mag} \lesssim 14$, including a flagged top-quality subsample of 6093 stars whose radial-velocity periods are fully compatible with the values derived from the G, $G_{\rm BP}$, and $G_{\rm RP}$ photometric time series. The radial-velocity time series contain a mean of 24 measurements per source taken unevenly over a duration of about three years. We identify the great majority of the sources (88%) as genuine LPV candidates, with about half of them showing a pulsation period and the other half displaying a long secondary period. The remaining 12% of the catalog consists of candidate ellipsoidal binaries. Quality checks against radial velocities available in the literature show excellent agreement. We provide some illustrative examples and cautionary remarks.

Conclusions. The publication of radial-velocity time series for almost ten thousand LPV candidates constitutes, by far, the largest such database available to date in the literature. The availability of simultaneous photometric measurements gives a unique added value to the *Gaia* catalog.

Key words. stars: AGB and post-AGB – stars: variables: general – stars: carbon – methods: data analysis – catalogs – techniques: radial velocities

1. Introduction

Evolved stars of low and intermediate mass show various kinds of light variability summarized in the class of long period variables (LPVs). Within this class, there are radially pulsating objects showing small to large light amplitudes in various pulsation modes and with various degrees of periodicity, but also stars whose light variability is due to the presence of a binary or due to eclipses of orbiting dust clouds. For disentangling the various causes for variability in these stars, sometimes even occurring in combinations, contemporaneous monitoring of radial-velocity (RV) variations has proven to be a useful approach.

Early measurements of RV variations in LPVs date back to the 1920s (Joy 1926). It was noted already then that emission and absorption lines in Miras show different kinds of velocity variations. This was supported by several further studies, all using lines in the blue part of the spectrum, but the observed variability pattern did not allow for a conclusive description of the pulsation in these stars (Joy 1954; Reid & Dickinson 1976). A major step forward was achieved by the first monitoring of RV changes in the near-infrared. The landmark paper by Hinkle et al. (1982) revealed the photospheric kinematics for the Mira χ Cyg, allowing for components related to stellar pulsation and to mass outflow to be identified, respectively. Line doubling of high-excitation CO lines was found near light maximum and, together with the appearance of hydrogen emission lines at those phases, interpreted as a trace of shock fronts. Combining velocity data from the violet to the radio regime allowed for a stratigraphy of a Mira's atmosphere to be constructed out to its circumstellar layers (Wallerstein 1985).

Measurements of velocity amplitudes in Mira variables have played a key role in the discussion on the pulsation mode of these stars (see Wood & Sebo 1996, for a summary). In addition, these observations constrained dynamical models of LPV atmospheres and led to today's understanding that the levitation of the outer layers of the stellar atmosphere driven by pulsation is essential for driving an efficient mass loss during this evolutionary phase (e.g., Höfner & Olofsson 2018).

Since the periods of LPVs can reach values of a few hundred days, obtaining velocity curves at high resolution with a good phase coverage remained challenging. The total number of Miras with such datasets available in the literature is still limited to a few tens (Hinkle et al. 1984; Hinkle & Barnbaum 1996; Lebzelter et al. 1999, 2005a,b; Alvarez et al. 2001). However, this sample covers a wide range in period, metallicity, and chemistry, revealing a consistent pattern in the velocity variations with s-shaped velocity curves in the near-infrared and peak-to-peak velocity amplitudes, depending on the lines used to trace the variation, between 20 and 30 km s^{-1} (Lebzelter & Hinkle 2002; Nowotny et al. 2010).

For physical and observational reasons, most of these studies were done in the 1.6 or 2.2 μm range relying on the first and second overtone lines of CO. These lines trace parts of the stellar atmosphere close to the pulsation driving zone (Nowotny et al. 2010). Within the spectrum, they are located close to the maximum of the spectral energy distribution of Miras and in an area with comparably low line blending and telluric absorption. Atomic lines in the same spectral range show a behavior very similar to the molecular lines (Hinkle & Barnes 1979). Velocity time series from the 4000 Å region show a much less expressed pattern with an amplitude of only 8 km s^{-1}. In the 4000 to 6800 Å range covered in the study of Alvarez et al. (2001), amplitudes around 20 km s^{-1} were measured, and thus the lines in this range compare well with the near-infrared range.

The semiregular variables (SRVs) show significantly smaller light amplitudes and most of them are pulsating in an overtone mode (Wood & Sebo 1996). Consequently, RV amplitudes are expected to be smaller for these stars, which has been confirmed observationally for SRVs with light amplitudes ranging from 0.1 to more than 2.5 mag (Lebzelter 1999; Lebzelter et al. 2005a). For the small amplitude and short period end, velocities of 1–5 km s^{-1} have been reported. Some stars have characteristics somehow between SRVs and Miras, such as W Hya (with a period of 390 days and an amplitude of more than 2 mag in V), and reach velocity amplitudes around 10 km s^{-1}. Semiregular light variability is typically reflected in the velocity variations.

From the point of view of RV variations, the ellipsoidal variables form a group of special interest among the LPVs (Soszyński et al. 2004). From their location in the period-luminosity diagram of LPVs, these stars are also known as sequence E stars. They are close binaries with one object being a red giant and the other one typically being a main sequence star. While there is no visible eclipse, regardless of it being due to an angle of orbital inclination that is too steep or due to the red giant being orders of magnitude brighter than the companion, there is a gravitational distortion of the red giant, which fills the Roche lobe. This produces an elongated shape of the object, and as the star rotates, brightness variations are observed due to this asymmetry.

As a consequence, the light and RV curves of these stars show two light cycles, but only one velocity cycle within one orbital period (Nicholls et al. 2010). Nie & Wood (2014) presented an extensive database of RV curves for 81 ellipsoidal variables. About 20% of these systems show eccentric orbits, a fraction twice as high as derived from light-curve analysis alone (Nie et al. 2017), which stresses the importance of RV data for the understanding of these variables. During their further evolution, the unseen companion will gain mass from the red giant leading to a common envelope system at some point. Ellipsoidal variables are assumed to be precursors of close binary planetary nebulae (Nicholls et al. 2010).

Another group of binaries among the LPVs are the symbiotic stars consisting of a red giant and a degenerated star such as a white dwarf or a neutron star. In the case of D-type symbiotics, the evolved star is a Mira (Hinkle et al. 2013). Radial-velocity changes thus combine pulsation and orbital motion. However, orbital periods of these systems are typically longer than decades (Seaquist & Taylor 1990) and they are therefore difficult to detect even in long velocity time series.

Finally, RV curves play a critical role in the explanation of the mysterious sequence-D stars. These LPVs show radial pulsation in some overtone modes combined with a secondary period that is typically ten times longer. Fundamental mode pulsation has been excluded as the cause of this secondary period (Wood 2000). Binarity and strange modes were suggested as alternative solutions. Interestingly, these long periods seem to form a period-luminosity sequence by themselves.

The origin of this kind of variability remains a matter of debate. Nicholls et al. (2010) showed that sequence-D stars are not ellipsoidals. From an attempt to model the velocity curves of a small sample of sequence-D stars, Hinkle et al. (2002) concluded that binarity is unlikely the cause of the variation because almost all of the objects analyzed show extremely similar values for the orbital parameters K, e, and ω. Soszyński & Udalski (2014) and Soszyński et al. (2021) suggest from a careful analysis of light curves and infrared data that sequence-D variability can be explained by an orbiting dust cloud in combination with a low-mass companion in a close circular orbit. On the other hand, Saio et al. (2015) show that the sequence-D period-luminosity relation agrees with expectations from oscillatory convective modes.

The observation of reliable RV curves of LPVs plays an important role for interpreting various aspects of these stars and their evolution. Observational challenges have limited the collection of large datasets up to now. Considering the variety of objects found among LPVs, the small existing dataset remains insufficient.

Since its second data release, *Gaia* has provided high-quality data for the study of the variability of LPVs, with the publication of photometric time series in the G, $G_{\rm BP}$, and $G_{\rm RP}$ bands of ~150 000 candidate LPVs in the second data release and over 2 million candidate LPVs in the *Gaia* Data Release 3 (DR3), respectively (Mowlavi et al. 2018; Lebzelter et al. 2023). Moreover, *Gaia* has the unique capability of simultaneously obtaining photometric and spectroscopic measurements owing to its Radial Velocity Spectrometer (RVS), thereby substantially boosting the possibility to investigate stellar variability. This feature was first exploited with the publication of RV time series of Cepheids and RR Lyrae as part of *Gaia* DR3 (Ripepi et al. 2023; Clementini et al. 2023). Here we extend this dataset to an additional 9614 sources that are part of the *Gaia* DR3 catalog of LPV candidates. In Sect. 2, we describe the procedures involved in the construction of this FPR catalog, while we present its content in Sect. 3 and discuss its quality in Sect. 4. In Sect. 5, we give an overview of the catalog, while Sect. 6 is dedicated to a summary and to conclusions.

Table 1. Summary of the steps involved in the construction of the catalog.

Steps	Sub-steps	Selection	# sources
Starting sample	`DR3-LPV`	2nd *Gaia* LPV catalog	1 720 588
	`DR3-RV`	radial velocity published in DR3	33 812 183
	`DR3-LPV-RV`	`DR3-LPV` and `DR3-RV`	501 308
Pre-filter	`bright`	`DR3-LPV-RV` and $G_{\rm RVS} < 12$ mag	249 600
	`vis-periods`	`DR3-LPV-RV` and `rv_visibility_periods_used` ≥ 12	180 850
	`small-err-rv`	`DR3-LPV-RV` and $\varepsilon_{V_R} < 0.175 \times$ `rv_amplitude_robust`	224 696
	`filter-0`	`bright` and `vis-periods` and `small-err-rv`	110 654
`processing`	Outliers removal, computation of best-fit model and period of the RV time series.		
Post-filter	`num-outliers`	`filter-0` and number of outliers in the RV time series $N_{V_R}^{\rm out} \leq 1$	105 715
	`no-trend`	`filter-0` and RV model's polynomial degree $D_p = 0$	90 942
	`high-snrv`	`filter-0` and $S/N_{V_R} > 1.5$	58 725
	`filter-1`	`no-trend` and `num-outliers` and `high-snrv`	44 216
	`Prv-low-limit`	`filter-1` and $P_{V_R} > 35$ days	24 118
	`Prv-up-limit`	`filter-1` and $P_{V_R} < \Delta t_{V_R}$	43 621
	`filter-2`	`Prv-low-limit` and `Prv-up-limit`	23 523
	`Prv-sim-Pg`	`filter-2` and $P_{V_R} \simeq P_G$	6 392
	`Prv-sim-Pbp`	`filter-2` and $P_{V_R} \simeq P_{G_{\rm BP}}$	5 768
	`Prv-sim-Prp`	`filter-2` and $P_{V_R} \simeq P_{G_{\rm RP}}$	6 534
	`Prv-sim-any-Pph`	`filter-2` and ($P_{V_R} \simeq P_G$ or $P_{V_R} \simeq P_{G_{\rm BP}}$ or $P_{V_R} \simeq P_{G_{\rm RP}}$)	7 551
	`Prv-sim-2Pg`	`filter-2` and $P_{V_R} \simeq 2P_G$	1 701
	`Prv-sim-2Pbp`	`filter-2` and $P_{V_R} \simeq 2P_{G_{\rm BP}}$	1 646
	`Prv-sim-2Prp`	`filter-2` and $P_{V_R} \simeq 2P_{G_{\rm RP}}$	1 703
	`Prv-sim-any-2Pph`	`filter-2` and ($P_{V_R} \simeq 2P_G$ or $P_{V_R} \simeq 2P_{G_{\rm BP}}$ or $P_{V_R} \simeq 2P_{G_{\rm RP}}$)	2 242
	`filter-3`$^{(a)}$	(`Prv-sim-any-Pph` or `Prv-sim-any-2Pph`)	9 614
Additional statistics	`Prv-sim-Pph+2Pph`	`filter-3` and (`Prv-sim-any-Pph` and `Prv-sim-any-2Pph`)	179
	`Prv-sim-Pph-only`	`filter-3` and (`Prv-sim-any-Pph` and (not `Prv-sim-any-2Pph`))	7 372
	`Prv-sim-2Pph-only`	`filter-3` and ((not `Prv-sim-any-Pph`) and `Prv-sim-any-2Pph`)	2 063
	`Prv-sim-all-Pph`	`filter-3` and ($P_{V_R} \simeq P_G$ and $P_{V_R} \simeq P_{G_{\rm BP}}$ and $P_{V_R} \simeq P_{G_{\rm RP}}$)	4 899
	`Prv-sim-all-2Pph`	`filter-3` and ($P_{V_R} \simeq 2P_G$ and $P_{V_R} \simeq 2P_{G_{\rm BP}}$ and $P_{V_R} \simeq 2P_{G_{\rm RP}}$)	1 194
	`top-quality`$^{(b)}$	(`Prv-sim-all-Pph` or `Prv-sim-all-2Pph`)	6 093

Notes. The exact criterion by which two periods are considered "similar" (indicated by $P_1 \simeq P_2$) is described in Sect. 2.3.2. $^{(a)}$All sources belonging to the `filter-3` subset are published as part of the FPR. $^{(b)}$The sources belonging to the subset `top-quality` are identified by the flag `flag_rv`.

Several Appendices complete the main body of the text. Appendix A gives additional details on the classification of LPV candidates presented in Sect. 3. Appendix B illustrates cases where the median RV differs significantly from the systemic RV. Appendix C analyzes the impact of the Java bug mentioned in Sect. 2.2 on the LPV results published in DR3. Finally, Appendix D gives some example queries to retrieve the data of the present catalog from the *Gaia* archive.

