Analysis of mass-transferring binary candidates in the Milky Way

G. Garcia-Moreno[1,2] ⓘ, N. Blagorodnova[1,2,3] ⓘ, F. Anders[1,2,3] ⓘ, M. Weiler[1,2,3] ⓘ, H. Wichern[4] ⓘ, N. Britavskiy[5,6] ⓘ, and S. de Wet[4] ⓘ

[1] Departament de Física Quàntica i Astrofísica (FQA), Universitat de Barcelona (UB), c. Martí i Franquès, 1, 08028 Barcelona, Spain
[2] Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona (UB), c. Martí i Franquès, 1, 08028 Barcelona, Spain
[3] Institut d'Estudis Espacials de Catalunya (IEEC), Edifici RDIT, Campus UPC, 08860 Castelldefels (Barcelona), Spain
[4] DTU Space, Technical University of Denmark, Elektrovej 327, 2800 Kgs. Lyngby, Denmark
[5] Royal Observatory of Belgium, Avenue Circulaire, Ringlaan 3, 1180 Brussels, Belgium
[6] University of Liège, Allée du 6 Août 19c (B5C), 4000 Sart Tilman, Liège, Belgium

ABSTRACT

Mass transfer between stars in binary systems profoundly impacts their evolution, yet many aspects of this process—especially the stability, mass loss, and eventual fate of such systems—remain poorly understood. One promising avenue to constrain these processes is through the identification and characterisation of systems undergoing active mass transfer. Inspired by the slow brightening preceding stellar merger transients, we worked on a method to identify Galactic mass-transferring binaries in which the donor is a Hertzsprung gap star. We constructed an initial sample of Hertzsprung gap stars using the *Gaia* EDR3 contribution `Starhorse` catalogue, and we identified candidate mass-transferring systems by selecting sources that exhibit Balmer emission features (as seen in the low-resolution *Gaia* XP spectra), mid-infrared excess (from WISE photometry), and photometric variability (inferred from the error in the *Gaia* G-band magnitude). This multi-criteria selection yielded a sample of 67 candidates, which we further analysed using complementary photometric and spectroscopic data, as well as information from cross-matched archival catalogues. Among our candidates, we identified at least nine eclipsing binaries and some sources that are potential binaries as well. Three sources in our sample are strong candidates for mass-transferring binaries with a yellow component, and three more are binaries with a Be star. Notably, three sources in our sample are strong candidates for hosting a compact companion, based on their ultraviolet or X-ray signatures. The main sources of contamination in our search are hot but highly reddened stars—primarily Oe and Be stars—as well as regular pulsating stars such as δ Scuti and Cepheid variables. As an additional outcome of this work, we present a refined new catalogue of 308 Hertzsprung gap stars, selected using improved extinction corrections and stricter emission-line criteria. This enhanced sample is expected to contain a significantly higher fraction of scientifically valuable mass-transferring binaries.

Key words. Hertzsprung-Russell and C-M diagrams – binaries: close – binaries: eclipsing – stars: emission-line, Be

1. Introduction

A large fraction of stars in the observable Universe are found in binary or multiple systems, with the likelihood of multiplicity increasing with stellar mass (Sana et al. 2012; de Mink et al. 2014; Moe & Di Stefano 2017). The evolution of most of these systems is significantly influenced by mass transfer between the components. How mass transfer unfolds and its eventual outcome depends strongly on the initial configuration of the binary and the nature of the mass transfer process. Broadly, two scenarios are considered: stable mass transfer, in which the mass exchange occurs without disrupting the system, and unstable mass transfer, where the accreting star is unable to absorb all the material from the donor, resulting in the formation of a common envelope that engulfs both stars (Kuiper 1941; Paczynski 1976).

Mass transfer is an accepted explanation for the creation of many astrophysical phenomena, such as the Algol-type binaries (Crawford 1955), the evolution of closely orbiting stellar binaries into X-ray binaries (Verbunt 1993), cataclysmic variable stars (Kraft 1964), supernovae (Wang & Han 2012; Yoon et al. 2010), accretion-induced collapse (Ivanova & Taam 2004), and gravitational wave sources (Schneider et al. 2001), among others. However, the details of mass transfer and the possible mass

and angular momentum loss of the process, as well as the fate of the evolving stars, are not well established and modelled yet.

In recent years, the study of mass transfer has gained renewed attention due to its fundamental link to the formation of compact binaries and, ultimately, gravitational-wave sources. To constrain these processes, detailed studies of systems currently undergoing mass exchange are crucial. These efforts span a wide range of approaches: observational analyses of interacting systems such as TZ Dra (Kahraman Aliçavuş et al. 2022), AU Monocerotis (Armeni & Shore 2022), and the contact binary V2840 Cygni (Pothuneni et al. 2023); modelling of well-studied binaries like β Lyrae A (Mourard et al. 2018; Brož et al. 2021); and theoretical work aimed at understanding the underlying physics of mass transfer and common envelope evolution (Cehula & Pejcha 2023; Podsiadlowski 2001; Ivanova et al. 2001; Marchant et al. 2021; Röpke & De Marco 2023).

The outcome of unstable mass transfer has recently been linked to a class of optical transients known as luminous red novae (LRNe), whose peak luminosities lie between those of novae and supernovae (Munari et al. 2002; Soker & Tylenda 2003; Tylenda et al. 2011; Pejcha 2014; Pejcha et al. 2016). These transients are generally interpreted as the observational signatures of stellar coalescence events, providing rare but direct evidence of binary interaction and merger processes.

Over the last two decades, LRNe have been studied both in our Galaxy and in nearby galaxies. Despite the proximity, the only five confirmed and studied stellar mergers in the Milky Way (MW) include V4332 Sagittarii (discovered in 1994; Martini et al. 1999), V838 Monocerotis (2002; Munari et al. 2002; Bond et al. 2003), OGLE-2002-BLG-360 (2002; Tylenda et al. 2013), CK Vul (1670; Kato 2003), and most notably, V1309 Scorpii (2008, Mason et al. 2010), whose pre-outburst photometry captured the final inspiral of a contact binary system (Tylenda et al. 2011). Additionally, about two dozen extra-galactic LRNe were found (see Pastorello & Fraser 2019, Blagorodnova et al. 2021, Reguitti et al. 2025, and references therein). These events share a number of defining characteristics: a rapid rise and redward evolution in colour, usually a double-peaked light curve followed by a plateau, and the presence of strong and narrow emission lines for hydrogen and low-ionisation elements.

Archival data taken prior to the outburst have shown that most confirmed LRN progenitors are predominantly yellow giants (YG) or yellow supergiants (YSG) with spectral types G, F or, to a lesser extent, A (Blagorodnova et al. 2017; MacLeod et al. 2017; Blagorodnova et al. 2021; Cai et al. 2022; Pastorello et al. 2023). These stars are evolving off the main sequence (MS) and crossing the Hertzsprung gap (HG) on their way to the red giant branch (RGB). This phase is characterised by a rapid increase in the stellar radius of the evolving star. Given the presence of a nearby companion, the donor can overfill its Roche lobe and initiate unstable mass transfer, finally resulting in a stellar merger and a transient. In the few cases where pre-outburst data exist, the progenitor shows a gradual, years-long brightening, interpreted as mass loss and the formation of an extended pseudo-photosphere before coalescence. Despite these key discoveries, LRN progenitors remain largely unexplored (Pastorello & Fraser 2019).

Guided by the location of known LRN progenitors on the Hertzsprung-Russell (HR) diagram, a recent study presented a method for identifying Galactic LRN candidates likely to outburst within the next 1–10 years (Addison et al. 2022). The selection of candidates was based on the position in the HR diagram and the variability of the sources. The study first identified a set of stars falling in a low stellar density area between the MS and the RGB from the *Gaia* DR2 and *Gaia* EDR3, and explored their variability in different time-domain surveys to select slowly brightening sources. From an initial set of approximately 10^7 initial sources, 21 LRN precursor candidates were selected. Most of them shared similar characteristics: the presence of Hα and (sometimes) Hβ emission lines together with the existence of an infrared (IR) excess. One of the limitations of that study was the lack of reliable estimates for Galactic extinction in *Gaia* DR2 and EDR3, implying that the true colours of the sources were rather uncertain.

The release of extinction estimates in *Gaia* DR3 and *Gaia* related catalogues, such as Starhorse (Anders et al. 2022), opened the possibility to revise the previous work with more accurate colours and astrophysical parameters. In this paper, we make use of these catalogues to present a selection and characterisation of HG stars in the Milky Way (MW) with characteristics compatible with known LRN precursors and mass-transferring systems. We first make a tailored selection of candidates, followed by an analysis using archival and new follow-up to find binary systems during the unstable mass transfer phase, which could become LRNe in the near future. In addition, we aim to characterise the variability of stars in the HG to understand what other systems can mimic the behaviour expected for LRN pre-

cursor systems. Finding unstable mass-transferring binaries will be of great interest to the theoretical and observational communities studying the stability of mass transfer and its impact on the final fate of binary stars.

2. Sample selection

Our initial selection strategy made use of *Gaia* DR3 (Gaia Collaboration et al. 2016, 2023) data and the *Gaia* EDR3 StarHorse (Anders et al. 2022) catalogue, in combination with mid-IR photometry from the ALLWISE catalogue (Cutri et al. 2021). The selection steps are listed below:

1. To select HG stars, we used StarHorse, which (among other astrophysical parameters) contains absolute magnitudes in the G band and de-reddened $BP-RP$ colours. Based on *MESA Isochrones & Stellar Tracks* (Paxton et al. 2011; Dotter 2016), we selected sources with absolute magnitudes $M_G < 1.5$, to avoid contamination from lower MS stars with masses $\lesssim 2M_\odot$, and we also applied two cuts to select stars between the MS and the RGB. Finally, we removed a portion of the colour-magnitude diagram populated by Red Clump stars, known red giants in the horizontal branch. In Fig. 1a our initial parameter space is shown.

2. To select sources with emission in Hα or Hβ, we used the method developed by Weiler et al. (2023) to derive the equivalent width (EW) of the lines using the internally calibrated *Gaia* XP spectra (Carrasco et al. 2021; De Angeli et al. 2023) (see Fig. 1b, which shows the XP spectrum of one of our sources alongside the corresponding follow-up spectrum). We first tested this method with the candidates in Addison et al. (2022). The sources with Hα emission had positive EW values (as expected), but some of these sources showed high errors and low statistical significance in our EW calculation, even though they had clear emission in both the *Gaia* XP spectrum and a higher resolution follow-up spectrum. Therefore, we selected sources with a positive Hα or Hβ EW, without taking into account the EW error or the statistical significance.

3. To select sources with mid-IR excess associated with warm dust emission (likely from a disk), we used the ALLWISE catalogue. We applied the following recommended quality cuts[1]: cc_flags = 0000, making sure the sources are unaffected by known artifacts; ext_flag = 0, indicating that the sources' shape is consistent with a point-source, and ph_qual = A, B or C, selecting sources with a flux signal-to-noise ratio better than 2. We selected sources with $W1(3.4\,\mu m) - W4(22\,\mu m) > 1$, based on the distribution of sources in the $W1(3.4\,\mu m) - W4(22\,\mu m)$ vs $W1(3.4\,\mu m) - W2(4.6\,\mu m)$ colour-colour diagram (see Fig. 1c).

4. As the photometry listed in the main *Gaia* DR3 catalogue is a weighted average of the individual measurements taken in different transits (Riello et al. 2021), one expects that variable stars have higher photometric uncertainties (as seen in Andrew et al. 2021 or Maíz Apellániz et al. 2023). Thus, we placed our initial HG sample in the $\log_{10}(G_{err})$ vs G plane (see Fig. 1d) and selected sources above the 90th percentile.

5. As a final step, we removed the sources with a known type in the SIMBAD (Wenger et al. 2000) database, for which the variability was associated with pulsations rather than binarity (such as RR Lyrae or Cepheids).

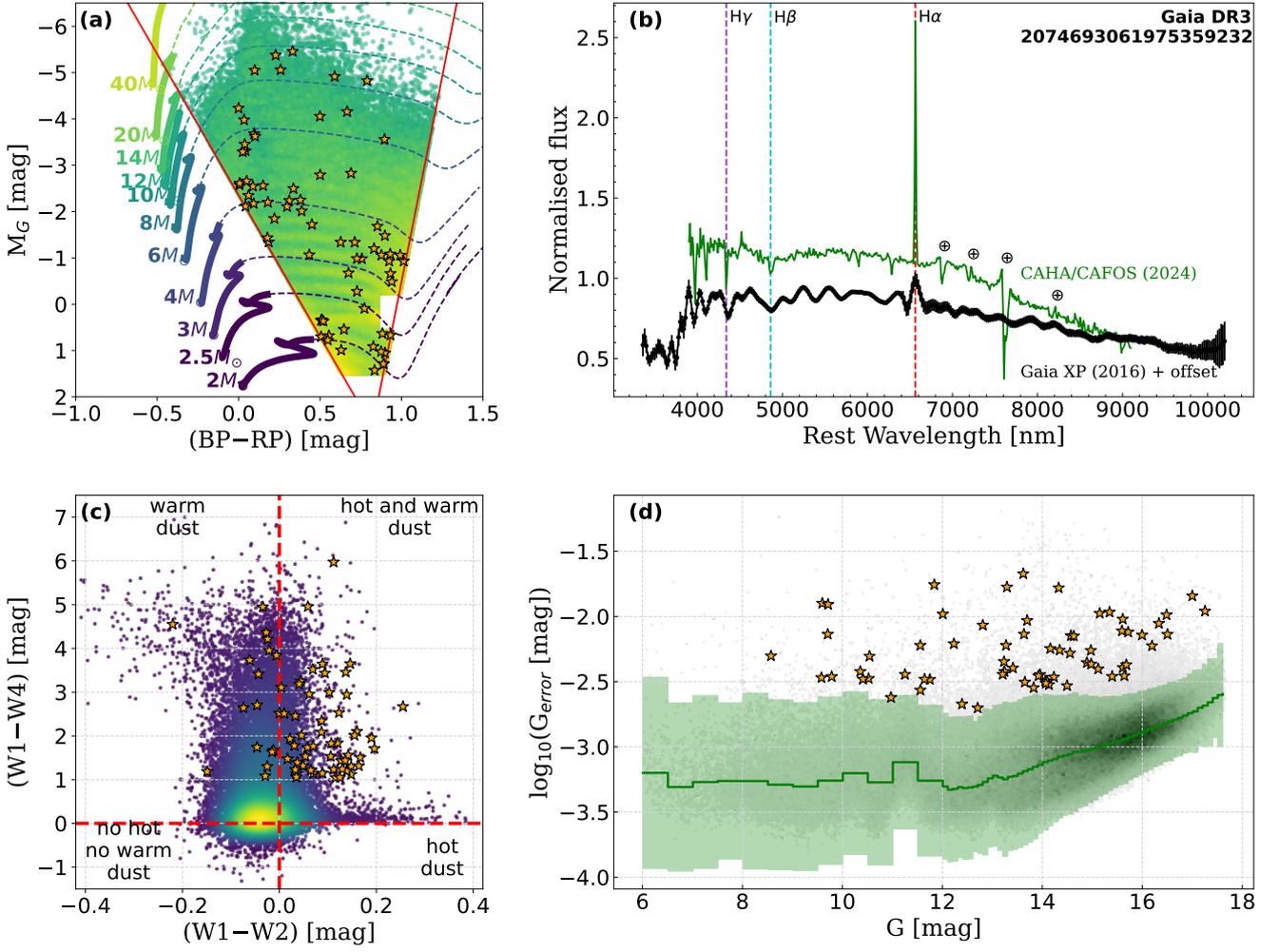[1] https://wise2.ipac.caltech.edu/docs/release/allwise/expsup/sec2_1a.html