2. Catalog construction

Our starting dataset is the 2nd *Gaia* catalog of LPV candidates (Lebzelter et al. 2023) published as part of the *Gaia* DR3 (Gaia Collaboration 2023). More precisely, we consider the sources that appear in the table `gaiadr3.vari_long_period_variable` of the *Gaia* Archive. All these sources have their photometric time series already published in DR3. More than 70% of them do not have median RV published in *Gaia* DR3, most likely because they are too faint (see Katz et al. 2023). Therefore, we discard these sources, and focus on the remaining 501 308 LPV candidates having RV data in *Gaia* DR3. Hereinafter we adopt the notation $V_R^{\rm DR3}$ to indicate the median RV published as part of *Gaia* DR3 (it corresponds to the field `radial_velocity` of the `gaiadr3.gaia_source` table in the *Gaia* archive[1]).

For the construction of the catalog, we proceed in three main steps. To begin with, we employ the quantities derived from the *Gaia* RVS, and published as part of *Gaia* DR3, to refine the input source list to be fed to the processing pipeline. We refer to this first step as "pre-filtering," and describe it in Sect. 2.1.

We then analyze the time series of the selected sources with an updated version of the pipeline used for variability processing in *Gaia* DR3 (Eyer et al. 2023; Lebzelter et al. 2023), as we describe in Sect. 2.2. Both the RV time series and the three photometric time series (in the *Gaia* G, $G_{\rm BP}$, and $G_{\rm RP}$ bands) undergo this "processing" step, that involves the detection and removal of outlier epochs, the computation of time series statistics, and the determination of the best-fit model.

Lastly, we employ the resulting quantities to further refine the sample of sources for publication. This final step is referred to as "post-filtering," and is described in Sect. 2.3. The filtering conditions and number of selected sources of each step and the corresponding sub-steps are summarized in Table 1.

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

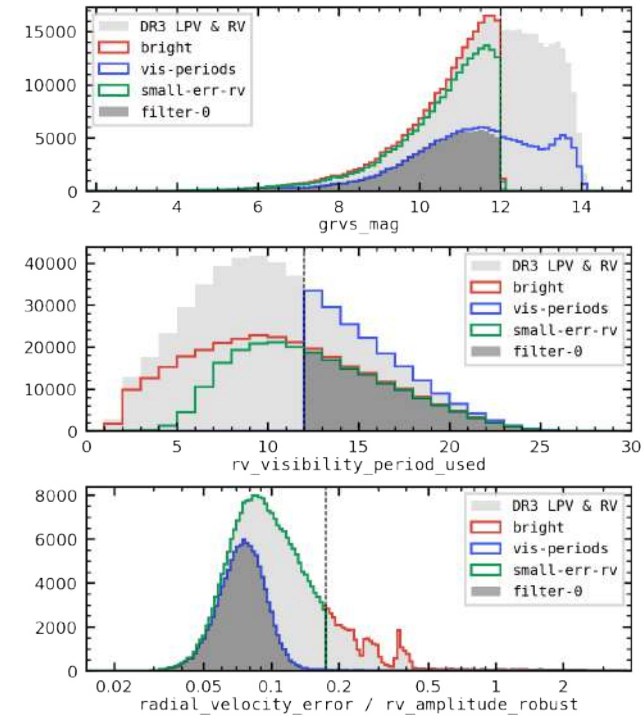

Fig. 1. Distribution of the pre-filtering parameters $G_{\rm RVS}$ (top panel), `rv_visibility_periods_used` (middle panel), and ε_{V_R}/`rv_amplitude_robust` (bottom panel), with vertical dashed lines indicating the filter limits. Different colors indicate the starting set (`DR3-LPV-RV`, gray filled histogram) and the individual pre-filters, labeled `bright` (red curve), `vis-periods` (blue curve), and `small-err-rv` (green curve) as in Table 1. The black filled histogram corresponds to the application of all three pre-filters (`filter-0`). We note that 50 349 sources in the starting sample of 501 308 sources lack a published value of $G_{\rm RVS}$ as it would be fainter than 14.1 mag, and that the quantity `rv_amplitude_robust` is not provided for sources with $G_{\rm RVS}$ > 12 mag (Sartoretti et al. 2023).

2.1. Pre-filtering

At this stage we aim to limit the sample to the objects with the highest-quality RV measurements by taking advantage of the information available from *Gaia* DR3 (namely in the `gaiadr3.gaia_source` table of the *Gaia* Archive). This is achieved by retaining only sources with large enough RVS flux, a sufficient number of RV measurements, and relatively small RV uncertainty ε_{V_R}. The relevant quantities and corresponding cuts involved in this process are illustrated in Fig. 1.

We begin by applying a filter that excludes the faintest objects in our dataset, using the median value $G_{\rm RVS}$ of the epoch $G^t_{\rm RVS}$ magnitudes (`grvs_mag` in the *Gaia* Archive, see Sartoretti et al. 2023), which are obtained by integration of the RVS epoch spectra. By requiring that $G_{\rm RVS}$ < 12 mag we limit our sample to "bright" stars (top panel of Fig. 1) following the distinction made for the DR3 RVS processing (Katz et al. 2023). Almost 250 000 sources meet this criterion. It is worth pointing out that several RV-related quantities published in DR3, such as the median RV and its uncertainty, are computed with different methods depending on whether the sources are brighter or fainter than $G_{\rm RVS}$ = 12 mag. Having required that $G_{\rm RVS}$ < 12 mag, these quantities are defined unequivocally for all sources in our sample. Namely, the RV is obtained as the median of the single-transit RVs, while the RV error is the uncertainty on the median of the epoch RVs, with a constant offset accounting for a calibration floor contribution (Sartoretti et al. 2022).

Then, we apply a condition to the number of data points in each RV time series. We note that the actual number of RV observations is not necessarily appropriate for this filtering step, as they often come in groups that span a relatively short period of time (often shorter than several days) because of the *Gaia* scanning law (see Eyer et al. 2017). This issue is often overcome through the concept of visibility period, that is a group of transits separated from other such groups by a gap of at least 4 days. The number of visibility periods used in the derivation of radial velocities is a parameter available from DR3, and we employ it to set the condition `rv_visibility_periods_used` ≥12. As will be explained in Sect. 2.2, a minimum number of 9 data points is necessary to obtain a time series model, but that may still not be enough for the model to be well-constrained. At the same time, raising too much the threshold would lead to the exclusion of too many sources, as can be appreciated from the middle panel of Fig. 1. We found that a good compromise could be attained by setting the threshold at 12. The condition on visibility periods is fulfilled by about 180 000 sources in our starting dataset.

The uncertainty ε_{V_R} on the median RV is provided in the `radial_velocity_error` data field of the `gaiadr3.gaia_source` table. Instead of setting an absolute upper limit to ε_{V_R}, we rather compare it with the amplitude of the RV curve (`rv_amplitude_robust`) estimated during DR3 processing after outlier removal[2]. We inspected visually the distribution of said parameters for the sources in our sample, before and after the application of the conditions on $G_{\rm RVS}$ and number of visibility periods (bottom panel of Fig. 1), and decided to construct the filtering condition in the form

$$\varepsilon_{V_R} < 0.175 \times \texttt{rv\_amplitude\_robust}, \quad (1)$$

which retains almost 225 000 objects from the starting sample. The combination of the three conditions described above results in a pre-filtered sample of 110 654 RV time series, that we input into the variability pipeline.

2.2. Time series processing

Overall, the processing of the RV and photometric time series is performed in a very similar manner as it was done for the photometric time series of LPVs in DR3. Therefore, we briefly summarize the procedure, focusing on the specific parameters for RV analysis and the few differences due to updates to the pipeline, and refer the reader to Lebzelter et al. (2023, and references therein) for more details. The processing operations involve the detection and removal of outliers, followed by the calculation of time series statistics, and the derivation of the best-fit model.

The setup for detecting outliers in the photometric time series are unchanged with respect to DR3. For the RV time series, we exclude epochs with RV that:

– have an uncertainty larger than 5 km s^{-1};
– deviate from the median of the time series by more than 100 km s^{-1};
– deviate from the median of the time series by more than 10 times the median absolute deviation of the time series.

The choice of these parameters was guided by physical considerations concerning the typical RV amplitude for pulsation in LPVs, that is not expected to exceed several tens of km s^{-1}.

[2] We remark that the outlier removal procedure employed during DR3 processing is different from the one adopted in the pipeline used in the present work (Sect. 2.2), therefore the results are not necessarily the same (see Sect. 4.1.1).

However, we quickly realized that the sample contains a non-negligible fraction of high-quality RV curves likely originating from binarity, that we did not want to reject. Therefore, we have launched a few runs of the pipeline using rather permissive values, and tuned them by visual inspection of the distribution of the resulting time series statistics.

To describe the RV and photometric time series we adopt the same kind of mono-periodic model with frequency f_X (where the subscript $X \in \{V_R, G, G_{\rm BP}, G_{\rm RP}\}$ indicates the type of time series), that consists of the sum of a polynomial trend of degree $D_{p,X} \leq 1$ (i.e., no trend or a linear trend) and a Fourier series with up to $N_{h,X} = 3$ components (i.e., up to the second harmonic). Using a notation similar to that of Eyer et al. (2017), the model is defined as

$$y = \sum_{k=1}^{N_{h,X}} A_{k,X} \cos\left[2\pi k f_X (t - t_{0,X}) + \psi_{k,X}\right] + \sum_{i=0}^{D_{p,X}} c_{i,X} (t - t_{0,X})^i \quad (2)$$

where $A_{k,X}$ and $\psi_{k,X}$ are the amplitude and phase of the kth Fourier component, respectively, and $t_{0,X}$ is a reference epoch. To avoid overfitting, the number of Fourier components is limited by the condition $N_{h,X} < \pi/\Delta\phi_{\max,X}$ on the maximum phase gap $\Delta\phi_{\max,X}$ of the folded time series (cf. Eyer et al. 2017). While this approach is effective in most cases, it may fail for the few time series that end up having large and repeated gaps, and hence lack coverage of specific phase intervals, which makes them especially exposed to overfitting. A similar effect may result if the best-fit period is longer than the duration of the time series (see Sect. 2.3.2).

For each source, the RV time series and the three photometric time series are processed independently of each other. For each time series, we begin by computing the periodogram. It is computed over the frequency range $[7 \times 10^{-4}, 0.1]$ day^{-1}, with an even spacing in frequency of 0.33×10^{-4} day^{-1}. We take the period of the time series to be equal to the value corresponding to the highest peak of the periodogram. After a first determination of the best model in the form given by Eq. (2), we employ a nonlinear Levenberg-Marquardt optimization algorithm to improve the result.

It should be clear that the main peak of the periodogram identifies the strongest periodic signal in a time series, which is not necessarily the same as the period of the underlying physical process. In particular, in the case of ellipsoidal red giants, the light curve shows two minima per cycle of possibly different depths, that mimic a variation with a period half as long as the true orbital period. This effect is not present in the RV time series, so that the occurrence is not uncommon of sources whose RV period is twice the photometric period as determined from the periodogram. This will be further discussed in Sect. 3.

We performed safety checks by comparing the newly derived photometric periods with the ones derived from G-band light curves and published in Lebzelter et al. (2023). To our surprise, despite the pipeline setup and sequence of operations being identical, we found that in several cases the results are not exactly the same. We traced this discrepancy to a bug of Java version 8 affecting the nonlinear modeling of the time series, that disappeared during the upgrade to Java 17 performed after the conclusion of *Gaia* DR3 operations. We remark that the deviations are small, and affect a minimal fraction of the sources. We provide a deeper analysis of this issue in Appendix C.

2.3. Post-filtering

We tackle the post-filtering in two successive sub-steps. The first one involves the properties of the cleaned RV curves (i.e., after outlier removal) revealed by the time series statistics as well as the parameters of the best-fit model, with the exception of the frequency. The latter quantity is considered in the second sub-step, aimed at excluding the objects whose best RV period (P_{V_R}) is uncertain. We construct a filtering criterion by comparing with each other the periods derived from the RV and photometric time series.

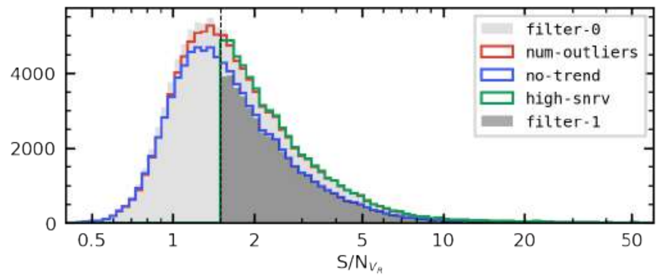

Fig. 2. Distribution of the RV signal-to-noise ratio, S/N_{V_R}, for the pre-filtered sample (gray histogram) and with the conditions involved in the `filter-1` post-filtering step, labeled `num-outliers` (red curve), `no-trend` (blue curve), and `high-snrv` (green curve) and described in Table 1. The dark gray histogram shows the combination of the three conditions. The vertical dashed line indicates the $SN_{V_R} = 1.5$ threshold.

2.3.1. Selection on RV time series properties

To begin with, we assess the impact of outlier removal on the RV time series of the pre-filtered sample, and examine the number $N_{V_R}^{\rm out} = N_{V_R}^{\rm raw} - N_{V_R}$ of rejected epochs as a parameter for constructing an additional filter, where N_{V_R} is the number of epochs in the cleaned RV time series and $N_{V_R}^{\rm raw}$ is the number of valid measurements in the original RV time series (i.e., excluding `NaN` values). The majority of the time series (about 88%) are unaffected, while $N_{V_R}^{\rm out} = 1$ for about 7.5% of the sources, and the remaining 4.5% of RV curves had at least two rejected epochs. Visual inspection of time series and folded RV curves with varying number of outliers reveals satisfying results for $N_{V_R}^{\rm out} \leq 1$, as well as a rapid degradation with increasing $N_{V_R}^{\rm out}$. We therefore restrict our sample by requiring that no more than one RV epoch is excluded during outliers removal, a condition that selects 105 715 sources from the pre-filtered sample.

One of the properties that we found to be often associated with low-quality fits is the adoption of a first-degree polynomial in the RV curve model. About 18% of the RV time series in the pre-filtered sample are modeled this way. This is done automatically by the variability pipeline when the inclusion of a linear trend results in a better fit with the underlying data. However, the combination of a relatively small number of epochs and characteristic time scales comparable with the duration of the time series make this approach poorly suited for LPVs. In contrast with variable objects with shorter periods, for which the inclusion of a linear trend can significantly improve the characterization of the time series, in the case of LPVs it tends to pick up the signal associated with long periods, while in some cases erroneously detrends RV curves with poor phase coverage. We therefore retain only the time series modeled without a linear trend, which correspond to 90 952 sources in the pre-filtered dataset.

Finally, we apply a threshold to the signal-to-noise ratio of the cleaned RV time series at $S/N_{V_R} = 1.5$ (see Fig. 2). In the pre-filtered sample there are 58 725 sources above that limit. By combining the three conditions described above, we reduce the pre-filtered sample down to 44 216 sources.

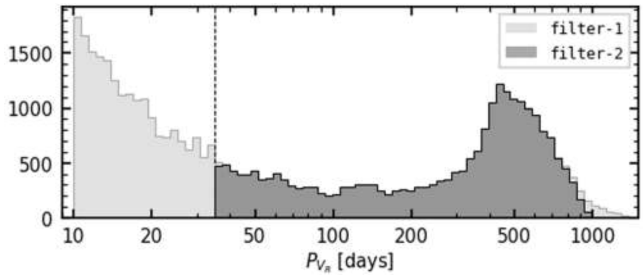

Fig. 3. Distribution of the RV periods of the sample after applying the post-filtering conditions `filter-1` (gray histogram, see Table 1) and after applying the conditions on the periods themselves (`filter-2`, black histogram). The vertical dashed line indicates the lower period limit at 35 days.

2.3.2. Selection on the periods

We follow the approach described in Lebzelter et al. (2023) to bound the range of RV periods, setting a fixed lower limit to 35 days, and excluding the cases in which P_{V_R} is longer than the duration Δt_{V_R} of the RV time series. We recall that the adoption of such a lower limit by Lebzelter et al. (2023) for the photometric time series was aimed at minimizing the contamination from spurious signals. These conditions further reduce our dataset to 23 523 sources. The distribution of RV periods before and after the application of these conditions is shown in Fig. 3.

Finally, we examine how close the RV period is to the periods obtained from modeling the photometric time series. To do so, we follow the method described by Mowlavi et al. (2023, their Sect. 4.1). We quantify the similarity between the RV period P_{V_R} and a photometric period $P_{\rm ph}$ (either P_G, $P_{G_{\rm BP}}$, or $P_{G_{\rm RP}}$) by the quantity

$$r_{V_R,{\rm ph}} = \frac{|P_{V_R} - P_{\rm ph}|}{P_{V_R}} \frac{\Delta t_{V_R}}{P_{V_R}}, \tag{3}$$

which represents the maximum phase deviation a signal with period $P_{\rm ph}$ can accumulate with respect to P_{V_R} during the observation duration Δt_{V_R}. In order to better understand the meaning of Eq. (3), we note that $\delta\phi_{V_R,{\rm ph}} = |P_{V_R} - P_{\rm ph}|/P_{V_R}$ is the difference between the two periods normalized to P_{V_R}, whereas $\Delta t_{V_R}/P_{V_R} = n^{\rm cyc}_{V_R}$ is the number of cycles with period P_{V_R} covered by the RV time series. Let us assume that the RV and photometric curves are in phase at the very beginning of the RV time series. Unless the two periods are identical, after one RV cycle P_{V_R} the two curves show a phase offset that is exactly equal to $\delta\phi_{V_R,{\rm ph}}$. After two RV cycles the phase deviation is twice as large, and so on. At the very end of the RV time series, the phase offset is $\delta\phi_{V_R,{\rm ph}} \times n^{\rm cyc}_{V_R} = r_{V_R,{\rm ph}}$. It is easy to see that this is also the maximum possible phase deviation for given P_{V_R}, $P_{\rm ph}$, and Δt_{V_R}.

From Eq. (3) it is clear that $r_{V_R,{\rm ph}}$ is defined asymmetrically, and that $r_{V_R,{\rm ph}} \neq r_{{\rm ph},V_R}$. However, the closer the values of the period being compared, and the smaller the asymmetry is. The distribution of the values of $r_{V_R,{\rm ph}}$ and $r_{{\rm ph},V_R}$ for all three photometric periods show that the two quantities rapidly converge when they are smaller than unity. We thus construct our "period similarity" condition in the form

$$r^{\rm max}_{V_R,{\rm ph}} = r^{\rm max}_{{\rm ph},V_R} = \max(r_{V_R,{\rm ph}}, r_{{\rm ph},V_R}) < 1. \tag{4}$$

Figure 4 shows the distribution of $r_{V_R,{\rm ph}}$ versus the number of cycles $n^{\rm cyc}_{V_R}$ covered by the RV time series. A large number of sources accumulate along two slanted stripes in the diagram, that

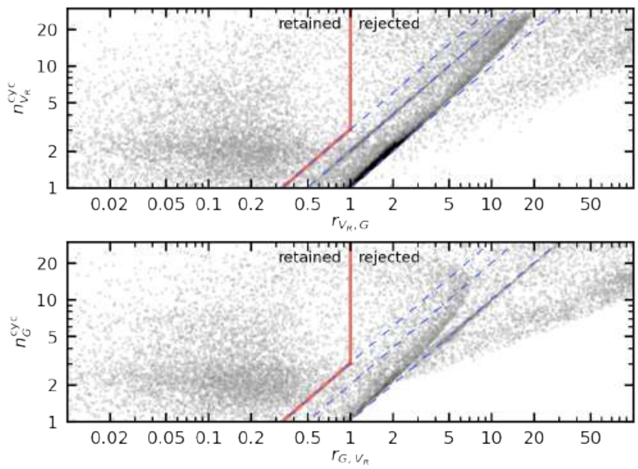

Fig. 4. Number of observed cycles versus phase deviation at the last cycle, comparing the RV and G-band curve models for the sample post-filtered down to `filter-2` (see Table 1). More precisely, the top panel shows the phase deviation $r_{V_R,G}$ with respect to the last RV cycle and the number $n^{\rm cyc}_{V_R}$ of RV cycles, while the same quantities are referred to the G-band time series in the bottom panel. In each panel, the thick red line marks the upper limit to the phase difference employed in post-filtering (Eq. (4)), while the dashed lines indicate $n^{\rm cyc}/r$ = 3, 2, and 1. Data points to the right of the thick red line are rejected. A similar picture emerges when the $G_{\rm BP}$ or $G_{\rm RP}$ time series are considered in place of G.

are only partially rejected by the condition defined by Eq. (4). These stripes correspond to $r_{V_R,{\rm ph}} \simeq n^{\rm cyc}_{V_R}$ and $r_{V_R,{\rm ph}} \simeq 2n^{\rm cyc}_{V_R}$, respectively. It is easy to show that the former case corresponds to $P_{\rm ph} \simeq 2P_{V_R}$ or $P_{\rm ph} \ll P_{V_R}$, and the latter to $P_{\rm ph} \simeq 3P_{V_R}$ (or $P_{\rm ph} \simeq -P_{V_R}$, which is not possible as periods have positive values). They clearly indicate situations of incompatibility between pairs of periods, and should be excluded. To do so, we require that

$$\delta P_{V_R,{\rm ph}} = \delta P_{{\rm ph},V_R} = \frac{|P_{V_R} - P_{\rm ph}|}{\min(P_{V_R}, P_{\rm ph})} < \frac{1}{3}, \tag{5}$$

which completes our condition on period similarity (Eq. (4)). The combination of Eqs. (4) and (5) for the G-band and RV periods is equivalent to taking only the data points that are on the left of the red lines in both panels of Fig. 4. Therefore, our definition of period similarity is given by

$$P_{V_R} \simeq P_{\rm ph} \quad \Longleftrightarrow \quad \left[\left(r^{\rm max}_{V_R,{\rm ph}} < 1\right) \text{ and } \left(\delta P_{V_R,{\rm ph}} < \frac{1}{3}\right)\right]. \tag{6}$$

It should be noted that, if a source displays variability due to binarity, the main peak in the periodogram of any one of its photometric time series can be half of the true orbital period, and hence of the RV period. Therefore, the requirement that $P_{V_R} \simeq P_{\rm ph}$ could lead to exclude these sources. In order to avoid this, we are also interested in using a requirement in the form $P_{V_R} \simeq 2\,P_{\rm ph}$, which means that any occurrence of $P_{\rm ph}$ in Eq. (6) is replaced by $2\,P_{\rm ph}$. Our final requirement is therefore

$$(P_{V_R} \simeq P_{\rm ph} \text{ or } P_{V_R} \simeq 2\,P_{\rm ph}) \text{ for any } P_{\rm ph} \in \{P_G, P_{G_{\rm BP}}, P_{G_{\rm RP}}\}. \tag{7}$$

Figure 5 displays the period distributions for the sources displaying compatibility between the RV period and one or more photometric periods according to Eq. (6). Comparing with the distribution in Fig. 3 we note the effectiveness of this type of

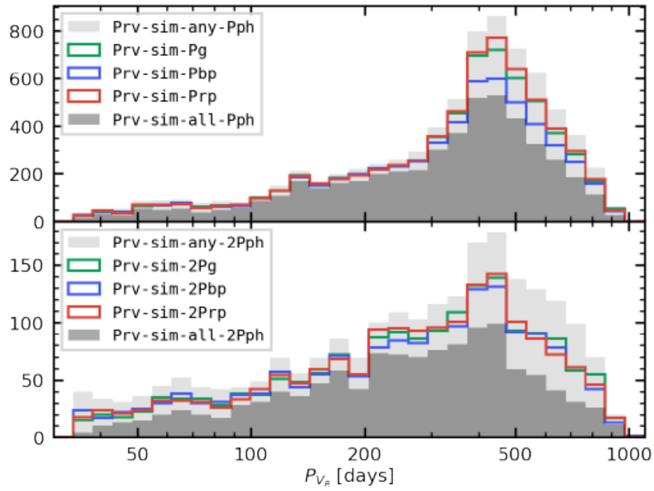

Fig. 5. Period distribution, after applying `filter-2`, for the subsets with RV period similar to one or more photometric periods (top panel), or twice of it (bottom panel). The light gray and dark gray histograms represent the sets in which the RV period is similar to at least one photometric period or to all of them, respectively. The colored curves represent the sets in which the RV period is similar to P_G (green), $P_{G_{BP}}$ (blue), or $P_{G_{RP}}$ (red).

selection at rejecting periods shorter than ~200 days, where a higher rate of occurrence of spurious frequencies is expected (see Holl et al. 2023). This is also true when the RV period is compared with twice one of the photometric periods.

We note that the number of sources with $P_{V_R} \simeq P_{G_{RP}}$ is slightly higher than that with $P_{V_R} \simeq P_G$, which in turn are more numerous than the objects with $P_{V_R} \simeq P_{G_{BP}}$. The same trend is present when the comparison is made against twice the photometric period, but is less pronounced. This could be indicative of a color-dependence of the photometric variability features, typical of pulsating stars. The fact that this feature becomes less conspicuous when $P_{V_R} \simeq 2P_{ph}$ would support the interpretation that the variability of these sources is extrinsic and associated with binarity.

We construct the final filter by applying the condition defined by Eq. (7) inclusively to the three photometric periods, that is, we require that the RV period is similar to at least one of them. The final dataset consists of 9614 sources.

2.4. Top-quality sample

By the criteria defined above, we identified a subset of the FPR sample consisting of sources displaying a high degree of compatibility between the RV and photometric variability. Namely, there are 6093 sources whose RV period is consistent with each one of the three periods derived from the photometric time series. This means that these sources fullfil the condition

$$(P_{V_R} \simeq P_{ph} \text{ or } P_{V_R} \simeq 2\,P_{ph}) \quad \text{for all} \quad P_{ph} \in \{P_G, P_{G_{BP}}, P_{G_{RP}}\}. \qquad (8)$$

These sources are identified by the field `flag_rv =True` in the *Gaia* Archive (see Sect. 2.5), and form a subset that we dubbed the "top-quality sample" (TQS).

Such a high consistency between the RV and photometric periods is a strong indication that a signal originating from the same physical process is being detected in all four time series, with two important consequences. On the one hand, these sources can be used to investigate a given type of variability in its different aspects (physical motion, changes in brightness and color) with a good degree of confidence that they all trace the same phenomenon. Given the multi-periodic nature of LPVs, this is far from trivial. On the other hand, there is a comparatively small probability that the periodic signal picked up by the variability processing pipeline is spurious.