Fig. 1: Initial selection of our sample of mass-transferring candidates, based on a) the *Gaia* EDR3 `StarHorse` colour-magnitude diagram, b) the *Gaia* XP spectra, c) the `ALLWISE` colour-colour diagram, and d) the *Gaia G*-band variability. In *a*, *c*, and *d*, we show the final sample of 67 candidates as orange stars. The red lines in *a* show the cuts used to select stars between the MS and the RGB based on *MESA Isochrones & Stellar Tracks*. In *b*, we show the *Gaia* XP spectrum of one of the sources in our sample as an example, and we compare it with our follow-up spectrum taken with CAFOS (see Section 3.2), and the ⊕ symbols represent telluric absorption (absorption by the Earth's atmosphere). The red dashed lines in panel *c* indicate the zero levels and divide the plot into four regions: warm dust, hot dust, both hot and warm dust, and no hot or warm dust. The shaded green region in *d* corresponds to the 0.9 percentile of the initial HG sample.
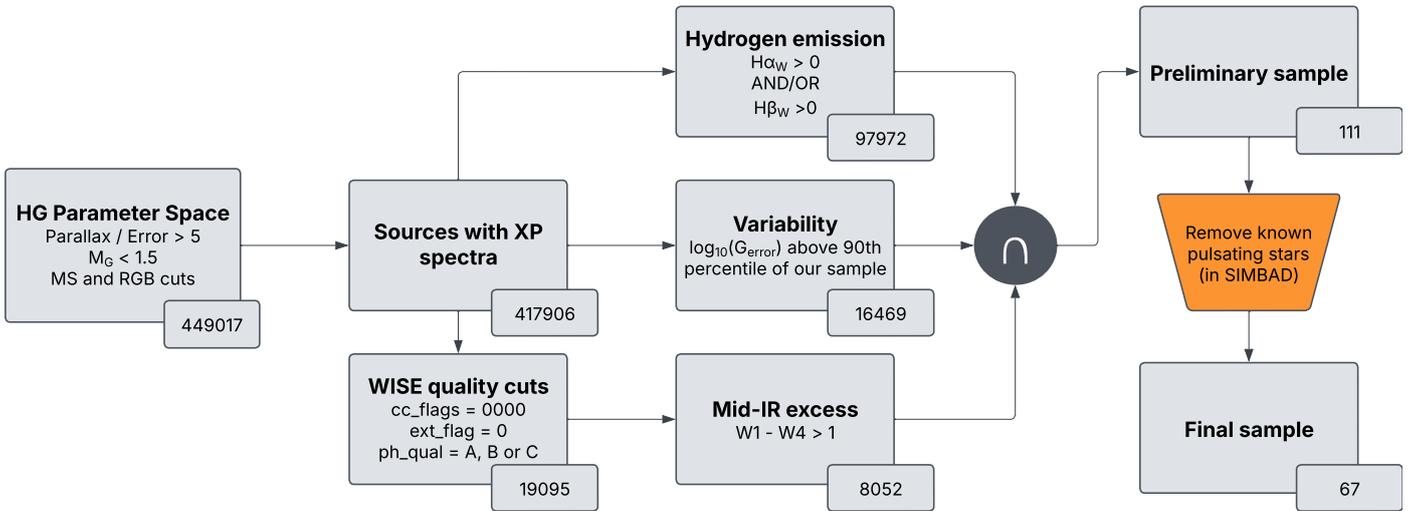

Fig. 2: Overview of the selection process with the number of sources at each stage of the selection.

Table 1: Constraints and parameter space cuts made for the selection of our candidates.

Selection step	Constraint
1	parallax / parallax_error > 5
	$M_G < 1.5$ mag
	$M_{BP} - M_{RP} > (21/128)M_G + (149/386)$
	$M_{BP} - M_{RP} < (-7/124)M_G + (603/620)$
2	$H\alpha_w > 0$ nm OR $H\beta_w > 0$ nm
3	$W1 - W4 > 1$ mag
4	$\log_{10}(G_{err})$ above 90th percentile

Notes. The magnitudes used are the absolute magnitudes provided by `StarHorse`, and $H\alpha_w$ and $H\beta_w$ refer to the equivalent width of the respective lines derived using the internally calibrated *Gaia* XP spectra.
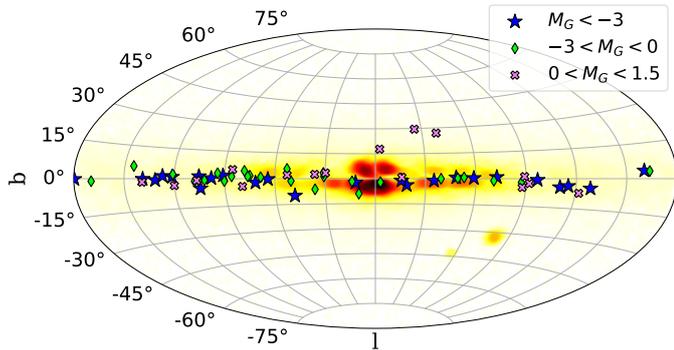

Fig. 3: Sky map showing the position of the 67 sources in our initial sample with reference to the Galactic density of stars, constructed with a *Gaia* DR3 random data set containing 100 000 sources. Different ranges of absolute magnitude are shown with different markers.

We first applied step 1 to select the initial HG sample, then performed steps 2, 3, and 4 on this initial sample separately. Finally, step 5 was applied to the intersection of steps 2, 3, and 4, resulting in our final sample of 67 sources selected for further analysis. This process is summarised in Fig. 2, and a summary of the constraints used is presented in Table 1. The 67 sources in the final sample are highlighted in Fig. 1, and their positions in the sky are shown in Fig. 3. In Table A.1 we gathered information on the targets in the final sample, such as their sky positions, distances and apparent magnitudes. In the following section, we describe the observational data used to characterise this sample.

3. Observational data

3.1. Photometry

In optical wavelengths, we retrieved light curves from the *Zwicky Transient Facility* (ZTF) (Masci et al. 2019), the *All Sky Automated Survey for SuperNovae* (ASAS-SN) (Hart et al. 2023), the *Asteroid Terrestrial-impact Last Alert System* (ATLAS) (Tonry et al. 2018; Heinze et al. 2018), and the *Transiting Exoplanet Survey Satellite* (TESS) (Ricker et al. 2015). In the near-infrared, we retrieved light curves from the *NEO Wide-Field Infrared Survey Explorer* (NEOWISE) (Mainzer et al. 2011). We cleaned all the photometric data using the recommended flags in the documentation of each catalogue. The TESS processed light curves were retrieved from catalogues that used different reduc-

tion pipelines. The main ones used in this work were the *TESS Light Curves From Full Frame Images* (SPOC) (Caldwell et al. 2020) and the *Quick-Look Pipeline* (QLP) (Huang et al. 2020; Kunimoto et al. 2021). In addition, we also inspected TESS light curves from the *TESS Asteroseismic Science Operations Center* (TASOC) (Lund et al. 2021), the *TESS FFI-Based Light Curves from the GSFC Team* (GSFC-ELEANOR-LITE) (Powell et al. 2022), and the *Cluster Difference Imaging Photometric Survey* (CDIPS) (Bouma et al. 2019). In Fig. B.1 we show the light curves of all the sources in our sample, in 50-day bins.

3.2. Spectroscopy

Spectra were obtained using the following instruments:

– The low-resolution instrument CAFOS (Meisenheimer 1994) at the *Calar Alto Observatory* (CAHA) 2.2 m telescope (Sánchez et al. 2007, 2008), with Grism G-200 ($R \sim 300$, 4 000–8 500 Å). We reduced the data using a custom-developed `Python` pipeline, adapted by our team.

– The HERMES high-resolution spectrograph (Raskin et al. 2011) ($R = 85 000$, 3 770–9 000 Å) at the *Mercator* telescope[2] (Proposal ID: 5 (ROB service time), PI: Britavskiy, N.). These spectra were flux calibrated using the flux retrieved from the *Gaia* DR3 low-resolution XP spectra.

– The FIES instrument (Telting et al. 2014) at the the *Nordic Optical Telescope* (NOT) (Djupvik & Andersen 2010), with the medium-resolution setup ($R = 46 000$, 3 700–8 300 Å) and with the low-resolution setup ($R = 25 000$, 3 700–8 300 Å) (Proposal ID: P69-854, PI: Wichern, H.; Proposal ID: P70-203, PI: Garcia-Moreno, G.).

– The ALFOSC instrument at the NOT, with Grism G4 ($R = 710$ and $R = 360$, 3 200–9 600 Å), Grism G7 ($R = 650$, 3 650–7 110 Å), and Grism G19 ($R = 1 940$, 4 400–6 950 Å) (Proposal ID: P70-203, PI: Garcia-Moreno, G.; Proposal ID: P71-209, PI: Blagorodnova, N.). We reduced the data using the `Python` package `PypeIt` (Prochaska et al. 2020).

– Low-resolution ($R = 1 800$, 3 700–9 000 Å) archival spectra from the *Large Sky Area Multi-Object Fibre Spectroscopic Telescope* (LAMOST) (Cui et al. 2012).

– The low-resolution Mookodi spectrograph and imager mounted on the *South African Astronomical Observatory* (SAAO) 1 m Lesedi telescope (Worters et al. 2016; Erasmus et al. 2024), with a 4″ slit ($R \approx 175$, 4 000–8 000 Å). We reduced the data using a Mookodi-specific pipeline developed using the `ASPIRED` toolkit (Lam et al. 2023), which performs standard long slit reductions including trace fitting, spectral extraction, wavelength calibration and flux calibration.

In Fig. 4 we show single epoch spectra of the 26 candidates that we characterised spectroscopically, and the normalised flux of the $H\alpha$ and the $H\beta$ profiles in velocity space is shown in Fig. 5. Details of the observations can be found in Table A.2.

4. Analysis

To verify the spectral type and binary nature of our candidates, we performed a light curve analysis and a spectroscopic analysis, in addition to a multi-wavelength characterisation using ultraviolet (UV) and X-ray catalogues from the literature. In this section, we describe the methodology of this analysis.
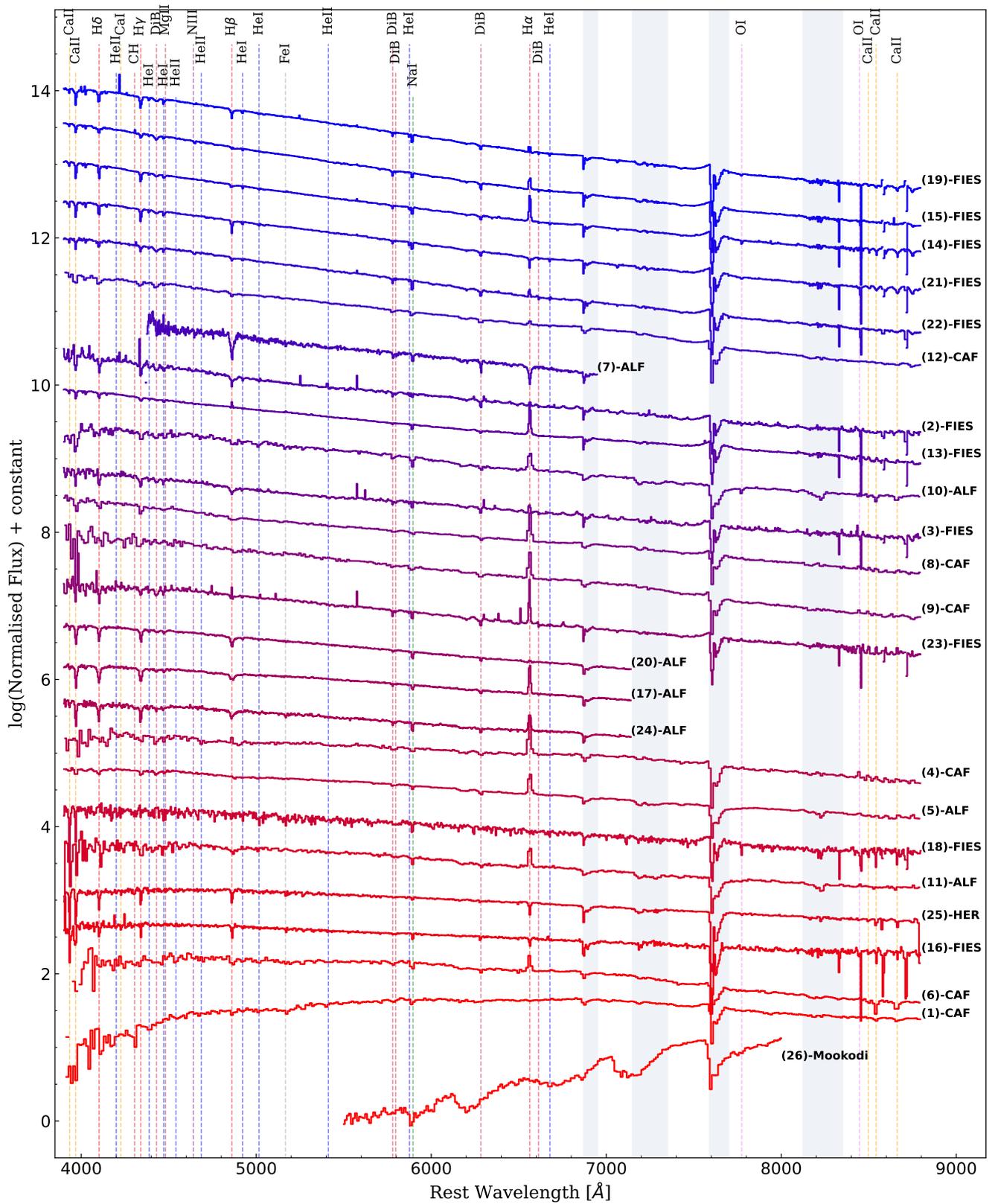
---

[2] `www.mercator.iac.es`

Fig. 4: Optical spectra for the subsample of spectroscopically characterised sources. Each spectrum, except for the Mookodi spectrum, is de-reddened for visualisation purposes using the best extinction values (see Sect. 6.1, Tab. A.1), with Fitzpatrick (1999) dust extinction function and $R_V = 3.1$. The main emission and absorption lines are indicated. The areas with strong telluric absorption are shown by the shaded rectangles. On the right, the number in parenthesis indicates the index associated with each source (see Table A.2), and this index is followed by the instrument used to take each spectrum: CAFOS (CAF) from the *Calar Alto Observatory* 2.2 m telescope, HERMES from the *Mercator* telescope, FIES and ALFOSC (ALF) from the *Nordic Optical Telescope*, and Mookodi from the *South African Astronomical Observatory* 1 m Lesedi telescope.
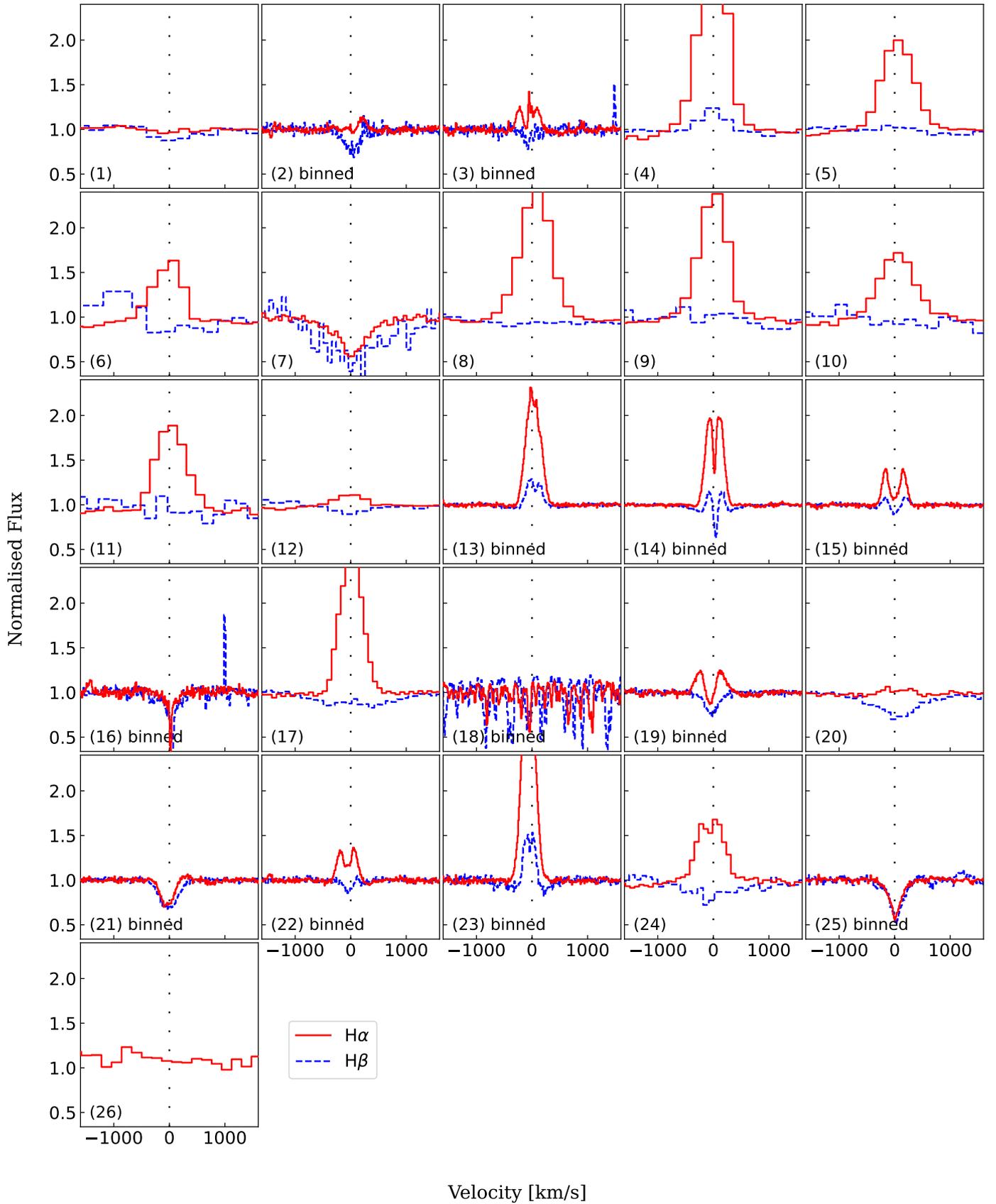
Fig. 5: Flux normalised Hα and Hβ velocity profiles for the subsample of observed sources. All sources are shown on the same scale. The index corresponding to each source (see Table A.2) is shown in the bottom left corner of each spectrum. Medium- and high-resolution spectra have been binned using $10\,km/s$ bins for improved visual representation.

4.1. Light curve analysis

4.1.1. Visual inspection

We visually inspected the retrieved light curves of every source in our sample. We labelled the observed features in the light curves using the following categories:

- Different kinds of variability (Diff): sources showing mixed, distinct variability patterns, including different oscillation modes or a combination of periodic and non-periodic variations, for example.
- Outbursts (Outb): sources showing sudden brightening once or more than once, which could be linked to outbursts.
- Slow-rising in brightness (SR): sources that are slowly increasing in brightness. This increase can be observed in the optical bands, in the IR bands, or both. We only give this label to sources that have been increasing in brightness during the observed time frame or are brightening at later times and do not show any signs of a previous brightening.

An example lightcurve for each of these three labels is shown in Fig. 6.

4.1.2. Periodicity analysis

We performed a periodicity search using `FINKER` (Stoppa et al. 2024), a nonparametric kernel regression code, applied to the ZTF, the ASAS-SN, and the ATLAS light curves. Frequencies from $0.001\,\mathrm{d}^{-1}$ (1 000 days) to $100\,\mathrm{d}^{-1}$ (14.4 minutes) with steps of $0.0004\,\mathrm{d}^{-1}$ were tested for all sources. Alternative frequency grids were also explored, but they did not yield improved results. We performed an additional search for periods as short as five minutes for some sources where no successful longer period was found. `FINKER` allows uncertainty estimation in frequency determination via bootstrapping, but in our case, the errors given by this method were insignificant, and we do not report them. Lomb-Scargle periodograms (Lomb 1976; Scargle 1982) were used for TESS light curves, which helped us find the shorter periods in our sample. Lomb-Scargle periodograms were also applied to the ZTF, the ASAS-SN, and the ATLAS light curves, but no significantly better results when compared with `FINKER` were obtained. In all attempts, we removed data points that were more than 5σ away from the median magnitude of all the observations. Outlier removal was performed carefully to avoid excluding short-duration eclipses. Finally, all periods were tested by visually inspecting the folded light curves.

We labelled the sources based on their periodic variability as sinusoidal (Sin), non-sinusoidal pulsations (Puls), or eclipsing (ECL). Sources were labelled as eclipsing only when the eclipses could be clearly identified. Consequently, contact binaries were classified as sinusoidal variables, since in most cases their eclipses are indistinguishable from the sinusoidal variability of pulsators.

4.2. Spectroscopic analysis

To verify the spectral type of our stars and their location in the HG, we carried out a spectroscopic follow-up campaign targeting the most promising sources with periodic or peculiar light curves. While higher-resolution spectra would allow a more detailed study of specific spectral features, the low resolution obtained for some systems was adequate for our primary objectives: assessing whether the stars are "yellow" and verifying the presence of Balmer emission.
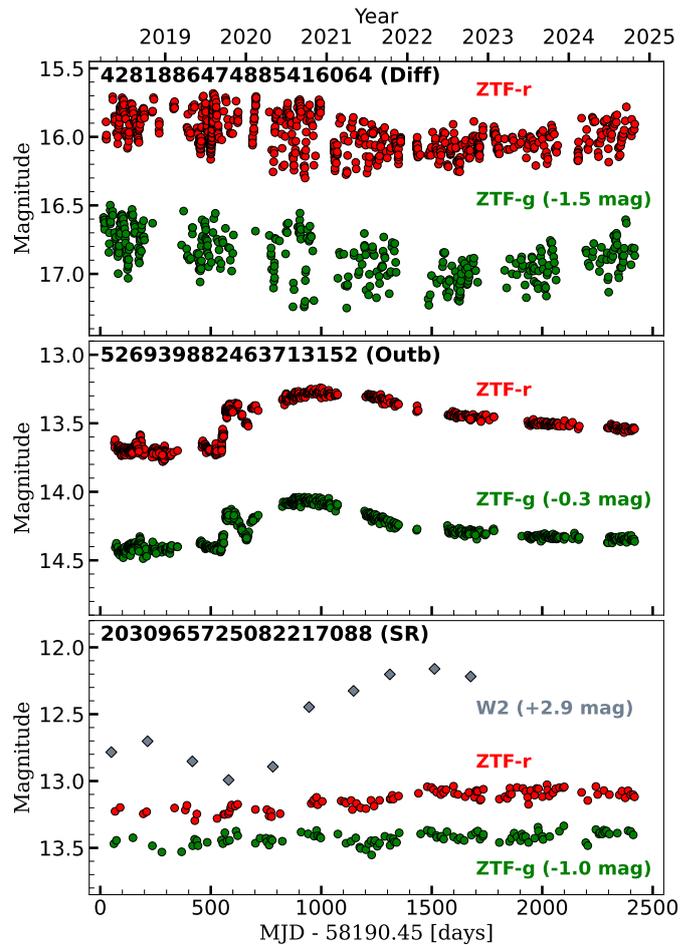

Fig. 6: Light curves of three sources in our sample illustrating the types of variability defined in this work (see Sect. 4.1.1). Circles represent the ZTF g band (green) and r band (red), while grey diamonds correspond to the NEOWISE W2 band. For clarity, the g-band and the W2-band light curves are vertically offset in all three panels, and the SR example is binned to better show its long-term evolution. The reference Modified Julian Date (MJD) corresponds to the start of the ZTF survey.

4.2.1. Spectroscopic classification

Our sources were given a spectral class by visually inspecting their spectra, and comparing them with theoretical spectra and spectra from standard stars, following Gray & Corbally (2009). We only report temperature types (O-type, B-type, etc.) in our classification, as our only aim is to determine if the spectral type is compatible with being a yellow star. Additionally, to obtain a complementary spectral classification, we used `NutMaat` (El-Kholy & Hayman 2024), a `Python` implementation of the `MKClass` code (Gray & Corbally 2014). These classifications were also compared with the spectral classification provided by *Gaia* DR3.

4.2.2. Hydrogen line profiles

We inspected the spectral region around Hα and Hβ and labelled each source, given the properties of their hydrogen profiles:

- P-Cygni profile (PCyg)
- Emission or absorption in Hα (Hα abs/emi): from the 26 sources with new spectra, this label tells which show emis-

sion or absorption in Hα, also indicating if it has a two-peak (2p) profile.

4.2.3. Extinction estimation

For the spectra with the high and medium resolution, we estimated the extinction from the absorption caused by the diffuse interstellar bands (DIBs) at 5 780 Å and 6 614 Å following Carvalho & Hillenbrand (2022), where they derive an empirical relation between the line-of-sight extinction and the EW of these two DIBs. To convert line-of-sight extinctions to visual extinctions, we used $R_V = A_V/E(B-V) = 3.1$. The resulting A_V values were then converted to extinctions in the *Gaia* passbands using the `gaia_edr3_photutils` package[3]. To use this method, we needed precise measurements of the EWs of the DIBs, which we could not get with our low-resolution spectra.

4.3. Characterisation with auxiliary catalogues

We cross-matched our sample with the UV catalogue *GALEX* (Martin et al. 2005) (with a search radius of 10 arcseconds), and X-ray catalogues from *Swift* (Evans et al. 2020) (with a search radius of 10 arcseconds), *Chandra* (Evans et al. 2010) (with a search radius of 50 arcseconds), *eROSITA* (Merloni et al. 2024) (with a search radius of 10 arcseconds), and *4XMM* (Webb et al. 2020) (with a search radius of 12 arcseconds). In all cases, the search radius was chosen to be enough to contain the positional error of the catalogues. We labelled the sources with X-ray (Hard X-ray) or UV (Far UV) detections.

To test for known variable stars, we also cross-matched our sample with a catalogue containing a cross-match of over 8 million *Gaia* sources with variable objects in 152 different catalogues from the literature (Gavras et al. 2023, or GVXM from now on). This catalogue provides over 100 variability (sub)types and periods, if available.

5. Results

5.1. Light curve analysis

We observed variability in the light curves of all 67 sources in our sample, but not all show signs of periodicity. Among all the sources, 19 show multiple types of variability, the majority of them likely associated with more than one pulsation mode (e.g., sources 2164630463117114496 or 4281886474885416064); ten show a fast decrease in magnitude (e.g., sources 526939882463713152 or 2027563492489195520), and 11 sources present a slow-increase in brightness over several hundreds or thousands of days (e.g., source 4263591911398361472). For some of the latter, the gradual brightening may be linked to long-term variability; future light curve data will help determine whether this is the case.

The folded light curves of the 36 candidates for which we determined a period are shown in Fig. B.2 and Fig. B.3. Periods of one day or less were found for 19 candidates, periods between one and eight days for ten candidates, and periods longer than ten days for the remaining seven, with the longest period being ≈106.76 days. Among these 36 sources with periodic variability, 26 show sinusoidal variability, nine show eclipses, and we identify pulsations for one (see Fig. 7).