Other than this self-consistency, the sources in the TQS have on the average the same properties as the remaining FPR sources, with the only exception that they include a larger fraction of sources identified as binary variables (see Sect. 3). Indeed, binary candidates are assigned to the TQS with a higher frequency (~80%) than other sources in the FPR (~60%). These trends are likely to be attributed to the fact that, owing to its geometric nature, binary-induced variability shows smooth variations compared with the pulsation of LPVs, known to display irregularities.

We compared the TQS and the other FPR sources in terms of the distributions of several quantities from *Gaia* DR3. The sources in the former set display slightly better astrometry (smaller errors in sky coordinates and proper motions), but the two sets are equivalent in terms of relative parallax uncertainty. These properties are to be attributed to a slightly higher number of visibility periods used in the astrometric solution. The uncertainty associated with both photometric and RV measurements as reported in the *Gaia* DR3 source table is slightly higher for the TQS sources, which simply reflects the fact that they tend to exhibit larger variability amplitudes. The TQS sources follow essentially the same brightness distributions of all other FPR sources in all three *Gaia* bands, except they are slightly brighter in G_{BP} and fainter in G_{RP}, and as a result they appear to have a slightly bluer color which reflects the larger fraction of binaries in the TQS compared to other FPR sources, see Sect. 5. The two sets do not show any particular difference in their RV distributions.

Some more significant differences between the two sets are found in terms of the variability parameters (Fig. 6). The requirement of period consistency effectively excludes from the TQS sources with short G-band periods (due to the lower limit at 35 days on P_{V_R}). Moreover, the amplitude distribution of TQS sources tends to be skewed toward slightly larger values compared with the full FPR sample, as they are associated with a higher signal-to-noise ratio (and hence a higher chance of picking the same periodicity in RV and in photometry). Other amplitude-related parameters (such as standard deviation, interquartile range, or Stetson variability index) show similar trends. For similar reasons the TQS sources display smaller Abbe values (Mowlavi 2014; Mowlavi et al. 2017) than other sources in all *Gaia* bands, indicating smoother light curves. However, such a difference is not present for the Abbe value computed for the RV time series.

Finally, we inspected the mean value of the uncertainties associated with single epochs (either of the RV or photometric time series) and the mean of the absolute residuals of the time series models, and found differences between the distribution associated with the TQS and with other sources that are consistent with the different amplitude distributions. Therefore, we remark that we consider this subsample to be of superior quality within the FPR because of its content of coherent physical information rather than in terms of actual quality of measurements.

2.5. Data fields

The present catalog follows the same scheme as the 2nd *Gaia* catalog of LPV candidates (Lebzelter et al. 2023), and has

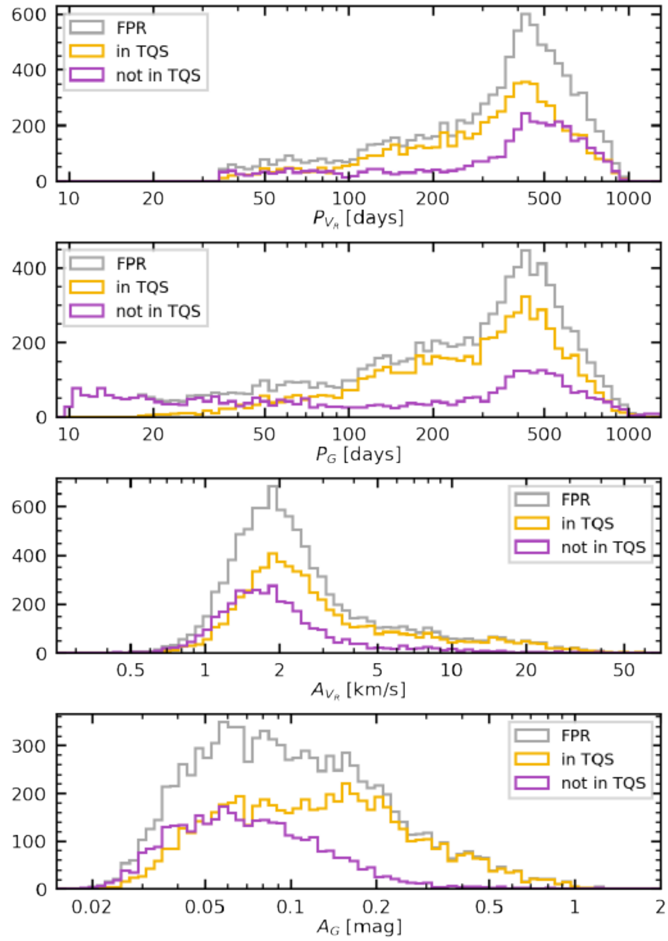

Fig. 6. Distribution of the RV and *G*-band variability parameters for the all the FPR sources (gray curves), and distinguishing between whether they are in the top-quality sample (TQS, orange curves) or not (purple curves). Panels from top to bottom show the distributions of RV periods (P_{V_R}), *G*-band periods (P_G), RV semi-amplitudes (A_{V_R}) and *G*-band semi-amplitudes (A_G).

therefore the same data fields, with the addition of the corresponding fields for the RV variability. More precisely, the fields `solution_id`, `source_id`, `median_delta_wl_rp`, and `isCstar` are left unchanged, while the fields `frequency`, `frequency_error`, and `amplitude` have their values replaced with the newly derived parameters of the best-fit model for the *G*-band time series (see Sect. 2.2 for the reason of the updated values). Finally, the following four data fields are added.

FREQUENCY_RV: *Frequency of the RV curve (double, Frequency [day^{-1}]).* This field provides the frequency determined from the RV time series.

FREQUENCY_ERROR_RV: *Uncertainty on the RV frequency (float, Frequency [day^{-1}]).* This field provides the uncertainty on the frequency of the RV time series.

AMPLITUDE_RV: *Amplitude of the RV curve (float, Velocity [km s^{-1}]).* This field gives the half peak-to-peak amplitude (semi-amplitude in km s^{-1}, based on the best-fit model; see Sect. 2.2).

FLAG_RV: *Flag identifying the top-quality subsample (boolean).* This field identifies the sources whose RV period

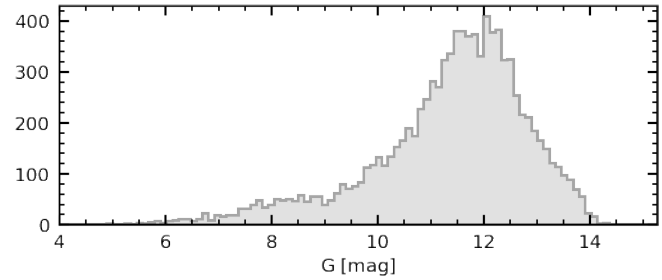

Fig. 7. *G*-band brightness distribution of the FPR sample.

is fully compatible with all three photometric periods (see Sect. 2.4).

The full RV time series for all sources in this FPR are available for download from the table `gaiafpr.vari_epoch_radial_velocity` in the *Gaia* archive, while the statistics for the cleaned RV time series are provided in the table `gaiafpr.vari_rad_vel_statistics`, following the same scheme adopted in *Gaia* DR3 for the RV time series of Cepheids and RR Lyrae (cf. Ripepi et al. 2023; Clementini et al. 2023). In Appendix D we provide some instructions on how to retrieve the FPR data.

Hereinafter we adopt the notation A_G and A_{V_R} to indicate the quantities `amplitude` and `amplitude_rv` published in this FPR, corresponding to the semi-amplitude of the fundamental component of the best-fit Fourier model of the *G*-band and RV time series, respectively. We note that roughly half of the *G*-band time series and more than 80% of the RV time series have been modeled with a single-component Fourier series, so the published value is exactly the semi-amplitude of the model. The remaining time series have harmonic components whose amplitude is typically much smaller than that of the fundamental component, so that the semi-amplitude of the latter is still representative to the semi-amplitude of the full Fourier model. Therefore, for simplicity, we often refer to A_G (not to be confused with the *G*-band extinction) and A_{V_R} as “semi-amplitude of the time series models”, whereas their formal meaning should be clear.

3. Catalog content

This FPR provides epoch RVs for 9614 candidate LPVs, of which the G, $G_{\rm BP}$ and $G_{\rm RP}$ time series are available in DR3 as part of the second *Gaia* catalog of LPV candidates. Figure 7 shows the *G*-band distribution of these sources, which cover the range $6 \lesssim G/\text{mag} \lesssim 14$. The RV time series have between 12 and 90 measurements, with an average of 24 epochs, unevenly sampling a time interval of about 3 yr. More precisely, the RV time series have a mean duration of 905 days, spanning a range between about 500 and 1000 days, but with a distribution skewed toward longer durations (typically $\gtrsim$800 days). The number of epochs in the RV and *G*-band time series, as well as the number of visibility periods adopted for deriving median RVs in DR3 (see Sect. 2.1), are illustrated in Fig. 8. We note that these numbers correspond to the cleaned time series, that is after outlier removal, and are the same values given in the *Gaia* archive summary tables, whereas the published *Gaia* light curves include the outlier epochs as well (flagged to indicate whether they have been rejected by the variability pipeline).

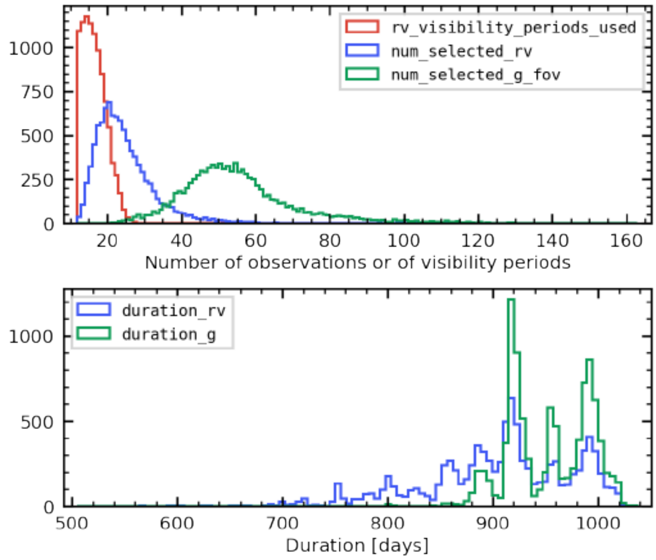

Fig. 8. Distribution of the number of observations (top) and duration (bottom) of the RV and G-band time series of the FPR sources. The red line indicates the number visibility periods used to derive the median RV published in DR3 (a single visibility period may contain multiple epochs, see Sect. 2.1). The blue and green lines indicate the number of measurements in the cleaned RV and G-band time series (top) or their duration (bottom).

Besides the filtering steps, in Table 1 we provide a summary of a few interesting subsets of the final sample, obtained by further comparing the periods derived from the RV and photometric time series. Following the criteria defined in Sect. 2.3.1, the period comparison for a given source has three possible outcomes: (1) $P_{V_R} \simeq P_{\rm ph}$, (2) $P_{V_R} \simeq 2\,P_{\rm ph}$, or (3) $P_{V_R} \not\simeq P_{\rm ph}$ and $P_{V_R} \not\simeq 2\,P_{\rm ph}$. These conditions are not necessarily the same for each of the three photometric periods. For instance, a source might be such that $P_{V_R} \simeq P_G$, $P_{V_R} \simeq 2\,P_{G_{\rm RP}}$, and at the same time $P_{V_R} \not\simeq P_{G_{\rm BP}}$ and $P_{V_R} \not\simeq 2\,P_{G_{\rm BP}}$. However, due to the filters we applied, either conditions (1) or (2) must be verified for at least one of $P_{\rm ph} \in \{P_G, P_{G_{\rm BP}}, P_{G_{\rm RP}}\}$ for all sources published in the FPR. These conditions allow us to distinguish between three types of sources:

- only 1:1 compatibility: $P_{V_R} \simeq P_{\rm ph}$ for at least one photometric period, but none of the other periods meets the condition $P_{V_R} \simeq 2\,P_{\rm ph}$;
- only 2:1 compatibility: $P_{V_R} \simeq 2\,P_{\rm ph}$ for at least one photometric period, but none of the other periods meets the condition $P_{V_R} \simeq P_{\rm ph}$;
- "mixed" compatibility: $P_{V_R} \simeq P_{\rm ph}$ for at least one photometric period, and $P_{V_R} \simeq 2\,P_{\rm ph}$ for at least one of the other photometric periods (as in the example above).

The majority of the FPR sources fall in the first category, consisting of 7372 sources (about 77%). There is no direct indication that the variability of these sources results from binarity, as none of the photometric period is close to a 2:1 ratio with respect to the RV period. Of course, this does not prove that they are not binary variables. However, it is reasonable to assume that most of these sources are probably pulsating stars, at least for the purpose of assessing the relative fractions of these types of variables in the FPR. Similarly, the 2063 sources (about 21%) belonging to the second category in the list above are probably binary variables. More precisely, as they are selected among bright red giants, these sources are most likely ellipsoidal variables (ELL), and will be referred as such hereinafter. Further evidence supporting this statement will be provided in Sect. 5.