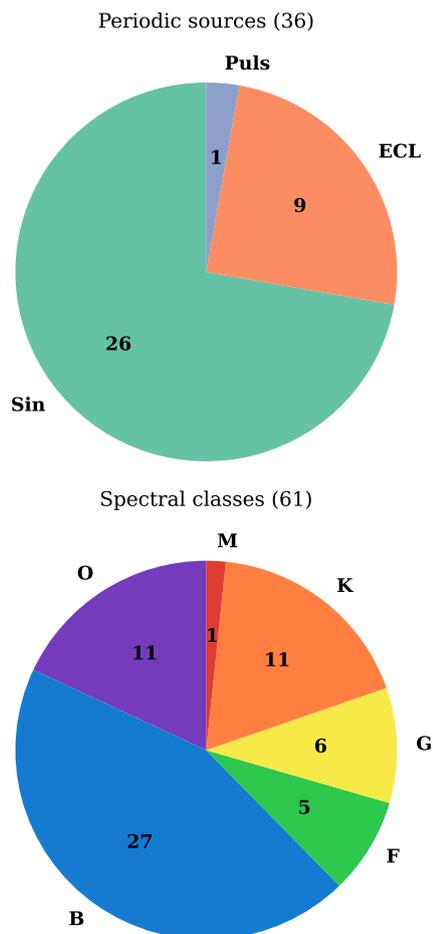
[3] `https://github.com/fjaellet/gaia_edr3_photutils/`

Fig. 7: Top: pie chart with the types of periodic variables identified in our sample. Bottom: pie chart with the spectral classes in our sample, combining the results of our classification and the classes reported in the *Gaia* DR3.

We note that periods detected exclusively in TESS light curves should be interpreted with caution. Given the large plate scale of TESS (21″ pixel^{-1}), light-curve contamination from nearby bright variables is possible when sources lie within the same pixel. This is a known issue discussed, for example, by Pedersen & Bell (2023) in the context of fast yellow pulsating supergiants. In our case, the short period of ≈ 0.171 days initially found for source *5962956195185292288* using TESS data was later traced to an eclipsing binary located ≈ 42″ away, identified as *Gaia* DR3 *5962956229545027712*.

5.2. Spectroscopic analysis

The estimated spectral types we report from our visual inspection can be grouped into three main groups: hotter stars (O- and B-type), yellow stars (F- and G-type), and colder stars (K- or M-type). Out of the 26 candidates with spectra, we identified 17 as hotter stars, most of them Be stars. On the other side, we identified one source as a K-type star, and one as an M-type star. Overall, only four sources were finally assigned a spectral type compatible with being a YG or YSG: three F-type stars and one G-type star. The spectra obtained for three of the sources showed no clear features (apart from the main hydrogen lines), and we

did not assign a spectral class to them. The `NutMaat` classifications rated as *good* or better were consistent with our spectroscopic results, yielding the same temperature types as those we derived.

A comparison of our spectral classification with the one available in the *Gaia* DR3 showed consistent results for 18 stars. For two, *Gaia* identified them as G-type stars, whereas we classified them as B-type based on the presence of He ɪ absorption lines. Another source was classified as an O-type star by *Gaia*, but our analysis indicates that its spectrum matches a B-type better. Additionally, *Gaia* had spectral types for 38 sources in our sample for which we lacked spectra. These 38 spectral types are distributed as follows: 8 O-type, 13 B-type, 2 F-type, 5 G-type, and 10 K-type stars. All the spectral types identified are summarised in Fig. 7.

We note that our spectral classification was based on single stellar spectra. In the case of binaries, this approach is reasonable when one component dominates the system's light, which we consider to be the case for our sources. Indeed, most of our spectra showed no evidence of composite features, suggesting that the contribution of a secondary component is negligible. However, this effect could explain the peculiar spectra observed in the three candidates for which no spectral type was assigned (sources *2061252975440642816*, *2164630463117114496*, and *2175699216614191360*).

From the Hα and Hβ profiles shown in Fig. 5, we identified Hα emission in 19 sources (ten of which had a two-peak profile), Hα absorption in five sources, and no discernible features in two of them. Five sources had Hβ emission in addition to Hα emission. Additionally, in Fig. B.4 we present a comparison between the Hα profiles obtained from the *Gaia* XP low-resolution spectra and our follow-up spectra, degraded to match the XP resolution. In most cases, the profiles show good agreement, while in others they differ significantly, suggesting that the Hα emission may be transitory or variable over time.

Following the method described in Sect. 4.2.3, we estimated the extinction for each source, resulting in A_V values reported in Table 2. Updated extinction values have enabled us to revisit the locations of nine sources in the *Gaia* colour-magnitude diagram, as shown in Fig. 8. The discussion of these results and a comparison of our extinction estimates with those from two external catalogues are discussed in Sections 6.1 and 6.2.

5.3. Characterisation with auxiliary catalogues

Our cross-match with UV and X-ray catalogues allowed the identification of three sources with X-ray emission, and one source with UV excess. The source with the UV detection in *GALEX* was *6123873398383875456*, which presented UV excess in the far UV channel (1350–1750 Å). This detection matched the optical counterpart at 0.44 arcseconds (within positional errors and with no other optical counterpart nearby).

The X-ray detections came from *4XMM* and *Swift*: one detection in *Swift*'s *2SXPS* catalogue matching source *4054010697162430592* at 0.93 arcseconds (within positional errors and with no other optical counterpart nearby), and two detections in *4XMM*-DR14, matching source *5866345647515558400* at 1.1 arcseconds and source *2083649030845658624* at 1.9 arcseconds, both within positional errors. However, *2083649030845658624* had other optical counterparts within a few arcseconds (also within positional errors), so it is not clear if the X-ray emission is from our source.

Table 2: Results of our extinction estimates using diffuse interstellar bands (DIBs).

Index	Gaia DR3 ID	A_V (mag)	$A_{V,err}$ (mag)
2	187219239343050880	4.06	1.13
3	2006088484204609408	3.72	1.02
13	3355776901779440384	1.18	0.30
14	3369399099232812160	2.47	0.65
15	3444168325163139840	4.00	1.10
19	4611936956247755424	3.56	1.00
21	508419369310190976	3.21	0.87
22	512721444765993472	4.82	1.37
23	526939882463713152	4.80	1.33

Notes. The reported A_V values represent the average of the estimates derived from the 5 780 Å and 6 614 Å DIBs. Indices correspond to those reported in Table A.2. For details on the derivation of the A_V values, see Sect 4.2.3
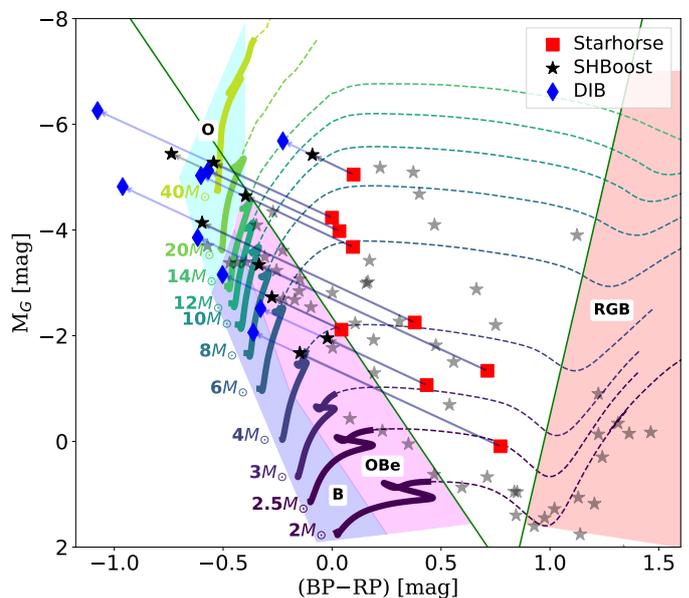

Fig. 8: Positions of the nine sources for which we re-estimated extinction (see Sect. 4.2.3) in the colour-magnitude diagram. Red squares indicate the positions derived using `StarHorse`, while blue diamonds show the positions computed using our extinction estimates based on diffuse interstellar bands, averaging the results from the 5 780 Å and 6 614 Å lines. For comparison, we also show the positions of all sources in our sample using the `SHBoost` catalogue (black stars; see Sect. 6.1). The schematic boxes denote the typical locations of different stellar types.

Our search for periods reported in the literature provided 18 catalogued periods. Comparing these with the periods derived in this work, we found that eight sources had consistent values. For one eclipsing binary, our analysis yielded half the reported period. We were able to reproduce the reported period for four sources, while for the remaining five, the periods published in GVXM could not be confirmed with our data. The unconfirmed periods correspond to relatively long timescales, which may not be reproduced with our data either due to the limited observational baseline or because the literature periods themselves are uncertain, as the cross-matched values in GVXM are not independently verified.

The main results of our photometric and spectroscopic analysis are summarized in Fig. 7, and all the results of our analysis are provided in Table A.3, where we present the spectral types identified in this work together with the types listed by *Gaia*, the main properties in the light curves and the spectra, the classes from the literature (from the GVXM and SIMBAD), and the periods found in this work and the literature. Literature periods that could not be confirmed with our data are marked with an asterisk in the last column of these tables.

6. Discussion

The goal of our selection was to identify mass-transferring binaries in the HG. However, our sample includes a variety of spectral types, most of which are contaminants. In this section, we first discuss how line-of-sight extinction affected our selection process, and then examine the nature of our candidates.

6.1. The effect of extinction

The most significant limitation of our method for selecting our sample of HG stars was the lack of reliable extinction estimates in the literature. As a result, for more than half of the sources in our sample, the extinction was underestimated, implying that the MS O- and B-type sources appeared fainter and redder. Consequently, many of our candidates were stars such as Oe or Be stars, which share some characteristics with our objects of interest but are generally not expected to lie in the HG Cochetti et al. (2020). Therefore, obtaining precise extinction estimates is essential for reliably selecting sources within the narrow region of the colour-magnitude diagram occupied by the HG.

Luckily, during the preparation of this work, several new astrophysical parameter catalogues derived from *Gaia* XP spectra became available (e.g., Zhang et al. 2023; Fallows & Sanders 2024; Huson et al. 2025). Of particular interest for us was the new SHBoost catalogue (Khalatyan et al. 2024), which contains stellar parameters (effective temperature, surface gravity, mass, metallicity) and line-of-sight extinctions computed using *Gaia* DR3 XP spectra, derived through supervised machine learning trained on a large high-quality dataset comprising the StarHorse results for spectroscopic stellar surveys (Queiroz et al. 2023) as well as complementary datasets of hot stars, white dwarfs, hot subdwarfs, among others. Particularly relevant for this work is that SHBoost contains more reliable characterisation (including extinction) for high-mass stars than the StarHorse catalogue, based on *Gaia* EDR3.

6.2. Revisiting our first selection with SHBoost

To assess the position of our sources in the colour-magnitude diagram with better extinction values, we cross-matched our sample with the SHBoost catalogue to obtain more accurate absolute magnitudes and colours. A comparison with the extinction corrected magnitudes from DIB spectroscopy is shown in Fig. 8, where we highlight two key observations: first, many of the Oe and Be stars now lie on the MS or just beyond the terminal-age MS, while the cooler stars (K- and M-types) are located on the RGB; second, all but one of the extinction values derived from our spectroscopic analysis are higher than those reported in SHBoost, with A_V values larger by approximately 0.7 mag on average. However, the values remain comparable within uncertainties, and we therefore consider the SHBoost estimates to be reliable. Figure 8 also shows that from the initial 67 candidates

selected based on the EDR3 StarHorse catalogue, only 25 remain in our HG parameter space. Given these results, we can now discuss the true nature of the sources in our sample.

6.3. Nature of our candidates

6.3.1. Oe and Be stars

Oe and Be stars are rapidly rotating, moderately massive stars (M ∼ 3.6–20 $M_\odot$) with Balmer emission associated with a gaseous envelope, and with an equatorial disk likely formed through pulsations or by radiatively-driven winds (Porter & Rivinius 2003; Rivinius et al. 2013). Moreover, these stars show IR excess caused by free–free and free–bound emission from the gas in the disk (Gehrz et al. 1974). Oe and Be stars are also known to exhibit different kinds of photometric variability (Diago et al. 2009; Labadie-Bartz et al. 2017).

Among the O- and B-type stars identified in our sample, 15 are strong candidates for Oe or Be stars based on the presence of Balmer emission. In some cases, Oe and Be stars also display Ca ɪɪ triplet and Paschen emission (e.g., Polidan 1976; Banerjee et al. 2021), an emission we identify in, for example, sources *2013187240507011456* and *2074693061975359232*. Only two B-type stars in our sample do not show Balmer emission in our follow-up spectra. As noted by Negueruela et al. (2004), the emission lines in Oe and Be stars are variable and can transition from emission to absorption as the star temporarily exits the Be phase. This may explain the case of source *508419369310190976*, for which Hα emission is visible in the *Gaia* XP spectrum (see Fig. B.4), but it is absent in our follow-up spectrum.

Additionally, for the 44 sources for which we could not assign a spectral classification, either because we did not have follow-up spectra or because the spectra did not show enough features, 21 were classified as O- or B-type stars by *Gaia*. Based on the selection criteria of our search and the results obtained, there is a strong reason to believe that most, if not all, of these sources are likely Oe or Be stars.

6.3.2. Regular variables

Regular variables cannot be discarded as possible contaminants in our search. We discuss source *6123873398383875456* in Sect. 6.3.4, which we think is most likely a close binary misclassified as an RRc, but we found two other sources classified as regular pulsator stars in the literature: *4606866489658625228*, classified as a δ Scuti by Chen et al. (2020), and *4076568861833452160*, classified as a Cepheid by Jayasinghe et al. (2019). These kinds of variables reside in the instability strip, which crosses the HG, so it is expected to find them in our parameter space. Examples of different kinds of variables placed in the *Gaia* HR diagram can be found in Gavras et al. (2023).

Mid-IR excess is also expected in Cepheid variables (Kervella et al. 2006; Hocdé et al. 2020), and Balmer emission has been observed in the type II Cepheid W Virginis (Kovtyukh et al. 2011). Pre-MS δ Scuti variable stars have been observed before (Díaz-Fraile et al. 2014), and strong IR-excess and emission are expected in this type of star, linked to active accretion processes. Therefore, the selection of these variables with our search criteria is possible, and a new identification method will be needed to trace and remove them from our sample, especially if we want to distinguish them from close binaries, which can have similar light curves.
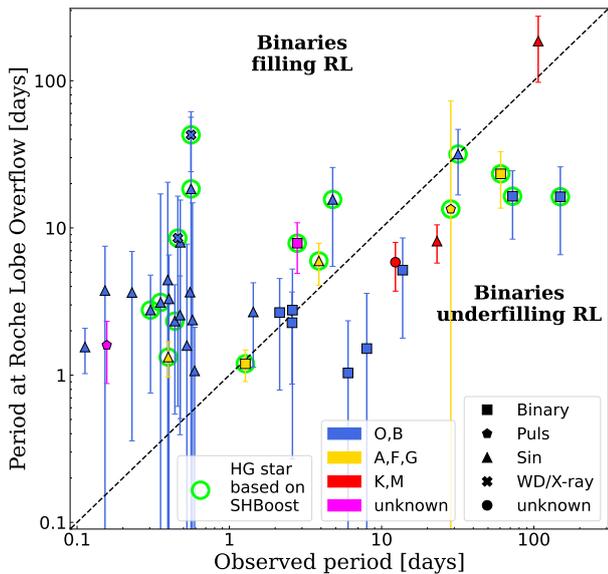
Fig. 9: Comparison of the observed period and the theoretical period of a binary in which one star fills its Roche lobe (see Sect. 6.3.4). Different marker shapes correspond to different types of stars, and colours represent spectral classes (using our classification when available, and *Gaia*'s otherwise). Sources that lie, based on SHBoost values, within our defined Hertzsprung gap parameter space are highlighted with circles.

6.3.3. Binaries with compact companions

As described in Sect. 5.3, we identify one source with far UV excess and three sources with hard X-ray emission. From the three sources with X-ray detections, it is not clear if the X-ray detection of source *2083649030845658624* corresponds to its optical counterpart, whereas sources *4054010697162430592* (reported as a single-lined spectroscopic binary in Makarov & Unwin 2015) and *5866345647515558400* are good candidates for X-ray binaries.

Interestingly, the source with far UV excess—source *473575777103322496*—is identified as a Be star, has sinusoidal variability with a period of 0.45878 days (Fig. B.3), and shows two peaks in the IR light curve, one between 2014 and 2016 and a second one between 2018 and 2020 (Fig. B.1). We suggest this source is a Be star with a hot companion (a subdwarf OB star or a white dwarf, for example), which is potentially accreting mass during a second mass transfer stage (as discussed in Gies et al. (2023) for γ Cas stars), based on the UV emission and the observed variability.

6.3.4. Mass transferring systems

One possible method for assessing whether our candidates were undergoing mass transfer in a binary system was to compare their observed orbital period with the theoretical period at which Roche lobe overflow would occur (P_{RLO}). Therefore, we take advantage of the improved parameters given by SHBoost to calculate this period,

$$P_{RLO} \approx 0.35 \left(\frac{R^3}{M}\right)^{1/2} \left(\frac{2}{1+q}\right)^{0.2} \text{ days.} \quad (1)$$

The masses are provided directly by SHBoost, while the radii are computed using the log g values also given in SHBoost.

We adopted a fixed mass ratio of $q = 0.5$ but accounting for the dependence on q in the error bars by also computing the period at $q = 0.01$ (indicating that 99% of the system's mass resides in the primary star) and at $q = 1$ (representing an equal-mass binary). The dependence of P_{RLO} on q is weak, and adopting different q values does not significantly affect our results. It is to be noted that the masses in SHBoost are computed assuming single stars. In case the second component contributes to the total light, this may affect the final P_{RLO}, adding an additional uncertainty that we can not verify straightforwardly.

We compared the P_{RLO} to the periods found in this work, and we show this in Fig. 9, where we also mark different kinds of variables found among our candidates, their spectral class, and whether they remain in the HG using the extinctions from the SHBoost catalogue. Points above the diagonal correspond to sources with an observed period shorter than the period they would have if they were a binary and one of the components had filled its Roche lobe, whereas the opposite is true for points below the diagonal. Binaries in this region, with observed periods longer than P_{RLO}, are most likely detached binaries. Nonetheless, an interesting example of a source in this area—the G-type star *2060841448854265216*—shows, in its folded light curve, eclipses akin to an Algol star with an accretion disk around the gainer (see B.2). Algol binaries are systems where the mass gainer is the more massive star that is still in the MS, and the donor is the cooler, fainter and larger star (Peters 2001). In this case, we could be seeing the less evolved star (or a combination of the two), which would mean P_{RLO} is underestimated. Consequently, we would not rule out *2060841448854265216* as a potential mass-transferring binary.

The most relevant sources for our study are those located near the line where the observed period equals P_{RLO}, as these represent the strongest candidates for ongoing mass transfer in binary systems. Among them, perhaps the most intriguing is source *6228685649971375616*, a yellow HG star with a folded light curve characteristic of an interacting binary and an orbital period of 1.274 days. The light curves in different bands are shown in Fig. 10, together with the folded ATLAS o-band and ZTF g-band light curves. We also present the *Gaia* XP spectrum and compare it with a model spectrum of a late F-type star ($T_{eff} = 6100$ K). The model was generated using pysynphot (STScI Development Team 2013) with the Castelli–Kurucz Atlas (Castelli & Kurucz 2003), adopting log $g = 3.75$, as reported by SHBoost (see Table A.1). The comparison shows a good agreement between the observed and model spectra, confirming that this is indeed a yellow source. Although no clear Hα emission is detected in the *Gaia* XP spectrum, the absence of a well-defined absorption feature suggests that the expected absorption is partially filled.

If this source is indeed evolving toward a merger (and a subsequent LRN), and assuming its progenitor behaves similarly to that of V1309 Sco (Tylenda et al. 2011), we would expect to observe three key signatures in its light-curve evolution: (1) a steady, monotonic brightening over several years (of order ∼ 1 mag in the I band for V1309 Sco) prior to outburst; (2) a progressive decrease in the orbital period as the components spiral in; and (3) a transition in the folded light-curve morphology from an EW-type with two maxima per cycle to an asymmetric single maxima profile (see figures 1–3 in Tylenda et al. 2011). Such behaviour is not observed in *6228685649971375616*. A slight increase in brightness is seen in the ATLAS bands (of about ∼ 0.1 mag over approximately 10 years), but this trend is not evident in other bands. Moreover, we detect no significant evolution in the folded light curve, which consistently shows two maxima
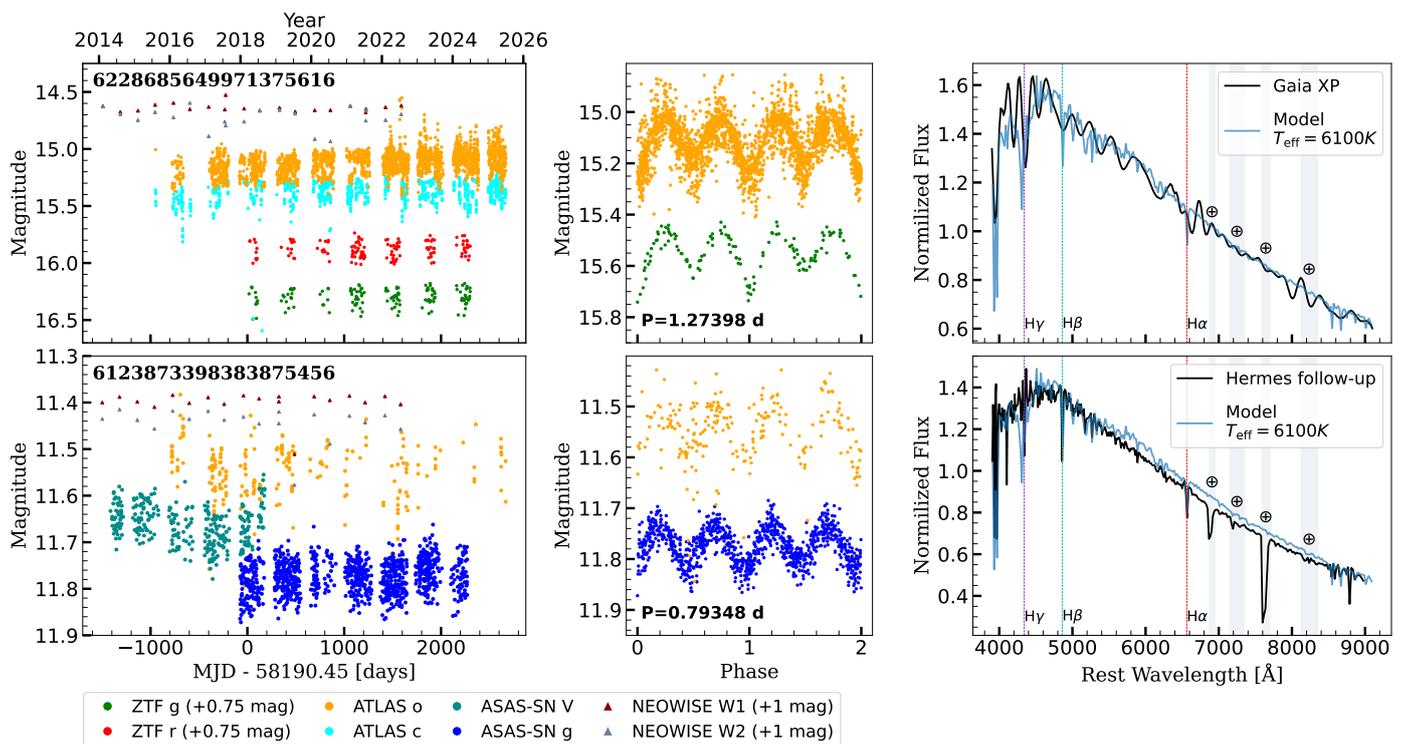
Fig. 10: Light curve (left column), phase-folded light curve (middle column), and spectrum (right column) of our two best candidates for mass-transferring merger progenitors: *6228685649971375616* (top row) and *6123873398383875456* (bottom row). For both sources, a model spectrum from the Castelli–Kurucz Atlas (Castelli & Kurucz 2003), with $T_{\rm eff}$ = 6100 K and $\log g$ = 3.75, is shown for comparison. The ZTF and NEOWISE data in the light curves (left column) are offset for visual clarity, and the reference Modified Julian Date (MJD) in the light curve corresponds to the start of observations by ZTF. Although we report a sinusoidal period of 0.39674 d for *6123873398383875456*, the phase-folded light curve is shown using twice this period, consistent with our interpretation of the source as a contact eclipsing binary. The $\oplus$ symbols and shaded areas in the spectra (right column) represent telluric absorption (absorption by the Earth's atmosphere).

and two minima per cycle across all epochs. We measured the period across different epochs, but any variations detected were smaller than the associated uncertainties. Investigating potential period evolution would require high-cadence observations over an extended timespan, which are not available for this source. If this object is indeed similar to the progenitor of V1309 Sco, we do not expect its outburst to occur in the near future.

While for the source discussed above we could clearly distinguish the two minima corresponding to the eclipses, this is not the case for some contact binaries. Contact binaries and regular pulsators can exhibit similarly sinusoidal light curves with short periods. Therefore, some of our candidates showing sinusoidal variability may in fact be close binary systems. One such case is *6123873398383875456*, an F-type star identified in the ASAS-SN catalogue of variable stars (Jayasinghe et al. 2019) as an RRc-type pulsator, although it is not included in the RR Lyrae table of *Gaia* DR3. Notably, this source has a re-normalised unit weight error (RUWE) of nearly 6.5 in the *Gaia* DR3 catalogue, a value well above the nominal threshold and often indicative of unresolved binarity (Castro-Ginard et al. 2024). The folded light curve displays two symmetric minima with slightly different depths but the same periodicity. Although this behaviour could be attributed to period doubling in an RR Lyrae star, it could also correspond to the two eclipses of a contact binary. If the latter interpretation is correct, the star would lie just above the 1:1 line in Fig. 9, further supporting the hypothesis that it is a mass-transferring binary.

Similarly to *6228685649971375616*, source *6123873398383875456* does not show the behaviour expected for an LRN progenitor approaching outburst. In Fig. 10, we show its light curves across different bands, along with the folded ATLAS o-band and ASAS-SN g-band light curves. We also present the follow-up spectrum, which we compare to a model of a late F-type star (the same model described above).

The case of source *6228685649971375616*, however, appears to be an exception rather than the rule, as we do not find any evidence of binarity in other sinusoidal sources. As discussed in the subsections above, regular variables and OBe stars can also be selected by our criteria. These likely account for the majority of stars with sinusoidal light curves located in the leftmost region of the P_{RLO}-observed period diagram.

Interestingly, among all the O- and B-type stars in our sample, we identify at least seven in binary systems, including three Be stars: *527155253604491392*, *2006088484204609408*, and *2166378312964576256*. The formation channel of Oe and Be stars remains a topic of debate, although several studies suggest that close binary interactions are an important channel for their formation (e.g., Pols et al. 1991; Shao & Li 2014; Dodd et al. 2024). These three Be stars in binary systems are therefore particularly compelling targets for investigating this formation channel. Two of them—*527155253604491392* and *2166378312964576256*—appear to be in wide binaries, possibly having acquired their Be characteristics through past binary interaction, as suggested by their location in Fig. 9. In contrast,

2006088484204609408 may represent an actively interacting binary system, where the accreting star is currently evolving into a Be star.

7. New sample of HG objects

The analysis of our initial sample showed that uncertainties in the extinction were the main cause of including contaminants in our selection process. Given the appearance of new catalogues, we redid the same selection steps described in Sect. 2, but this time on the `SHBoost` input catalogue instead of `StarHorse`. In this case, we only looked at Hα EW values for the selection of hydrogen emitters, and we removed false positives by visual inspection of the calibrated XP spectra. We did not remove any further sources that might have a classification in `SIMBAD` or other catalogues. The other selection steps were kept the same.

The filtering process provided a new preliminary sample composed of 444 sources. After visually inspecting the XP spectra of the preliminary sample, we finished with a new sample composed of 308 HG stars that fit our criteria. We will explore this new sample in future research. We present this new sample in a table made available in its entirety via CDS/VizieR.