Finally, there exist 179 sources such that their RV period is consistent with one or two of the photometric periods, and twice the value of the remaining ones. We examined visually the time series of a random sample of these sources, and found that the cleaned light curves often show large phase gaps when folded with the RV period. Figure 9 shows a clear example with a lack of data near minimum light. All time series show a similar trend, and the best-fit model to the RV, G, and $G_{\rm BP}$ time series is visually convincing, yet the $G_{\rm RP}$ has a best-fit model with half the period found in the other time series. A similar situation can arise when the time series covers a small number of RV cycles, so it becomes difficult to constrain the period precisely. It is clear that this kind of mixed consistency between photometric and RV periods has artificial causes and, in principle, none of the two periods can be confidently taken to be the correct one. It is not possible to make any inference on the nature of these sources based only on their periods. However, the fact that they represent less than 2% of the FPR is encouraging.

These three categories give us a general idea of the fractions of ellipsoidal and pulsating variables in the FPR based on weak conditions on period consistency. Stronger conditions can be imposed by restricting the analysis to the TQS, which includes 4899 probable pulsators ($P_{V_R} \simeq P_{\rm ph}$ for each $P_{\rm ph}$) and 1194 probable ELL ($P_{V_R} \simeq 2P_{\rm ph}$ for each $P_{\rm ph}$). The two kinds of sources make up for about 80% and 20% of the TQS, respectively. These percentages are fully compatible with the values found in the previous paragraph.

3.1. Candidate ellipsoidal variables

Classifying the types of variables discussed above based only on the ratio between the RV period and photometric periods is not necessarily a good approach. In particular, it might be inappropriate if one or more of the cleaned time series end up having a small number of measurements, so that the corresponding period is poorly constrained. Therefore, we use the semi-amplitude A_{V_R} of the RV time series model, and the corresponding value A_G for the G-band model, to perform a deeper analysis. In doing so we consider only the TQS in the rest of this section, so to obtain as clean a picture as possible.

The G-band and RV semi-amplitudes derived for the sources in this sample are displayed in Fig. 10. Two groups are clearly separated in this diagram (in either panel). The first group shows G-band variations over a wide range ($0.02 \lesssim A_G/{\rm mag} < 2$), but is limited to relatively small RV amplitudes ($A_{V_R} \lesssim 10\,{\rm km\,s^{-1}}$, with only a few exceptions). The second group is characterized by large RV variations ($A_{V_R} \gtrsim 5\,{\rm km\,s^{-1}}$) and relatively small light amplitudes ($A_G \lesssim 0.2$ mag). We can readily interpret the former group as consisting of pulsating stars, whose brightness changes can become very large (owing to strong absorption by molecules that form efficiently in the expanding phase of the cycle, Reid & Goldston 2002) while they can hardly attain pulsation velocities larger than ${\sim}20\,{\rm km\,s^{-1}}$ (Nowotny et al. 2010). In contrast, orbital velocities can easily exceed that value in binaries, but their G-band variations do not exceed a few tenths of magnitude. This interpretation is supported by the fact that the vast majority of sources with $P_{V_R} \simeq P_{\rm ph}$ are found in the former group (black points in the top panel of Fig. 10), whereas most sources in the latter group have $P_{V_R} \simeq 2\,P_{\rm ph}$ (black points in the bottom panel), although some contamination is present in both.

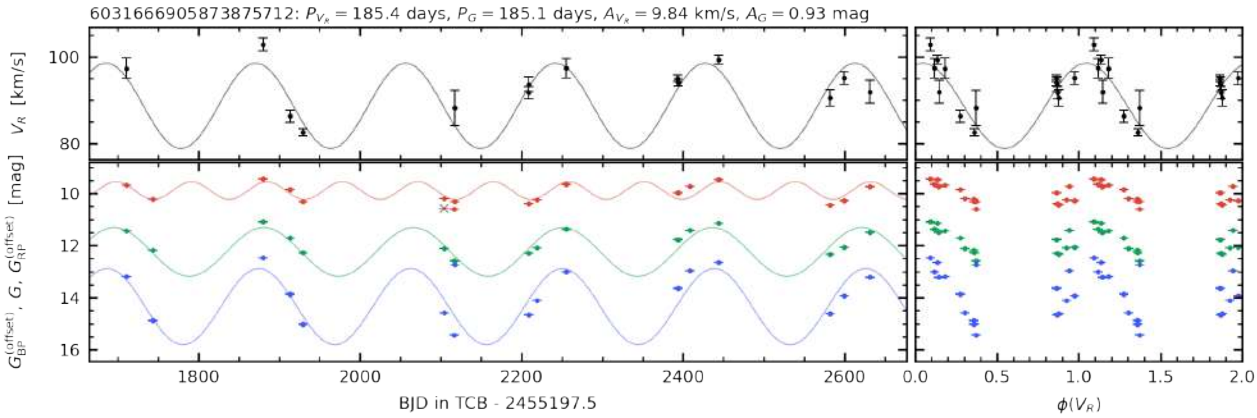

Fig. 9. Example time series for a source with mixed consistency between the photometric and RV time series. This source has $P_G \simeq P_{G_{BP}} \simeq P_{V_R}$, while $P_{G_{RP}} \simeq 0.5 P_{V_R}$. The panels in the top row show the RV data and model, while the photometric data and corresponding models are shown in the panels in the bottom row (in red, green, and blue for the G_{RP}, G, and G_{BP} bands, respectively). For visualization purposes, an arbitrary offset is applied to the G_{RP} and G_{BP} time series. The *Gaia* DR3 source ID of this object is indicated in the title, together with the period and semi-amplitude of the best-fit G-band and RV time series models. The panels on the right show the four time series folded by the RV period.

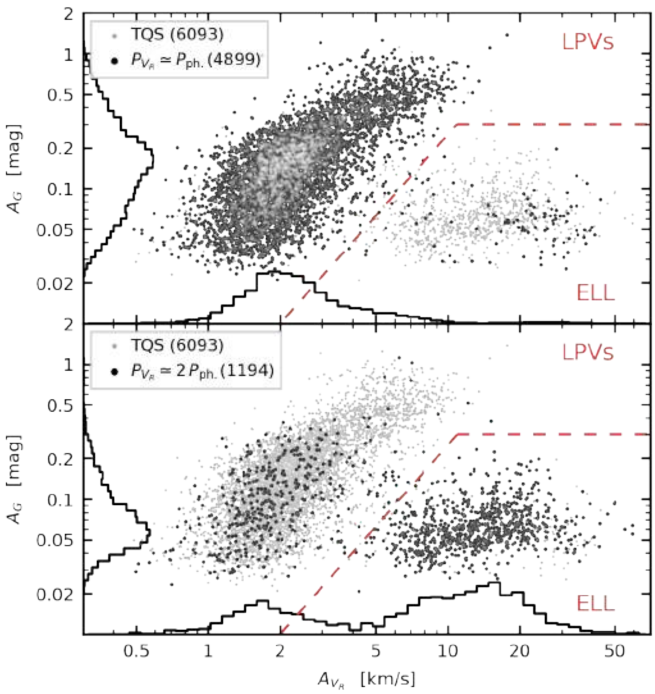

Fig. 10. Comparison between the semi-amplitudes A_G, A_{V_R} of the best-fit models of the G-band and RV time series for the TQS (light gray symbols in the background). The darker symbols indicate sources whose RV period is consistent with the photometric periods in a 1:1 ratio (top panel) or in a 2:1 ratio (bottom panel). The dashed red line corresponds to Eq. (9), and the size of each sample is indicated in the legend.

Based on the distributions displayed in Fig. 10, we identify ELL candidates by the condition

$$A_G < 0.3\,\text{mag} \quad \text{and} \quad \frac{A_G}{\text{mag}} < 2.5 \times 10^{-3} \cdot \left(\frac{A_{V_R}}{\text{km s}^{-1}}\right)^2 \tag{9}$$

which corresponds to the region in Fig. 10 below and to the right of the dashed red line. We prefer Eq. (9) to a condition based on the RV-to-photometric period ratios as it is based on physical arguments, and allows us to identify ellipsoidal variables more confidently. For instance, we note that there are several sources in Fig. 10 (top panel) having $20 \lesssim A_{V_R}/\text{km s}^{-1} \lesssim 50$ that are unlikely to be pulsators, but would be classified as such based only on the ratio between their RV period and photometric periods. Further evidence in support of this approach is given in Sect. 5.

At the same time, there are sources with $P_{V_R} \simeq 2\,P_{\text{ph}}$ that end up outside of the region associated with ELLs in Fig. 10 (bottom panel). While there is no a-priori reason why they should not be binaries, their distribution in this diagram is consistent with that of the sources with $P_{V_R} \simeq P_{\text{ph}}$ (top panel of Fig. 10), suggesting that they display the same kind of variability. Visual inspection of their RV and light curves indicates that the 2:1 period ratio is probably artificial. This most likely results from the fact that many stars in this part of the diagram are semi-regular variables with multi-periodic variability, not necessarily well-described by a single-period model.

3.2. Candidate LPVs: Pulsation and long secondary periods

For pulsating LPV stars the photometric amplitude of variability increases with the pulsation period, a trend that can be identified in the left and top sides of the diagram in Fig. 11 ($20 \lesssim P_G/\text{days} \lesssim 500$). A second group of stars can be seen in the bottom-right corner of the diagram, characterized by long periods and comparatively small G-band amplitudes. While it is likely that these sources are also pulsating LPVs, the dominant period picked up by the variability processing pipeline is probably a long secondary period (LSP; see, e.g., Fig. 16 of Lebzelter et al. 2023).

We tentatively identify LSPs in the period amplitude diagram by the condition

$$A_G < 0.35\,\text{mag} \quad \text{and} \quad \frac{A_G}{\text{mag}} < 10^{-7} \cdot \left(\frac{P_G}{\text{days}}\right)^{2.5}. \tag{10}$$

We remark that this criterion and the resulting classification is necessarily approximate, and is adopted only for the purpose of characterizing the content of the FPR. In principle, a knowledge of the absolute brightness is required in order to accurately differentiate between pulsation periods and LSPs so that one can construct a period-luminosity diagram. This cannot be done for the entire FPR sample because of the relatively large

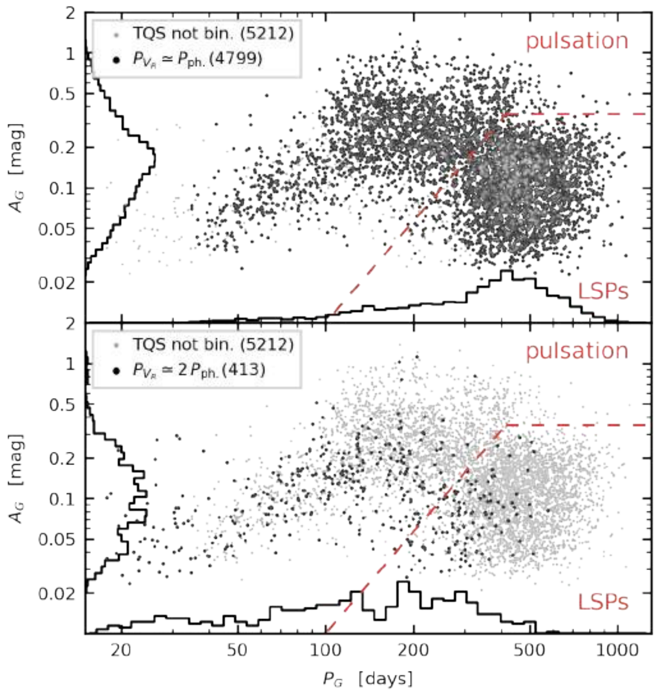

Fig. 11. Similar to Fig. 10, but comparing the period P_G and semi-amplitude A_G derived from G-band time series for TQS sources that are probable LPVs (not identified as ellipsoidal variable candidates). The dashed red line corresponds to Eq. (10), and the size of each sample is indicated in the legend.

uncertainties affecting the parallaxes of a number of sources, even though more than one third of them have relative parallax errors better than 10% (they are examined in more detail in Sect. 5.2).

We note that certain LPVs, such as some relatively massive AGB stars or red supergiants, have long pulsation periods and relatively small photometric amplitude, and their variability could thus mimic the LSP variation. These stars overlap with LSPs in the period-amplitude diagram (see e.g., Fig. 18 of Lebzelter et al. 2019), and might be misclassified by our criteria. However, given the rarity of these stars, this has a negligible impact in our quantification of the relative fraction of variability types in the FPR. They are further examined in Sect. 5.2.

We also note that pulsation and LSP usually coexist in LPVs that exhibit the LSP phenomenon. The knowledge of multiperiodicity and the amplitude associated with each period can then improve the classification derived from the period-amplitude diagram. However, as only one period is extracted from each time series in this FPR, we consider our selection appropriate enough for our purposes. Additional evidence to support this is provided in Appendix A.

3.3. Summary

Table 2 provides a summary of the number of sources identified as LPVs (either showing pulsation or LSP variability) or as ellipsoidal variables. The TQS consists by about 14% of ELLs, by about 38% of pulsating LPVs, and by about 48% of LPVs for which we detect LSP-like variability. If the conditions defined by Eqs. (9) and (10) are extended to the entire FPR sample, these percentages become about 12%, 42%, and 46%, respectively.

The numbers in Table 2 also show that, overall, the TQS includes roughly 60% of the LPVs (regardless of whether they show pulsation or LSPs), and 80% of the ELL candidates, indicating that the latter enter more easily in the TQS. This is consistent with the fact that the geometric origin of the variability of the latter results in much smoother and regular variations than those presented by the former, increasing the chances of consistency between photometric and RV periods.