8. Summary and conclusions

In this work, we aimed to identify Galactic mass-transferring binaries with quickly evolving donors in the HG with yellow spectral types that showed Balmer emission, mid-IR excess, and variability. After our initial selection, we obtained a sample of 67 candidates that we analysed using photometric data and spectroscopic data, along with multi-wavelength photometric archival data. We present a summary of the nature of our candidates (identified by their *Gaia* DR3 IDs):

- We computed periods for 36 sources, among which 26 showed sinusoidal variability, nine were eclipsing binaries, and one showed pulsations.
- The spectral types we identified from our follow-up spectra and the spectral types given by *Gaia* were in agreement in 18 out of 21 cases. Combining our spectral types and the ones reported by *Gaia*, the distribution of spectral classes in our sample was: 11 O-type, 27 B-type, 5 F-type, 6 G-type, 11 K-type, and 1 M-type star.
- Among the O- and B-type stars in our sample, 15 are strong candidates for Oe or Be stars based on the presence of Balmer emission. These reddened Oe and Be stars contributed the most to the contaminants in our sample.
- Regular variables cannot be discarded as possible contaminants, and we identified two in our sample: a δ Scuti (*460686648965862528*), and a Cepheid (*4076568861833452160*).
- Three sources in our sample showed associated hard X-ray emission. Two of them had good matches with their optical counterpart (sources *4054010697162430592* and *5866345647515558400*), and are strong candidates for X-ray binaries. Source *2083649030845658624* had other optical counterparts within the positional errors of the X-ray detection.
- Source *473575777103322496* shows far UV excess. This source was identified as a Be star, and we suggest it is in a binary system with a hot companion (subdwarf OB star or a white dwarf, for example), and it is potentially mass-transferring.

- The strongest candidates for mass-transferring binaries with a bright yellow primary are *2060841448854265216*, a G-type star whose light curve resembles that of an Algol-type eclipsing binary with an accretion disk; *6228685649971375616*, an F-type star in a close binary, and *6123873398383875456*, an F-type contact binary that was previously misclassified as an RRc Lyrae star.
- Among all the O- and B-type stars in our sample, we identify seven in binary systems, including three Be stars: *527155253604491392*, *2006088484204609408*, and *2166378312964576256*. These three Be stars in binary systems are particularly compelling targets for investigating the formation channel of Oe and Be stars through close binary interactions.

Most of the sources in our analysed sample are not yellow stars. Of the initial 67 candidates selected using the EDR3 `StarHorse` catalogue, only 25 remain within our HG parameter space when using the more reliable `SHBoost` extinction values. This underscores a key challenge in our original goal of identifying yellow HG stars as potential LRN progenitors: the large uncertainties in extinction estimates from *Gaia* and `StarHorse`, which were central to defining our selection space.

Comparing the observations of the two best candidates for stellar-merger progenitors in our sample— *6228685649971375616* and *6123873398383875456*—with those of the progenitor of V1309 Sco, the best-studied case of a stellar merger to date, suggests that if these systems are indeed potential merger progenitors, they are still far from the merger phase. Nevertheless, both remain interesting targets that deserve further investigation.

As an additional outcome of this work, we present a refined selection of 308 HG candidate stars using the improved extinction correction from the `SHBoost` catalogue. This new sample was derived using the same overall methodology but incorporates improved extinction estimates and visual inspection of the *Gaia* XP spectra, ensuring that each selected object displays, at a minimum, clear Hα emission. From this revised sample, we expect to identify a significantly larger fraction of scientifically valuable mass-transferring binary systems.

This new sample gives new opportunities to explore HG stars with interesting features. We encourage other teams to pursue follow-up observations of promising sources in this newly identified sample, and also of interesting sources in our initial sample. Additionally, forthcoming data releases, particularly *Gaia* DR4, expected by the end of 2026, will substantially enhance the search for HG mass-transferring binaries by providing improved spectrophotometric and astrometric data, including BP/RP spectra for all sources down to $G \approx 19$ mag, RVS spectra for sources brighter than $G \approx 14.5$ mag, and epoch photometry. These additions will offer a richer dataset for identifying and characterising evolved binaries and their emission features with greater precision.

References

Addison, H., Blagorodnova, N., Groot, P. J., et al. 2022, MNRAS, 517, 1884
Anders, F., Khalatyan, A., Queiroz, A. B. A., et al. 2022, A&A, 658, A91
Andrew, S., Swihart, S. J., & Strader, J. 2021, ApJ, 908, 180
Armeni, A. & Shore, S. N. 2022, A&A, 664, A103
Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, ApJ, 935, 167
Banerjee, G., Mathew, B., Paul, K. T., et al. 2021, MNRAS, 500, 3926
Blagorodnova, N., Klencki, J., Pejcha, O., et al. 2021, A&A, 653, A134
Blagorodnova, N., Kotak, R., Polshaw, J., et al. 2017, ApJ, 834, 107
Bond, H. E., Henden, A., Levay, Z. G., et al. 2003, Nature, 422, 405
Bouma, L. G., Hartman, J. D., Bhatti, W., Winn, J. N., & Bakos, G. Á. 2019, ApJS, 245, 13
Brož, M., Mourard, D., Budaj, J., et al. 2021, A&A, 645, A51
Cai, Y. Z., Pastorello, A., Fraser, M., et al. 2022, A&A, 667, A4

Caldwell, D. A., Tenenbaum, P., Twicken, J. D., et al. 2020, Research Notes of the American Astronomical Society, 4, 201
Carrasco, J. M., Weiler, M., Jordi, C., et al. 2021, A&A, 652, A86
Carvalho, A. S. & Hillenbrand, L. A. 2022, ApJ, 940, 156
Castelli, F. & Kurucz, R. L. 2003, in IAU Symposium, Vol. 210, Modelling of Stellar Atmospheres, ed. N. Piskunov, W. W. Weiss, & D. F. Gray, A20
Castro-Ginard, A., Penoyre, Z., Casey, A. R., et al. 2024, A&A, 688, A1
Cehula, J. & Pejcha, O. 2023, MNRAS, 524, 471
Chen, X., Wang, S., Deng, L., et al. 2020, ApJS, 249, 18
Cochetti, Y. R., Zorec, J., Cidale, L. S., et al. 2020, A&A, 634, A18
Crawford, J. A. 1955, ApJ, 121, 71
Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, Research in Astronomy and Astrophysics, 12, 1197
Cutri, R. M., Wright, E. L., Conrow, T., et al. 2021, VizieR Online Data Catalog, II/328
De Angeli, F., Weiler, M., Montegriffo, P., et al. 2023, A&A, 674, A2
de Mink, S. E., Sana, H., Langer, N., Izzard, R. G., & Schneider, F. R. N. 2014, ApJ, 782, 7
Diago, P. D., Gutiérrez-Soto, J., Auvergne, M., et al. 2009, A&A, 506, 125
Díaz-Fraile, D., Rodríguez, E., & Amado, P. J. 2014, A&A, 568, A32
Djupvik, A. A. & Andersen, J. 2010, in Astrophysics and Space Science Proceedings, Vol. 14, Highlights of Spanish Astrophysics V, 211
Dodd, J. M., Oudmaijer, R. D., Radley, I. C., Vioque, M., & Frost, A. J. 2024, MNRAS, 527, 3076
Dotter, A. 2016, ApJS, 222, 8
El-Kholy, R. I. & Hayman, Z. M. 2024, NutMaat: A Python library for classifying stellar spectra based on the MKCLASS package
Erasmus, N., Steele, I. A., Piascik, A. S., et al. 2024, Journal of Astronomical Telescopes, Instruments, and Systems, 10, 025005
Evans, I. N., Primini, F. A., Glotfelty, K. J., et al. 2010, ApJS, 189, 37
Evans, P. A., Page, K. L., Osborne, J. P., et al. 2020, ApJS, 247, 54
Fallows, C. P. & Sanders, J. L. 2024, MNRAS, 531, 2126
Fitzpatrick, E. L. 1999, PASP, 111, 63
Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1
Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023, A&A, 674, A1
Gavras, P., Rimoldini, L., Nienartowicz, K., et al. 2023, A&A, 674, A22
Gehrz, R. D., Hackwell, J. A., & Jones, T. W. 1974, ApJ, 191, 675
Gies, D. R., Wang, L., & Klement, R. 2023, ApJ, 942, L6
Ginsburg, A., Sipőcz, B. M., Brasseur, C. E., et al. 2019, AJ, 157, 98
Gray, R. O. & Corbally, C. J. 2014, AJ, 147, 80
Gray, R. O. & Corbally, Christopher, J. 2009, Stellar Spectral Classification
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357
Hart, K., Shappee, B. J., Hey, D., et al. 2023, arXiv e-prints, arXiv:2304.03791
Heinze, A. N., Tonry, J. L., Denneau, L., et al. 2018, AJ, 156, 241
Hocdé, V., Nardetto, N., Lagadec, E., et al. 2020, A&A, 633, A47
Huang, C. X., Vanderburg, A., Pál, A., et al. 2020, Research Notes of the American Astronomical Society, 4, 204
Hunter, J. D. 2007, Computing in Science & Engineering, 9, 90
Huson, D., Cowan, I., Sizemore, L., Kounkel, M., & Hutchinson, B. 2025, ApJ, 984, 58
Ivanova, N., Podsiadlowski, P., & Spruit, H. 2001, in Astronomical Society of the Pacific Conference Series, Vol. 229, Evolution of Binary and Multiple Star Systems, ed. P. Podsiadlowski, S. Rappaport, A. R. King, F. D'Antona, & L. Burderi, 261
Ivanova, N. & Taam, R. E. 2004, ApJ, 601, 1058
Jayasinghe, T., Stanek, K. Z., Kochanek, C. S., et al. 2019, MNRAS, 485, 961
Kahraman Aliçavuş, F., Handler, G., Aliçavuş, F., et al. 2022, MNRAS, 510, 1413
Kato, T. 2003, A&A, 399, 695
Kervella, P., Mérand, A., Perrin, G., & Coudé du Foresto, V. 2006, A&A, 448, 623
Khalatyan, A., Anders, F., Chiappini, C., et al. 2024, A&A, 691, A98
Kovtyukh, V. V., Wallerstein, G., Andrievsky, S. M., et al. 2011, A&A, 526, A116
Kraft, R. P. 1964, ApJ, 139, 457
Kuiper, G. P. 1941, ApJ, 93, 133
Kunimoto, M., Huang, C., Tey, E., et al. 2021, Research Notes of the American Astronomical Society, 5, 234
Labadie-Bartz, J., Pepper, J., McSwain, M. V., et al. 2017, AJ, 153, 252
Lam, M. C., Smith, R. J., Arcavi, I., et al. 2023, AJ, 166, 13
Lomb, N. R. 1976, Ap&SS, 39, 447
Lund, M. N., Handberg, R., Buzasi, D. L., et al. 2021, ApJS, 257, 53
MacLeod, M., Macias, P., Ramirez-Ruiz, E., et al. 2017, The Astrophysical Journal, 835, 282
Mainzer, A., Bauer, J., Grav, T., et al. 2011, ApJ, 731, 53
Maíz Apellániz, J., Holgado, G., Pantaleoni González, M., & Caballero, J. A. 2023, A&A, 677, A137
Makarov, V. V. & Unwin, S. C. 2015, MNRAS, 446, 2055
Marchant, P., Pappas, K. M. W., Gallegos-Garcia, M., et al. 2021, A&A, 650, A107

Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, ApJ, 619, L1
Martini, P., Wagner, R. M., Tomaney, A., et al. 1999, AJ, 118, 1034
Masci, F. J., Laher, R. R., Rusholme, B., et al. 2019, PASP, 131, 018003
Mason, E., Diaz, M., Williams, R. E., Preston, G., & Bensby, T. 2010, A&A, 516, A108
McKinney, W. et al. 2010, in Proceedings of the 9th Python in Science Conference, Vol. 445, Austin, TX, 51–56
Meisenheimer, K. 1994, Sterne und Weltraum, 33, 516
Merloni, A., Lamer, G., Liu, T., et al. 2024, A&A, 682, A34
Moe, M. & Di Stefano, R. 2017, ApJS, 230, 15
Mourard, D., Brož, M., Nemravová, J. A., et al. 2018, A&A, 618, A112
Munari, U., Henden, A., Kiyota, S., et al. 2002, A&A, 389, L51
Negueruela, I., Steele, I. A., & Bernabeu, G. 2004, Astronomische Nachrichten, 325, 749
Paczynski, B. 1976, in Structure and Evolution of Close Binary Systems, ed. P. Eggleton, S. Mitton, & J. Whelan, Vol. 73, 75
Pastorello, A. & Fraser, M. 2019, Nature Astronomy, 3, 676
Pastorello, A., Valerin, G., Fraser, M., et al. 2023, A&A, 671, A158
Paxton, B., Bildsten, L., Dotter, A., et al. 2011, ApJS, 192, 3
Pedersen, M. G. & Bell, K. J. 2023, AJ, 165, 239
Pejcha, O. 2014, ApJ, 788, 22
Pejcha, O., Metzger, B. D., & Tomida, K. 2016, MNRAS, 455, 4351
Peters, G. J. 2001, in Astrophysics and Space Science Library, Vol. 264, The Influence of Binaries on Stellar Population Studies, ed. D. Vanbeveren, 79
Podsiadlowski, P. 2001, in Astronomical Society of the Pacific Conference Series, Vol. 229, Evolution of Binary and Multiple Star Systems, ed. P. Podsiadlowski, S. Rappaport, A. R. King, F. D'Antona, & L. Burderi, 239
Polidan, R. S. 1976, in Be and Shell Stars, ed. A. Slettebak, Vol. 70, 401
Pols, O. R., Cote, J., Waters, L. B. F. M., & Heise, J. 1991, A&A, 241, 419
Porter, J. M. & Rivinius, T. 2003, PASP, 115, 1153
Pothuneni, R. R., Devarapalli, S. P., & Jagirdar, R. 2023, Research in Astronomy and Astrophysics, 23, 025017
Powell, B. P., Kruse, E., Montet, B. T., et al. 2022, Research Notes of the American Astronomical Society, 6, 111
Prochaska, J., Hennawi, J., Westfall, K., et al. 2020, The Journal of Open Source Software, 5, 2308
Queiroz, A. B. A., Anders, F., Chiappini, C., et al. 2023, A&A, 673, A155
Raskin, G., van Winckel, H., Hensberge, H., et al. 2011, A&A, 526, A69
Reguitti, A., Pastorello, A., & Valerin, G. 2025, arXiv e-prints, arXiv:2504.14592
Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2015, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 014003
Riello, M., De Angeli, F., Evans, D. W., et al. 2021, A&A, 649, A3
Rivinius, T., Carciofi, A. C., & Martayan, C. 2013, A&A Rev., 21, 69
Röpke, F. K. & De Marco, O. 2023, Living Reviews in Computational Astrophysics, 9, 2
Sana, H., de Mink, S. E., de Koter, A., et al. 2012, Science, 337, 444
Sánchez, S. F., Aceituno, J., Thiele, U., Pérez-Ramírez, D., & Alves, J. 2007, PASP, 119, 1186
Sánchez, S. F., Thiele, U., Aceituno, J., et al. 2008, PASP, 120, 1244
Scargle, J. D. 1982, ApJ, 263, 835
Schneider, R., Ferrari, V., Matarrese, S., & Portegies Zwart, S. F. 2001, MNRAS, 324, 797
Shao, Y. & Li, X.-D. 2014, ApJ, 796, 37
Soker, N. & Tylenda, R. 2003, ApJ, 582, L105
Stoppa, F., Johnston, C., Cator, E., Nelemans, G., & Groot, P. J. 2024, A&A, 686, A158
STScI Development Team. 2013, pysynphot: Synthetic photometry software package, Astrophysics Source Code Library, record ascl:1303.023
Telting, J. H., Avila, G., Buchhave, L., et al. 2014, Astronomische Nachrichten, 335, 41
Tonry, J. L., Denneau, L., Heinze, A. N., et al. 2018, PASP, 130, 064505
Tylenda, R., Hajduk, M., Kamiński, T., et al. 2011, A&A, 528, A114
Tylenda, R., Kamiński, T., Udalski, A., et al. 2013, A&A, 555, A16
Verbunt, F. 1993, ARA&A, 31, 93
Wang, B. & Han, Z. 2012, New A Rev., 56, 122
Webb, N. A., Coriat, M., Traulsen, I., et al. 2020, A&A, 641, A136
Weiler, M., Carrasco, J. M., Fabricius, C., & Jordi, C. 2023, A&A, 671, A52
Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, A&AS, 143, 9
Worters, H. L., O'Connor, J. E., Carter, D. B., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, ed. C. J. Evans, L. Simard, & H. Takami, 99083Y
Yoon, S. C., Woosley, S. E., & Langer, N. 2010, ApJ, 725, 940
Zhang, X., Green, G. M., & Rix, H.-W. 2023, MNRAS, 524, 1855