It is worth noting that the classification we adopted is generally consistent with ratios between RV and photometric periods. Among the sources in the TQS, 89% of the ones identified as ELLs show a 2:1 ratio between P_{V_R} and $P_{\rm ph}$, whereas about 86% and 96% of the LPV candidates showing pulsation or LSP variability, respectively, are consistent with a 1:1 ratio. The corresponding percentages concerning the full FPR are roughly consistent with these values, although some differences arise due to the occurrence of sources showing mixed consistency. Some example time series of ELL, pulsating LPVs, and LSPs from the TQS are displayed in Figs. 12–14, respectively.

4. Catalog quality

In the present section we check the quality of the FPR catalog. We examine the average RVs in Sect. 4.1, then we consider the RV variability in Sect. 4.2, and finally we compare with literature data in Sect. 4.3.

4.1. Radial velocity estimates

4.1.1. Median radial velocity: Consistency with *Gaia* DR3

The values $V_R^{\rm DR3}$ of RV published in *Gaia* DR3 are derived with two different methods depending on the source brightness. In particular, if a source has $G_{\rm RVS}$ brighter than 12 mag (as is the case for all FPR sources), $V_R^{\rm DR3}$ is computed as median values over the RV time series (Katz et al. 2023). It is worth comparing these values with the median values $\langle V_R \rangle$ resulting from the variability processing pipeline employed for the FPR sample (the field `median_rv` in the table `gaiafpr.vari_rad_vel_statistics`). Some small deviations are expected between the two values because of slight differences in terms of data and methods. Indeed, the definition of median value adopted by the *Gaia* variability processing unit and by the *Gaia* spectroscopic data processing unit are slightly different. To compute the median value of a dataset consisting of an odd number of values, both methods sort the data and take the middle value. In contrast, if the number of values is even, after sorting the variability processing pipeline takes the smaller of the two middle values, whereas the spectroscopic processing pipeline takes the mean of the two middle values. This means that the former method systematically results in smaller median values than the latter. Moreover, compared to the RV time series used to compute $V_R^{\rm DR3}$, the time series published in this FPR can have one fewer epoch as a result of outlier removal during the variability processing (Sect. 2.2).

In summary, we find differences between $\langle V_R \rangle$ and $V_R^{\rm DR3}$ for 305 sources whose RV time series had one outlier removed, and for 4672 sources whose time series had no outlier removed, but have an even number of epochs. These sources are displayed in Fig. 15, where the absolute value of the difference between $\langle V_R \rangle$ and $V_R^{\rm DR3}$, normalized to the uncertainty ε_{V_R} on $V_R^{\rm DR3}$, is shown against the number of epochs in the FPR time series. For the majority of the sources we find $|\langle V_R \rangle - V_R^{\rm DR3}| \simeq 0.2\,\varepsilon_{V_R}$, whereas only 104 sources have a difference between $\langle V_R \rangle$ and $V_R^{\rm DR3}$ that exceeds ε_{V_R}. These sources are usually characterized

Table 2. Number of sources assigned to different types (LPVs showing pulsation or LSP variability, or ellipsoidal variables) in the TQS and in the full FPR, distinguishing between the stars showing compatibility between the RV and photometric periods in a 1:1 or 2:1 ratio.

		TQS			FPR		
		All $P_{V_R} \simeq P_{\rm ph}$	All $P_{V_R} \simeq 2P_{\rm ph}$	Total	Any $P_{V_R} \simeq P_{\rm ph}$	Any $P_{V_R} \simeq 2P_{\rm ph}$	Total [(a)]
LPV	Pulsation	1992	304	2296	3266	906	4084
	LSP	2807	109	2916	4125	370	4421
ELL		100	781	881	160	966	1109
Total		4899	1194	6093	7551	2242	9614

Notes. [(a)]Sources that do not belong to the TQS may show photometric periods that are simultaneously compatible with P_{V_R} in both the 2:1 and 1:1 ratios, so the corresponding "total" column is not necessarily equal to the sum of the two previous columns.

by a relatively large RV amplitude or a small number of RV measurements. It is easy to see how such features can enhance the impact of outlier removal and methods differences on the calculation of the median RV, especially given the irregular time sampling of *Gaia* observations.

4.1.2. Mean and systemic radial velocities

The uneven sampling of *Gaia* time series also means that the median RV is not necessarily a good indicator of the systemic RV, that is the center-of-mass velocity, an accurate estimate of which generally requires the RV variability to be modeled. In principle, the zero-point V_R^0 of the RV time series model (i.e., the constant term c_{0,V_R} in Eq. (2), unpublished but computable with the published time series) is more representative of the systemic RV than an average over the time series. In order to assess this, we examine the fractional deviations between $\langle V_R \rangle$ and V_R^0 by taking the absolute value of their difference and scaling it to $|V_R^0|$. The distribution of this quantity is displayed in Fig. 16, together with the fractional difference with respect to V_R^0 of $V_R^{\rm DR3}$ and of the mean value $\overline{V_R}$ derived from variability processing. This diagram shows that in most cases the median values are compatible with V_R^0 to within few percents, whereas the mean value performs slightly better, typically within less than a 1%. Both indicators are consistent with V_R^0 within about 10% for the bulk of the sample.

As is discussed in Sect. 4.2, a subset of the FPR sources are also present in the *Gaia* DR3 non-single stars table, which provides center-of-mass velocities based on orbital solutions. We include in Fig. 16 a comparison between these values and V_R^0, that are found to be in excellent agreement, being usually compatible within 0.03%. We leave further comparisons with the non-single stars results for the next section.

While such fractional difference is helpful to have a quick glance at the general degree of agreement between various indicators, it fails to characterize the stars having a very small systemic RVs (comparable with or smaller than the RV uncertainty) which lead to the tail of the distribution at large values, as can also be appreciated from Fig. 17. Therefore, we further examine the sample by scaling the difference $|\langle V_R \rangle - V_R^0|$ by the mean value $\overline{\varepsilon}_{V_R}$ of the uncertainties on individual RV epochs on a given time series. The resulting quantity is displayed in Fig. 18 as a function of the semi-amplitude of the RV time series model. Its distribution peaks around one, indicating that for the majority of the time series the value $\langle V_R \rangle$ is compatible with V_R^0 within the mean RV uncertainty. However, there is a clear tendency for this compatibility to degrade with increasing RV amplitude. This is primarily due to the uneven and irregular time sampling of *Gaia* observations. While this is rarely an issue for sources that display little or no variability, it is something to be kept in mind when dealing with LPVs and ELLs, because:

- they have periods comparable with the observational baseline, so the *Gaia* time series may cover only a small number of variability cycles;
- they have large variability amplitudes, and so observations taken at a random phase of the variability cycle may be very far from the central value;
- the *Gaia* scanning law is such that a large number of observations may be concentrated within an interval of time much shorter than the typical time scales of LPVs and ELLs; these "clusters" of epochs have a very high statistical weight, but carry little more information than a single measurement taken in the middle of that time interval.

The combination of these effects increases the chance of selectively over-sampling or under-sampling a specific phase of the variability cycle, thereby skewing the median away from the mid-point of the RV curve. For similar reasons, the amplitude of variability may be underestimated by statistical indicators such as the standard deviation or the interquartile range. Some examples of RV curves showing these effects are provided in Appendix B.

We note that the top-quality sample shows a slightly larger difference, on the mean, between $\langle V_R \rangle$ and V_R^0 than the full FPR sample. This is due to the presence of a larger fraction of ellipsoidal variables in the former sample, which have large RV ranges (see Sect. 3 and Fig. 19).

Finally, we recall that the epoch RVs are measured from RVS data by comparing observed spectra with synthetic stellar spectra used as templates (see Sect. 6.4.8 of the *Gaia* DR3 documentation; Sartoretti et al. 2022). In this FPR the templates have $T_{\rm eff}$ ranging from 3100 K to 7500 K (typically between 3300 K and 3800 K) and log(g) in the range –0.5 to 5.0, and are restricted to O-rich stars (Katz et al. 2023). As LPVs can have stellar parameters outside these ranges and include C-stars, the matched template is not necessarily the most appropriate in terms of atmospheric parameters $T_{\rm eff}$ and log(g), which might impact the derived zero-point and median RVs. Yet, the analysis of the overall features of the FPR sample does not highlight any clear systematic discrepancy with respect to what would be reasonably expected. For instance, the distribution of RV as a function of galactic longitude (Fig. 20) does not show a wider spread for C stars than for O-rich LPVs. Likewise, the sky map presented in Sect. 5 gives essentially the same picture as Fig. 5 of Katz et al. (2023), despite being limited to a selection of stars (i.e., cool red giants and including C-stars) that are substantially more exposed to the risk of template mismatch than the bulk of

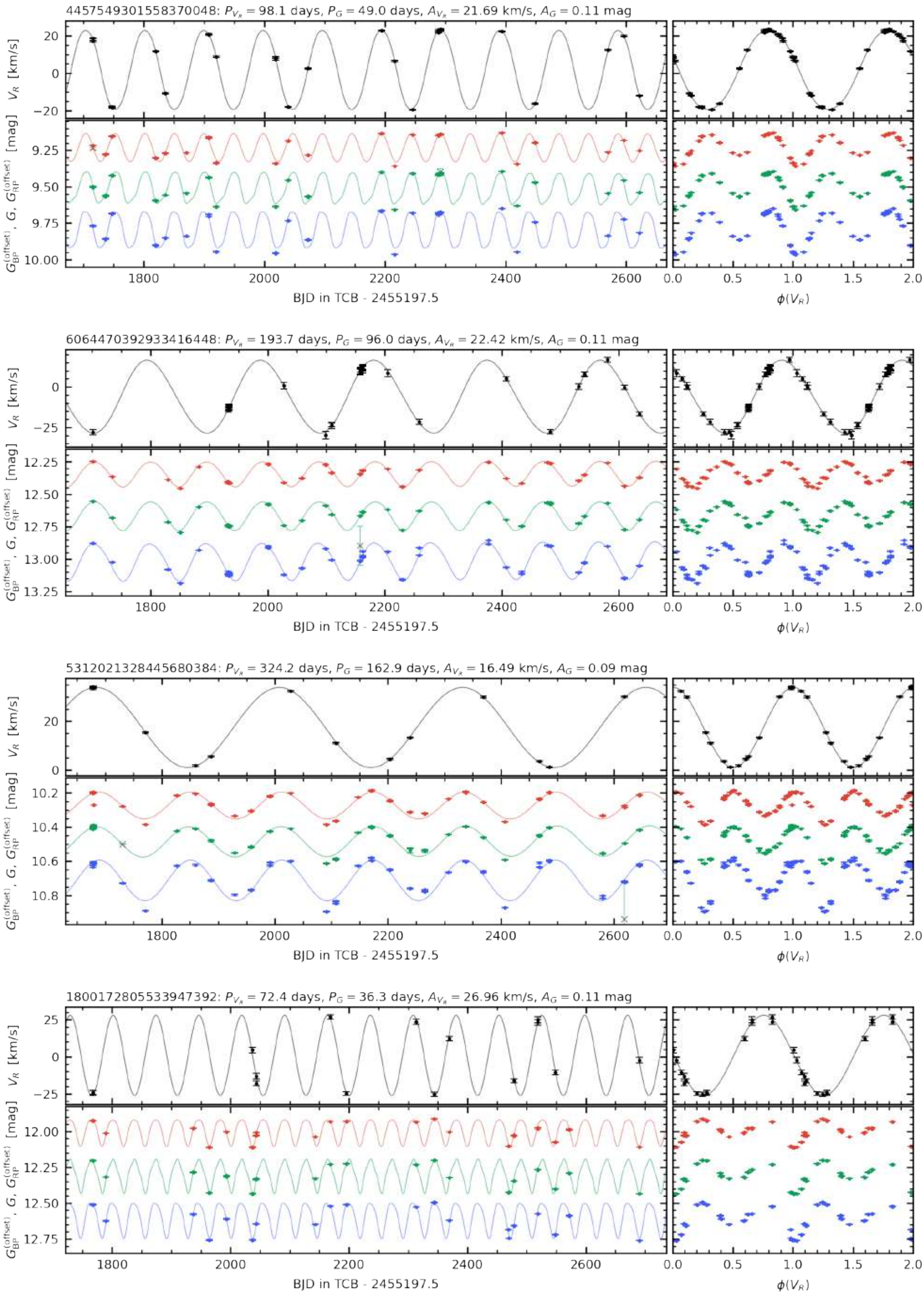

Fig. 12. Similar to Fig. 9, but showing some example time series of ELL candidates. All data in the panels on the right-hand column are folded by the FPR RV period.