Appendix A: Additional tables

Table A.1: Information of the candidates in our sample: sky position, distance, apparent magnitude, colour, line-of-sight extinction, mass, surface gravity and effective temperature.

Gaia source ID[1]	RA[1] (deg)	DEC[1] (deg)	Distance[3] (parsec)	G[1] (mag)	BP-RP[1] (mag)	E(B-V)[2] (mag)	E(B-V) (2024)[3] (mag)	M[3] (M$_\odot$)	log g[3] (cgs)	T$_{\mathrm{eff}}$[3] (K)
1872192393430508880	78.5501	37.0042	3091^{+106}_{-101}	13.95	1.51	$0.56^{+0.00}_{-0.05}$	1.14 ± 0.17	10.01 ± 4.57	3.90 ± 0.74	14127 ± 4248
4281036526730497228	7.3050	59.3159	3380^{+208}_{-208}	14.49	1.28	$0.57^{+0.05}_{-0.12}$	0.66 ± 0.14	6.00 ± 13.69	3.84 ± 0.31	9184 ± 1767
4606866489658625258	44.2189	57.2332	2338^{+45}_{-57}	15.72	2.19	$1.30^{+0.01}_{-0.01}$	1.32 ± 0.10	1.81 ± 0.50	3.63 ± 0.31	7258 ± 523
4611936956247755424	44.8338	58.7449	3214^{+928}_{-57}	10.97	1.29	$0.90^{+0.04}_{-0.04}$	1.26 ± 0.28	9.75 ± 37.69	3.12 ± 0.64	34393 ± 13005
4735757777103322496	60.9224	60.2468	3878^{+148}_{-106}	12.71	1.14	$0.29^{+0.04}_{-0.12}$	0.72 ± 0.14	2.91 ± 1.13	2.60 ± 0.60	10371 ± 713
5084193693101909076	30.6300	61.0035	4365^{+288}_{-952}	11.25	0.99	$0.65^{+0.06}_{-0.06}$	0.89 ± 0.21	7.62 ± 6.57	3.36 ± 0.47	19215 ± 7082
5127214447659934721	22.2065	63.8494	5380^{+396}_{-1055}	11.75	1.21	$0.85^{+0.02}_{-0.05}$	1.15 ± 0.16	12.01 ± 69.03	3.54 ± 0.60	18709 ± 6265
5269398824637131527	9.2763	63.7070	4364^{+947}_{-364}	13.63	1.64	$0.69^{+0.05}_{-0.45}$	1.33 ± 0.18	8.75 ± 2.86	3.16 ± 0.49	16675 ± 2881
5271552536049139271	8.1916	64.9694	3067^{+76}_{-76}	13.24	1.86	$0.76^{+0.06}_{-0.11}$	1.03 ± 0.14	4.38 ± 1.57	2.38 ± 0.40	7916 ± 770
1870955515858422656	312.9137	37.5701	2136^{+79}_{-20}	14.16	1.90	$0.75^{+0.05}_{-0.05}$	0.54 ± 0.07	1.25 ± 0.23	2.66 ± 0.20	4765 ± 73
2002117151282819840	341.3075	53.2094	2461^{+20}_{-20}	9.79	0.21	$0.14^{+0.06}_{-0.06}$	0.50 ± 0.11	7.74 ± 2.06	3.59 ± 0.44	15929 ± 1900
2006088484204609408	333.7223	55.6014	4913^{+318}_{-976}	14.23	1.39	$0.72^{+0.06}_{-0.22}$	0.96 ± 0.14	5.85 ± 5.01	3.46 ± 0.37	13909 ± 3858
2006912396372680960	339.2652	57.1667	2455^{+35}_{-33}	12.81	1.31	$0.43^{+0.06}_{-0.03}$	0.80 ± 0.16	5.97 ± 0.74	3.75 ± 0.27	9145 ± 1079
2007318661608788096	341.0511	57.8325	3240^{+612}_{-137}	15.61	1.82	$0.72^{+0.03}_{-0.03}$	0.65 ± 0.10	1.36 ± 0.29	2.85 ± 0.25	5072 ± 134
2013187240507011456	343.6522	58.0566	3153^{+178}_{-178}	14.35	1.87	$1.26^{+0.04}_{-0.04}$	0.98 ± 0.19	5.53 ± 27.20	3.39 ± 0.35	8676 ± 1173
2027563494289195520	298.5404	27.3946	7715^{+1371}_{-1371}	14.89	2.23	$1.34^{+0.28}_{-0.28}$	1.52 ± 0.15	6.20 ± 1.51	2.56 ± 0.63	10210 ± 1630
2030965725082217088	300.0884	31.6290	2677^{+2259}_{-42}	13.21	1.75	$1.15^{+0.11}_{-0.11}$	1.12 ± 0.16	5.45 ± 1.43	3.48 ± 0.54	10482 ± 799
2054338249889402880	305.1283	32.5679	3091^{+328}_{-170}	13.92	2.30	$1.65^{+0.07}_{-0.05}$	1.78 ± 0.50	7.22 ± 2.78	3.82 ± 0.55	19691 ± 2232
2060841448854265216	303.4874	37.5023	3668^{+30}_{-30}	14.57	2.30	$1.44^{+0.08}_{-0.08}$	1.40 ± 0.15	3.51 ± 0.56	2.17 ± 0.32	5089 ± 173
2061252975440642816	304.7922	38.2985	3869^{+1742}_{-948}	16.33	2.13	$0.97^{+0.04}_{-0.04}$	1.62 ± 0.15	2.07 ± 1.14	3.32 ± 0.39	6629 ± 614
2074693061975359232	302.9374	41.7822	2242^{+8}_{-8}	13.23	1.10	$0.44^{+0.03}_{-0.02}$	0.62 ± 0.17	3.69 ± 0.60	3.77 ± 0.25	8711 ± 628
2083649030845658624	307.8765	47.9782	2554^{+63}_{-49}	15.64	1.98	$0.89^{+0.01}_{-0.01}$	0.82 ± 0.18	1.39 ± 0.37	3.15 ± 0.33	4912 ± 216
2164630463117114496	319.5385	47.5318	3023^{+233}_{-49}	13.41	1.77	$0.79^{+0.11}_{-0.04}$	1.39 ± 0.16	5.47 ± 1.04	3.86 ± 0.65	11948 ± 1467
2166378312964576256	312.6099	46.0180	3524^{+70}_{-48}	14.63	2.52	$1.20^{+0.00}_{-0.00}$	1.88 ± 0.15	5.97 ± 0.81	2.43 ± 0.38	11232 ± 756
2169083000475385856	317.1013	50.1992	3797^{+276}_{-528}	13.83	3.22	$2.21^{+0.31}_{-0.31}$	2.24 ± 0.31	5.63 ± 1.24	1.71 ± 0.49	6603 ± 717
2173852964799716480	326.9270	53.9998	6441^{+338}_{-395}	13.65	1.80	$1.26^{+0.01}_{-0.07}$	1.46 ± 0.16	10.61 ± 4.60	3.26 ± 0.91	13979 ± 2477
2175699216614191360	319.3053	53.4088	4169^{+82}_{-395}	14.98	2.52	$1.73^{+0.08}_{-0.08}$	1.67 ± 0.12	4.64 ± 1.03	2.38 ± 0.47	9464 ± 557
2200434313577635840	336.7675	59.0930	3097^{+252}_{-170}	12.23	1.67	$1.19^{+0.03}_{-0.03}$	1.32 ± 0.19	$10.85 \pm 3.47 \times 10^7$	3.26 ± 0.63	19734 ± 3965
2934216142176785920	107.0585	-18.5093	3113^{+37}_{-37}	9.60	1.66	$0.81^{+0.06}_{-0.02}$	0.41 ± 0.07	5.14 ± 1.60	1.37 ± 0.41	4588 ± 357
3355776901779440384	99.6750	13.7071	6793^{+124}_{-1243}	9.71	0.41	$0.22^{+0.05}_{-0.03}$	0.30 ± 0.13	10.42 ± 1.33	3.47 ± 0.53	20345 ± 3824
3369390992328121760	97.8989	16.4806	3311^{+495}_{-48}	11.65	0.64	$0.42^{+0.03}_{-0.09}$	0.63 ± 0.13	6.89 ± 0.71	3.42 ± 0.59	10909 ± 1593
3441683251631390840	86.0562	29.1531	2014^{+109}_{-172}	10.53	1.03	$0.48^{+0.13}_{-0.13}$	1.03 ± 0.28	11.96 ± 50.69	3.81 ± 0.53	23089 ± 11545
4054010697162430592	266.6885	-32.6311	1131^{+13}_{-10}	9.71	1.43	$0.44^{+0.04}_{-0.01}$	0.96 ± 4.55	4.02 ± 0.78	1.84 ± 0.33	7136 ± 718
4076568618334521600	280.5377	-24.5018	5576^{+22}_{-29}	11.84	1.17	$0.36^{+0.01}_{-0.01}$	0.38 ± 0.13	1.85 ± 1.23	1.68 ± 1.24	5623 ± 528
4094491141885400576	275.0983	-19.7796	5890^{+180}_{-195}	13.27	2.59	$1.53^{+0.02}_{-0.02}$	1.65 ± 0.18	7.40 ± 1.50	1.72 ± 0.53	9585 ± 1492
4096527235637366912	275.4238	-17.6223	1970^{+34}_{-578}	10.36	0.74	$0.40^{+0.02}_{-0.07}$	0.58 ± 0.13	4.45 ± 1.47	3.63 ± 0.39	14307 ± 2117
4260141158544875008	280.6910	-1.3850	3172^{+18}_{-540}	15.61	3.11	$1.79^{+0.07}_{-0.05}$	1.43 ± 0.07	1.51 ± 0.29	2.19 ± 0.27	4485 ± 89

Table A.1: continued.