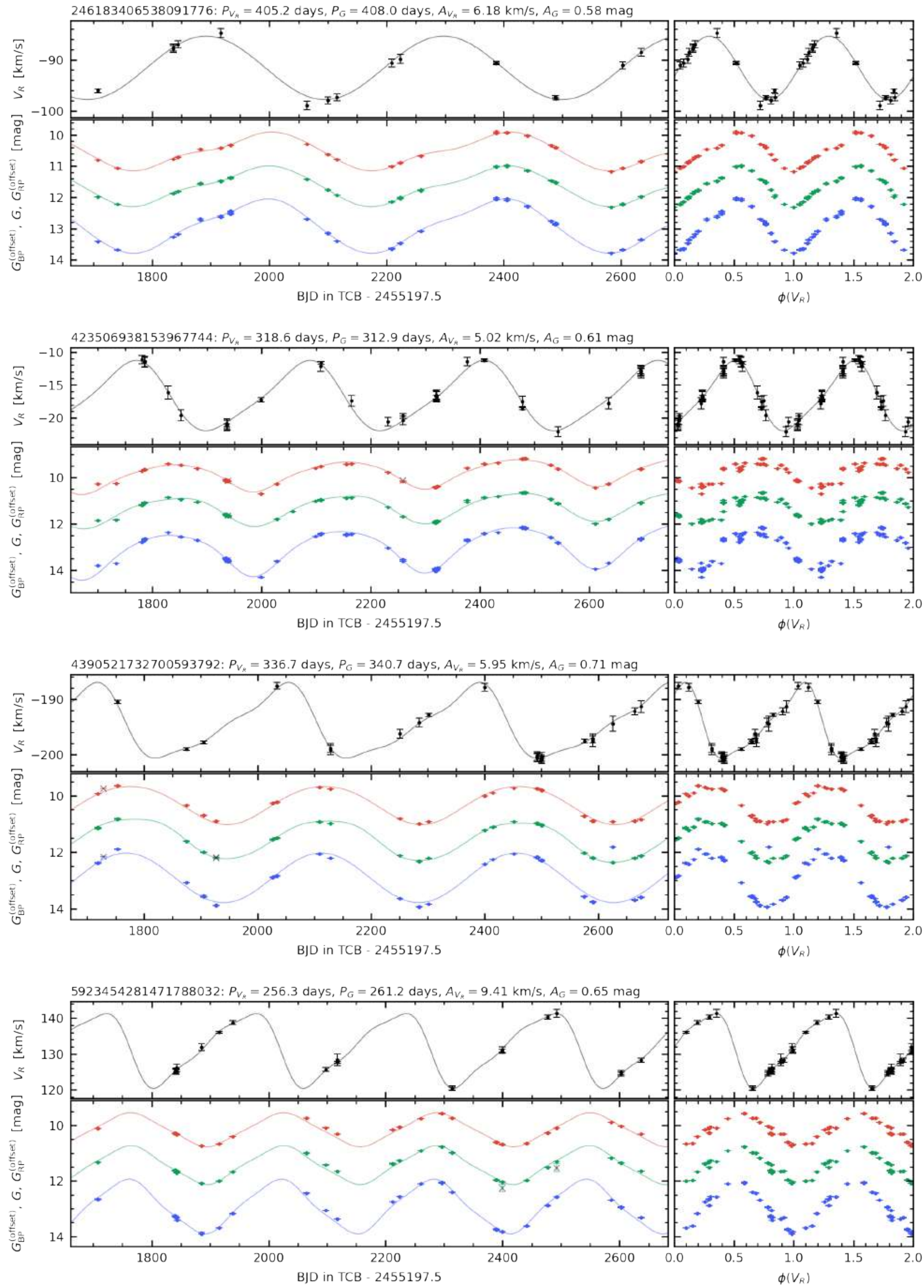

Fig. 13. Similar to Fig. 9, but for pulsating LPV candidates. All data in the panels on the right-hand column are folded by the FPR RV period.

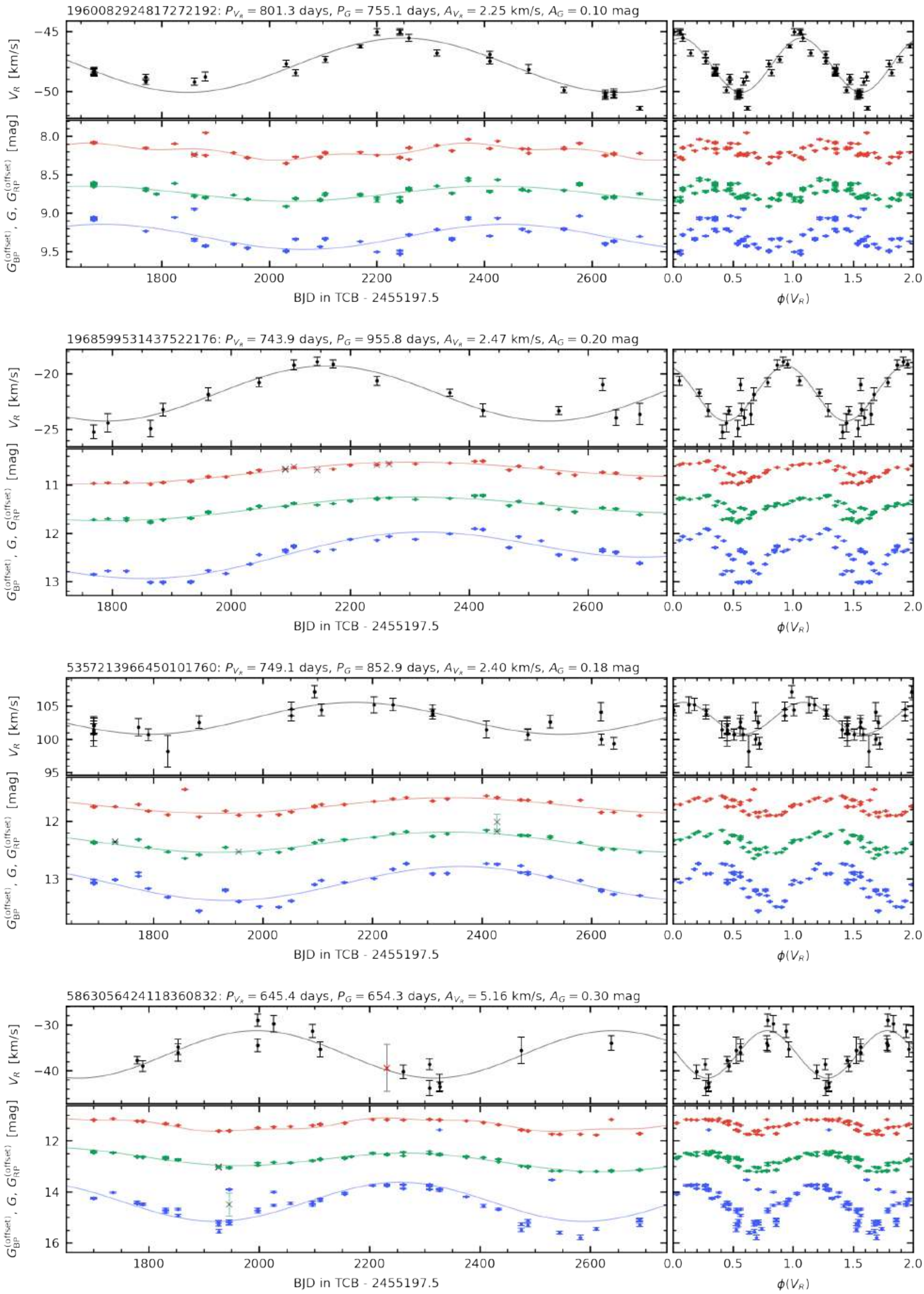

Fig. 14. Similar to Fig. 9, but for LPV candidates for which we likely detect a LSP. All data in the panels on the right-hand column are folded by the FPR RV period.

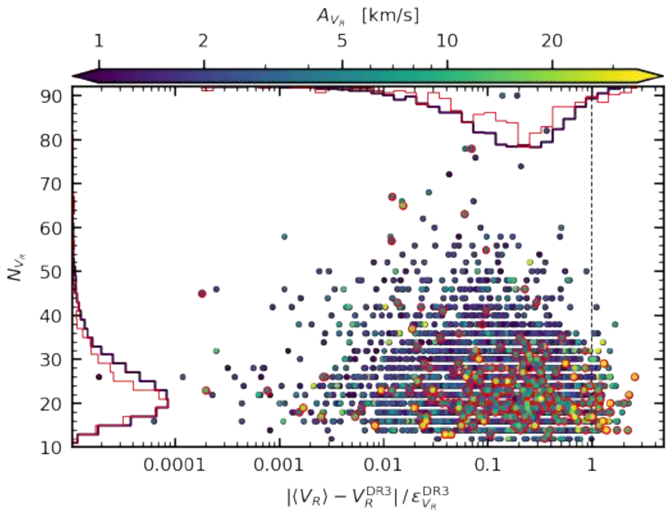

Fig. 15. Number N_{V_R} of epochs retained for RV variability processing against the absolute difference between the median RV derived by variability processing ($\langle V_R \rangle$) and published in *Gaia* DR3 (V_R^{DR3}), scaled to the RV uncertainty ε_{V_R} published in DR3. Data points are color-coded by the semi-amplitude of the RV time series model. The time series that had one RV epoch excluded by outlier removal during variability processing are circled in red. The value $\langle V_R \rangle$ for these sources is computed from one fewer epoch compared to V_R^{DR3}.

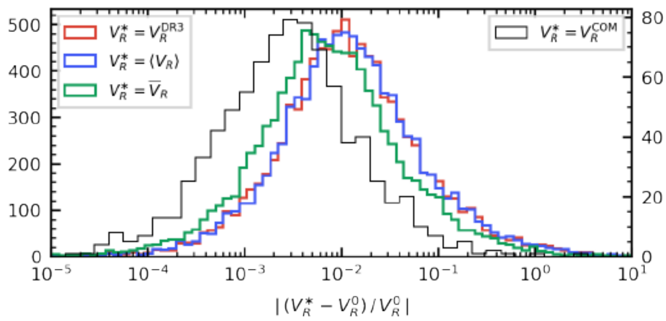

Fig. 16. Comparison of several average RV indicators with the zero point of the RV time series models. Different indicators are displayed in different colors (red: median value V_R^{DR3} published in *Gaia* DR3; blue: median value $\langle V_R \rangle$ computed by variability processing; green: mean value $\overline{V_R}$ computed by variability processing). The histograms show the distribution of absolute difference between each of the average values and V_R^0, normalized to the latter. The thin black histogram, limited to a subset of the FPR sample, compares V_R^0 with the center-of-mass velocity V_R^{COM} derived by the non-single stars processing pipeline for *Gaia* DR3 (see Sect. 4.1.2 for more details), and refers to the scale on the right-hand side axis.

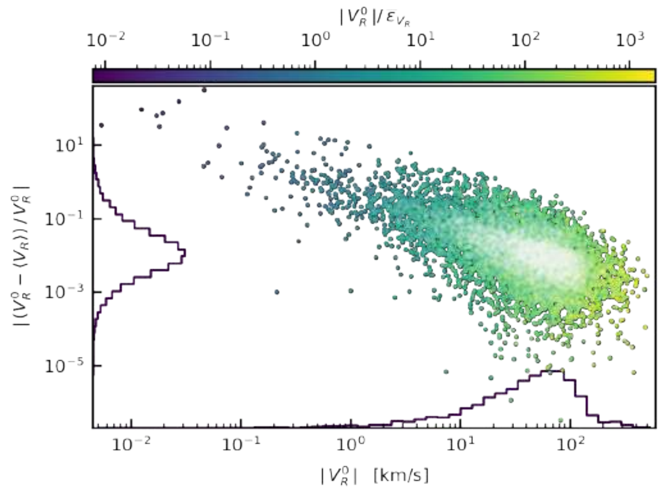

Fig. 17. Absolute difference between the median value $\langle V_R \rangle$ and the zero point V_R^0 RV estimates, scaled to the latter and shown against the absolute value $|V_R^0|$ of the latter. Data points are color-coded by the ratio $|V_R^0|/\overline{\varepsilon}_{V_R}$, showing that large discrepancies (top portion of the diagram) are associated with absolute values of the systemic RV comparable with or smaller than the RV uncertainty. A white shading indicates a more densely populated area of the diagram.

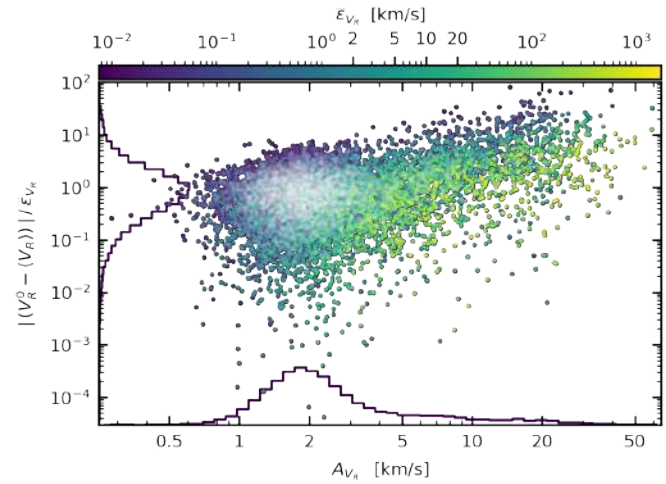

Fig. 18. Absolute difference between the median value $\langle V_R \rangle$ and the zero point V_R^0 RV estimates, scaled to the mean value of the uncertainties on individual RV epochs, and shown against the semi-amplitude A_{V_R} of the RV time series model. A white shading indicates a more densely populated area of the diagram.

sources used by these authors in their figure. More importantly, the variability properties of the RV curves, that are the object of this FPR, are not expected to be significantly affected by the template mismatches, as all epoch RV measurements should be equally impacted by the mismatch. The LPVs undergoing large changes in T_{eff} throughout the pulsation cycle may represent an exception as a different spectral template should be adopted for different epochs, but it is unclear how this might impact the RV variability data. Visual inspection of the RV curves and comparison with their photometric light curves give further support to these considerations.

4.2. Radial velocity variability

For the purpose of assessing the quality of the periods and amplitudes derived from RV time series of sources that are identified as ellipsoidal binary stars, we compare the FPR data with the results of non-single star (NSS) processing from *Gaia* DR3. In particular, we consider the data from the `nss_two_body_orbit` table from the *Gaia* archive (Gosset et al., in prep.). We find that 855 of the FPR sources are also found in that table. Based on the classification outlined in Sect. 3, we identify 353 of them as ELL candidates, 296 as pulsating LPVs, and 205 as LPVs with a LSP.

We compare our RV periods with the NSS values in Fig. 21, which displays the absolute difference between the two periods scaled to the latter. In most cases we find a good degree of period compatibility (typically within a few percents), even for

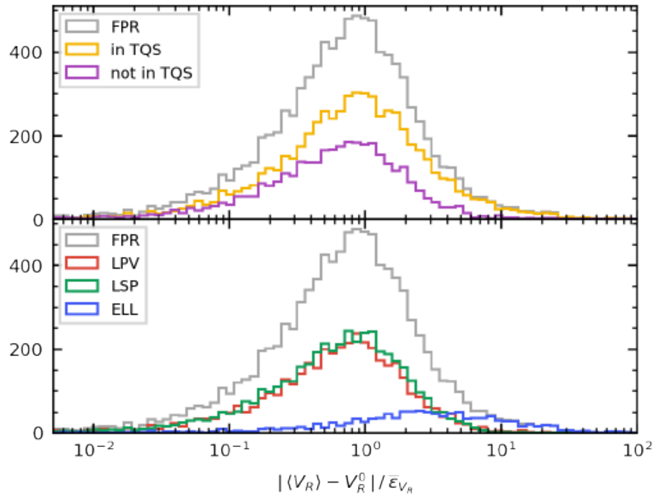

Fig. 19. Absolute difference between $\langle V_R\rangle$ and V_R^0, scaled by the mean of epoch RV uncertainties, for various subsets of the FPR sample. In the top panel, the orange curve corresponds to sources flagged for high consistency between RV and photometric periods, whereas all other sources are represented by the purple curve. In the bottom panel, the red, green and blue curve correspond to sources tentatively identified as pulsating LPVs, LPVs showing LSP variability, or ellipsoidal variables, respectively (see Sect. 3). The gray curves in both panels represent the whole FPR sample.

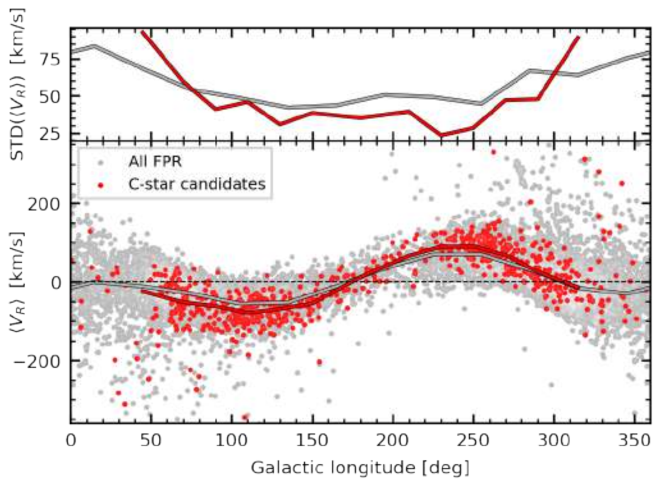

Fig. 20. Median RV ($\langle V_R\rangle$) of FPR sources as a function of their galactic latitude. Red sources are C-star candidates. The solid lines indicate median values over bins of galactic latitude. The curves show the median value and standard deviation of $\langle V_R\rangle$ in bins of galactic latitude (for C-stars these statistics are limited to galactic longitudes between 30° and 330°, excluding regions where they are scarce).

the sources that we do not classify as ELL candidates, although the agreement is better for the latter. In general, the comparison is slightly better for objects identified as pulsating LPVs compared with the LSP candidates. For sources that we identify as ellipsoidal variables, the periods are always within 10% of each other, and typically compatible within 0.1%. There is only one exception (whose time series are displayed in Fig. 22) for which we derive a 428.7 days period in stark difference with the NSS period of 0.35 days. As the RV time series only covers 12 visibility periods, it is hard to conclude which period is more realistic. Both values correspond to a strong signal in the periodogram, but we do not detect the latter as our processing is limited to periods

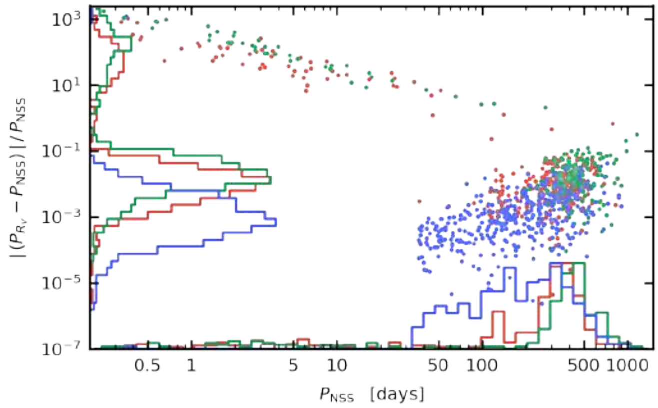

Fig. 21. Relative difference between the period P_{V_R} estimated by variability processing and the value derived from the non-single star pipeline, relative to the latter, for the FPR sources that have a counterpart in the *Gaia* DR3 table of orbital parameters of non-single stars. Data points are color-coded by variability type according to the classification presented in Sect. 3 (blue: ellipsoidal variables; red: pulsating LPVs; green: LPVs with LSP).

longer than 10 days. In any case, periods as short as 0.35 days are not expected for ellipsoidal red giants. A similar situation occurs for a number of other sources for which NSS results in relatively short periods, and hence there exists a large difference with respect to the FPR period (sequence of points in the upper part of Fig. 21 extending to low NSS periods). However, none of these sources is identified as a ELL candidate according to our classification criteria, which casts some doubt on the validity of the orbital model adopted for modeling their RV time series for the NSS processing.

We perform a similar analysis in terms of the RV semi-amplitude, comparing the value derived by the variability pipeline with the NSS results (Fig. 23). We find a qualitatively similar picture as for the periods, the compatibility between the NSS and our semi-amplitudes being typically within 2–3% for ELL candidates, 10% for pulsating LPV candidates, and 20% for LSP candidates. It is interesting to observe that the RV curve asymmetries frequently displayed by pulsating LPVs are often accounted for by a large eccentricity in the binary models adopted by NSS.

Finally, we note that the NSS data for the matched sources result from the assumption of a single-lined spectroscopic binary model, with the only exception of the source 2567779977831471232 for which an orbital astrometric binary model was adopted, obtaining a 928.7±85.3 days period. According to the classification presented in Sect. 3, we also identify that source as a binary (ELL) candidate, but we find a RV period significantly shorter (P_{V_R} = 695.3 days), although we note that the RV observation covers just about 800 days, hence the period cannot be accurately constrained.

4.3. Comparison with RV data from literature

We searched the literature for RV data to compare with this FPR, and found them to fall into two main categories. The first one concerns large-scale spectroscopic surveys providing extensive catalogs of RV data. The chances of finding matching objects against these source lists are relatively high, but at the same time they only allow for comparing average RVs, as they are based on single-epoch observations or a few epochs at most, and rarely

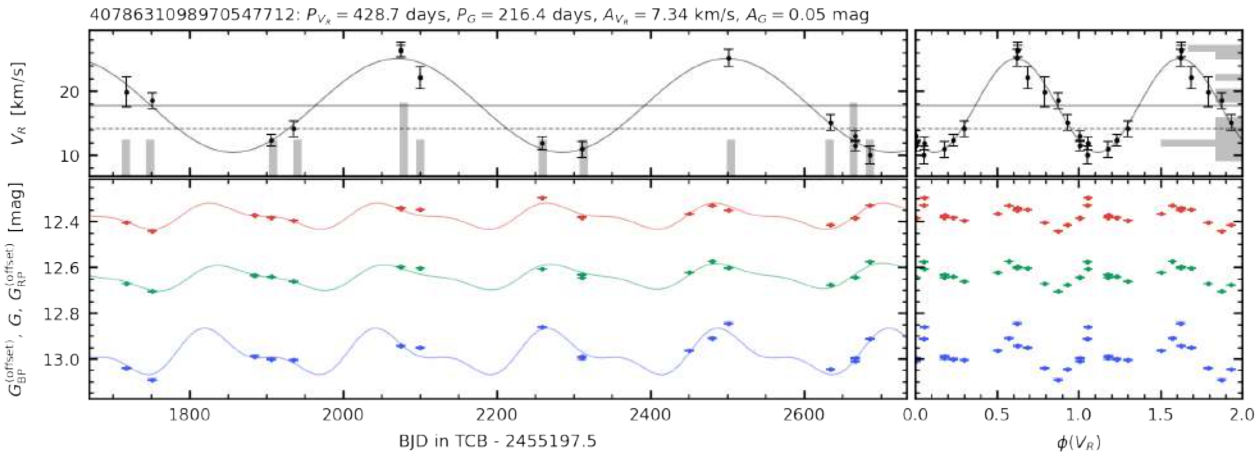

Fig. 22. Similar to Fig. 9, showing the *Gaia* RV and photometric curves of the ELL candidate showing a large mismatch between the FPR RV period (428.7 days) and the value published in the *Gaia* DR3 NSS table (0.35 days). For the purpose of visualization, the $G_{\rm BP}$ and $G_{\rm RP}$ time series are offset by an arbitrary amount. All data in the panels on the right-hand column are folded by the FPR RV period.

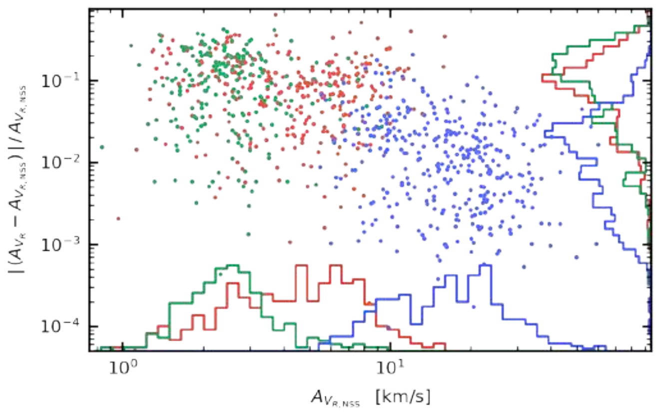

Fig. 23. Similar to Fig. 21, but comparing the RV semi-amplitudes rather than the periods.

provide RV time series. In Sect. 4.3.1 we present a comparison with a few such catalogs.

Besides these surveys there exist a number of smaller-scale observational programs targeting specific types of stars or fields of the sky. These studies involve a more focused analysis and validation of the RV data of the targets, and often result in the publication of the time series. We attempted to cross-match the FPR catalog with the source lists from various such literature works, but found only a few matches, that are examined in Sect. 4.3.2.

4.3.1. Comparison with RV surveys

We compare the FPR results with the data published by three spectroscopic surveys: the Apache Point Observatory Galactic Evolution Experiment survey (APOGEE, data release 16, Jönsson et al. 2020), the GALactic Archaeology with HERMES survey (GALAH, data release 3, Buder et al. 2021), and the RAdial Velocity Experiment survey (RAVE, data release 6, Steinmetz et al. 2020). For the former two, we rely on the cross-match with *Gaia* DR3 performed as part of their data release process (they provide best-match *Gaia* DR3 source IDs), whereas for the latter we adopt the results of the *Gaia* cross-match with external catalogs (`gaiadr3.ravedr6_best_neighbour` table in the *Gaia* archive).

Table 3. Number of FPR sources cross-matched with external spectroscopic surveys.

	FPR	APOGEE	GALAH [(a)]	RAVE
LPV	4084	34	58 (65)	503
LSP	4421	41	89 (95)	643
ELL	1109	14	20 (21)	80
Total	9614	89	167 (181)	1 226
in TQS	6093	60	108 (119)	784

Notes. [(a)]Numbers indicate the sources with a published value of RV in GALAH, while the total number of matches is given in parentheses.

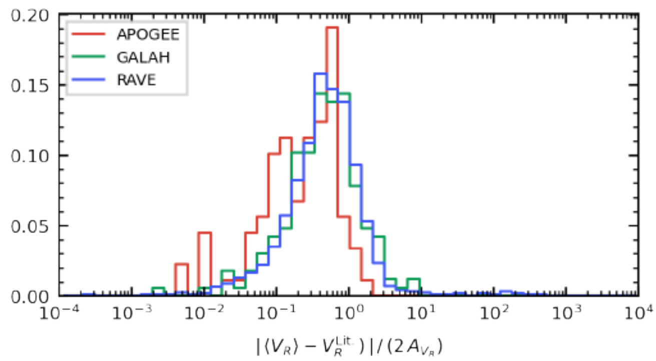

Fig. 24. Comparison between the median RV from this FPR with the average values provided by external catalogs (red: APOGEE; green: GALAH; blue: RAVE). We consider the absolute difference between the two values, normalized to the peak-to-peak model amplitude. We note that the histograms are normalized to their area.

A summary of the numbers of matched sources is provided in Table 3. We find 89 matches with APOGEE, 167 with GALAH, and 1226 with RAVE (of which 60, 119, and 784, respectively, are in the TQS). In Fig. 24 we provide an overview of the comparison between the median RV $\langle V_R\rangle$ from the FPR and the averages given by each of these surveys. Namely, that figure shows the distribution of the absolute difference between $\langle V_R\rangle$ and the literature value, divided by the peak-to-peak amplitude of the *Gaia* RV time series model. We employ the median RV rather