Gaia source ID[1]	RA[1] (deg)	DEC[1] (deg)	Distance[3] (parsec)	G[1] (mag)	BP-RP[1] (mag)	E(B-V)[2] (mag)	E(B-V) (2024)[3] (mag)	M[3] ($M_\odot$)	$\log g$[3] (cgs)	T_{eff}[3] (K)
4263591911398361472	289.9265	-0.0499	1031^{+8}_{-4}	8.57	0.44	$0.24^{+0.01}_{-0.00}$	0.36 ± 0.19	3.95 ± 1.38	3.42 ± 0.39	10853 ± 2581
4271992661242707200	278.4116	-1.1951	1260^{+96}_{-7}	15.59	3.89	$2.67^{+0.09}_{-0.09}$	2.22 ± 0.11	1.40 ± 0.29	2.55 ± 0.18	4744 ± 124
4272583560222995200	278.9687	-0.0870	1904^{+118}_{-111}	15.40	3.74	$2.34^{+0.08}_{-0.08}$	1.90 ± 0.10	1.66 ± 0.39	1.84 ± 0.26	4510 ± 124
4281886474885416064	282.0665	4.0900	1036^{+128}_{-88}	15.34	3.05	$2.01^{+0.03}_{-0.03}$	1.39 ± 0.12	1.07 ± 0.18	2.99 ± 0.16	4617 ± 130
4299045198336460800	298.9082	8.9623	7318^{+770}_{-46}	9.58	0.55	$0.23^{+0.25}_{-0.25}$	0.13 ± 0.07	2.86 ± 0.56	3.00 ± 0.71	6836 ± 226
4321276689423536384	292.2191	14.8869	2927^{+611}_{-1579}	15.12	2.78	$1.49^{+0.05}_{-0.05}$	1.26 ± 0.09	1.71 ± 0.40	2.14 ± 0.27	4598 ± 128
4515124540765147776	290.2013	18.4955	2496^{+155}_{-55}	16.48	4.09	$2.62^{+0.08}_{-0.08}$	1.96 ± 0.17	1.39 ± 0.35	1.97 ± 0.43	3879 ± 107
4519475166529738112	286.9374	20.3501	4917^{+1218}_{-80}	13.29	2.71	$1.45^{+0.04}_{-0.27}$	0.81 ± 0.26	1.12 ± 0.32	0.74 ± 0.31	3764 ± 92
5311969857556479616	138.7208	-51.7624	3876^{+156}_{-173}	14.33	1.92	$0.76^{+0.01}_{-0.07}$	1.49 ± 0.17	6.73 ± 2.96	3.10 ± 0.41	15212 ± 3692
5323384162646755712	134.6810	-53.1388	3757^{+528}_{-173}	16.00	1.75	$0.85^{+0.15}_{-0.20}$	0.69 ± 0.10	1.42 ± 0.42	2.70 ± 0.22	5446 ± 131
5326288831829996416	140.1741	-48.1982	2595^{+155}_{-125}	17.25	3.37	$2.06^{+0.00}_{-0.00}$	1.85 ± 0.14	1.75 ± 0.42	2.73 ± 0.19	4818 ± 123
5328449200388495616	133.0337	-48.5787	4153^{+48}_{-6}	16.19	1.98	$0.90^{+0.00}_{-0.00}$	0.75 ± 0.10	1.26 ± 0.36	2.62 ± 0.24	5025 ± 109
5338183383022960512	163.8516	-60.8902	5858^{+275}_{-351}	14.55	1.30	$0.82^{+0.14}_{-0.01}$	0.92 ± 0.18	3.82 ± 2.86	2.66 ± 0.37	10494 ± 2176
5350869719969619840	162.5043	-58.2992	2457^{+78}_{-181}	10.54	1.00	$0.69^{+0.05}_{-0.07}$	0.61 ± 0.12	3.71 ± 7.33	2.13 ± 0.69	7552 ± 626
5524022735225482624	130.8649	-43.5837	2594^{+56}_{-25}	13.70	2.33	$1.52^{+0.07}_{-0.07}$	1.88 ± 0.16	5.43 ± 1.31	3.55 ± 0.79	13174 ± 1351
5593826360487373696	117.0976	-34.3848	4244^{+219}_{-114}	12.01	2.36	$1.64^{+0.04}_{-0.04}$	1.56 ± 0.55	8.33 ± 125.53	2.02 ± 0.50	11664 ± 1513
5599309216965305728	114.9433	-29.8086	3229^{+416}_{-455}	10.41	0.25	$0.17^{+0.04}_{-0.04}$	0.42 ± 0.07	5.83 ± 1.17	2.72 ± 0.63	12988 ± 1080
5617186348318629248	107.9070	-25.5237	2845^{+182}_{-404}	14.06	1.27	$0.30^{+0.06}_{-0.06}$	0.33 ± 0.08	1.12 ± 0.20	2.82 ± 0.22	5309 ± 167
5866345647515558400	209.6089	-61.6174	2501^{+20}_{-4}	14.97	3.10	$1.85^{+0.04}_{-0.04}$	1.46 ± 0.12	1.76 ± 0.40	1.96 ± 0.21	4391 ± 78
5866474526572151936	211.2822	-60.8993	2710^{+20}_{-197}	12.39	1.41	$0.97^{+0.06}_{-0.06}$	1.16 ± 0.27	11.88 ± 26.53	3.13 ± 0.41	30216 ± 7958
5868425648663616768	201.6601	-61.9116	3706^{+595}_{-8}	15.68	3.25	$1.93^{+0.13}_{-0.44}$	1.67 ± 0.13	1.87 ± 0.67	1.89 ± 0.25	4470 ± 96
5880159842877908352	228.1479	-57.9014	2948^{+84}_{-20}	14.10	2.87	$2.24^{+0.02}_{-0.13}$	1.65 ± 0.10	2.81 ± 0.52	1.95 ± 0.36	7429 ± 293
5882737819707242240	235.7697	-56.5290	2067^{+213}_{-20}	11.56	1.31	$0.51^{+0.08}_{-0.08}$	1.16 ± 0.22	6.38 ± 1.45	3.29 ± 0.52	17763 ± 5569
5962956195185292288	257.2450	-46.5970	3989^{+129}_{-58}	13.62	3.80	$2.48^{+0.02}_{-0.02}$	1.76 ± 0.20	1.60 ± 0.48	0.86 ± 0.32	4340 ± 162
5965503866703572480	256.8225	-42.4636	3090^{+72}_{-77}	14.13	2.36	$1.71^{+0.00}_{-0.00}$	1.73 ± 0.15	7.65 ± 7.00	3.29 ± 0.44	18218 ± 5978
5966574133883852416	254.5209	-41.4761	2496^{+161}_{-36}	17.00	3.04	$2.02^{+0.32}_{-0.02}$	1.77 ± 0.15	1.96 ± 0.70	3.14 ± 0.29	5742 ± 268
6052463412400724992	248.7806	-20.7259	5158^{+63}_{-1054}	16.50	1.75	$0.66^{+0.02}_{-0.04}$	0.47 ± 0.08	0.91 ± 0.11	3.56 ± 0.34	5029 ± 157
6053890788968694656	189.2723	-62.4698	3559^{+62}_{-223}	14.07	2.37	$1.58^{+0.00}_{-0.00}$	1.80 ± 0.15	8.59 ± 4.78	3.28 ± 0.55	17548 ± 2141
6123873398383875456	212.0047	-33.2620	1316^{+139}_{-118}	11.56	0.64	$0.07^{+0.01}_{-0.07}$	0.04 ± 0.02	1.58 ± 0.26	3.75 ± 0.19	6469 ± 106
6228685649971375616	223.1954	-25.7899	7094^{+118}_{-3144}	15.15	0.83	$0.13^{+0.01}_{-0.05}$	0.08 ± 0.03	1.01 ± 0.14	3.75 ± 0.17	6112 ± 113

References. (1) Gaia Collaboration et al. (2023); (2) Anders et al. (2022); (3) Khalatyan et al. (2024)

Table A.2: Observation date, instrument and instrument details for the follow-up spectra obtained for a subsample of our candidates.

Index	Gaia DR3 ID	Date	Telescope/Instr.	Setup	Range (Å)	R
1	1870955515858422656	2024-04-14 03:38:36	CAHA/CAFOS	G-200/1.5″ Slit	4000–8500	≈300
2	1872192393430508880	2024-10-30 03:59:50	NOT/ALFOSC	G-19/0.5″ Slit	4400–6950	1940
	1872192393430508880	2024-10-30 05:17:55	NOT/ALFOSC	G-7/1″ Slit	3650-7110	650
	1872192393430508880	2012-02-19 12:01:14	LAMOST	Low res	3700–9000	1800
	1872192393430508880	2014-01-29 12:16:59	LAMOST	Low res	3700–9000	1800
	1872192393430508880	2014-10-17 19:17:55	LAMOST	Low res	3700–9000	1800
	1872192393430508880	2024-10-29 04:05:44	NOT/FIES-L	Low res	3700–8300	25000
	1872192393430508880	2024-10-29 04:36:32	NOT/FIES-L	Low res	3700–8300	25000
	1872192393430508880	2024-10-30 02:52:46	NOT/FIES-L	Low res	3700–8300	25000
	1872192393430508880	2024-10-30 03:23:33	NOT/FIES-L	Low res	3700–8300	25000
3	2006088484204609408	2024-10-30 00:11:18	NOT/ALFOSC	G-7/1″ Slit	3650-7110	650
	2006088484204609408	2024-10-28 20:57:13	NOT/FIES-L	Low res	3700–8300	25000
	2006088484204609408	2024-10-28 23:05:09	NOT/FIES-L	Low res	3700–8300	25000
	2006088484204609408	2024-10-29 00:33:53	NOT/FIES-L	Low res	3700–8300	25000
	2006088484204609408	2024-10-29 21:24:41	NOT/FIES-L	Low res	3700–8300	25000
	2006088484204609408	2024-10-29 21:55:28	NOT/FIES-L	Low res	3700–8300	25000
4	2013187240507011456	2024-04-14 03:45:35	CAHA/CAFOS	G-200/1.5″ Slit	4000–8500	≈300
5	2030965725082217088	2025-06-02 03:17:24	NOT/ALFOSC	G-4/1″ Slit	3200–9600	360
6	2060841448854265216	2024-04-14 03:55:16	CAHA/CAFOS	G-200/1.5″ Slit	4000–8500	≈300
7	2061252975440642816	2024-10-29 19:52:13	NOT/ALFOSC	G-19/0.5″ Slit	4400–6950	1940
	2061252975440642816	2024-10-29 20:02:23	NOT/ALFOSC	G-19/0.5″ Slit	4400–6950	1940
8	2074693061975359232	2024-04-13 04:06:39	CAHA/CAFOS	G-200/1.5″ Slit	4000–8500	≈300
9	2164630463117114496	2024-04-14 03:25:02	CAHA/CAFOS	G-200/1.5″ Slit	4000–8500	≈300
10	2166378312964576256	2025-06-02 02:56:27	NOT/ALFOSC	G-4/1″ Slit	3200–9600	360
11	2175699216614191360	2024-10-30 00:02:17	NOT/ALFOSC	G-7/1″ Slit	3650-7110	650
	2175699216614191360	2025-06-02 03:03:37	NOT/ALFOSC	G-4/1″ Slit	3200–9600	360
12	2200433413577635840	2024-04-14 04:10:48	CAHA/CAFOS	G-200/1.5″ Slit	4000–8500	≈300
	2200433413577635840	2024-05-27 01:16:01	NOT/FIES-M	Med res	3700–8300	46000
13	3355776901779440384	2024-10-29 05:46:33	NOT/FIES-L	Low res	3700–8300	25000
14	3369399099232812160	2024-10-30 06:03:12	NOT/ALFOSC	G-4/0.5″ Slit	3200–9600	710
	3369399099232812160	2024-10-29 05:11:32	NOT/FIES-L	Low res	3700–8300	25000
	3369399099232812160	2024-10-29 05:27:19	NOT/FIES-L	Low res	3700–8300	25000
	3369399099232812160	2024-10-30 04:37:57	NOT/FIES-L	Low res	3700–8300	25000
	3369399099232812160	2024-10-30 04:53:44	NOT/FIES-L	Low res	3700–8300	25000
15	3444168325163139840	2024-10-29 01:32:19	NOT/FIES-L	Low res	3700–8300	25000
16	4076568861833452160	2024-05-27 00:37:58	NOT/FIES-M	Med res	3700–8300	46000
17	428103652673049728	2024-10-30 00:27:35	NOT/ALFOSC	G-7/1″ Slit	3650-7110	650
18	4299904519833646080	2024-05-26 02:21:19	NOT/FIES-M	Med res	3700–8300	46000
19	461193695624775424	2024-10-29 03:50:56	NOT/FIES-L	Low res	3700–8300	25000
20	4735757771103322496	2024-10-30 00:38:10	NOT/ALFOSC	G-7/1″ Slit	3650-7110	650
21	508419369310190976	2024-10-29 01:17:57	NOT/FIES-L	Low res	3700–8300	25000
22	512721444765993472	2024-10-30 04:08:31	NOT/ALFOSC	G-4/0.5″ Slit	3200–9600	710
	512721444765993472	2024-10-29 22:29:36	NOT/FIES-L	Low res	3700–8300	25000
	512721444765993472	2024-10-29 22:40:23	NOT/FIES-L	Low res	3700–8300	25000
23	526939882463713152	2024-10-30 01:39:21	NOT/ALFOSC	G-4/0.5″ Slit	3200–9600	710
	526939882463713152	2024-10-29 01:49:16	NOT/FIES-L	Low res	3700–8300	25000
	526939882463713152	2024-10-29 02:20:04	NOT/FIES-L	Low res	3700–8300	25000
	526939882463713152	2024-10-30 01:46:22	NOT/FIES-L	Low res	3700–8300	25000
	526939882463713152	2024-10-30 02:17:09	NOT/FIES-L	Low res	3700–8300	25000
24	527155253604491392	2024-10-30 00:21:14	NOT/ALFOSC	G-7/1″ Slit	3650-7110	650
25	6123873398383875456	2024-05-11 23:09:48	Mercator/HERMES		3770–9000	85000
	6123873398383875456	2024-05-13 23:03:37	Mercator/HERMES		3770–9000	85000
26	5962956195185292288	2025-10-29 18:00:00	Lesedi/Mookodi	4″ Slit	4000–8000	≈175

Notes. The date of observations is in format YYYY-MM-DD HH:MM:SS UT. We aimed to achieve a signal-to-noise ratio (SNR) greater than 80 for the low-resolution spectra (CAFOS, ALFOSC, and Mookodi), and an SNR above 40 for the higher-resolution spectra (FIES and HERMES).

Table A.3: Summary of our results, where the spectral classification and properties observed in the light curve and spectra of our candidates, together with the periods found, are reported.

Gaia DR3 ID	Spec. class (Gaia)	Properties	Gaia-Xmatch Class (Subclass)	Simbad Class (Subclass)	Period (this work) (days)	Period (Gaia-Xmatch) (days)
1872192393430508880	B (B)	Diff,SR,Hα PCyg	RAD_VEL_VAR	EmLine* (*|Em*)		6.02047
4281036526730497728	Be (G)	Diff,Hα emi				0.15669
4606866489586252528	Be (B)	Sin,Diff	DSCT	delSctV* (dS*)	0.15669	
4611936956247755424	Be	Sin,2p Hα emi		Star (*|NIR)	0.39407	
4735757771103322496	B	Sin,Outb,Far UV,2p Hα emi			0.45878	
5084193693101909976	Oe (O)	Sin,Outb,Hα abs		EmLine* (*|Em*|NIR)	0.4011 , 1.4682	
5127214447659934720	Be (B)	Sin,2p Hα emi		Star (*|NIR)	0.57251	
5269398824637131520	Be (B)	Outb,2p Hα emi				
5271552536044913920	Be (B)	Puls,Diff,2p Hα emi	L	SB* (*|Em*|NIR|SB*)	1.00037	149.15505
1870955515858422656	K (K)	Sin,Diff			23.10592	
2002171512828198400	(B)	ECL,SR	ECL (EA)	EclBin (*|EB*|Em*|NIR)	2.58051	
2006088484204609408	Be (B)	ECL,Diff,SR,2p Hα emi	OMIT	EclBin (*|EB*)	2.59606	5.19160
2006912396372680960	(G)	Outb	SOLAR_LIKE (RS)	Variable* (*|MIR|NIR|Opt|V*)		1488.28626*
2007318661608788096	(K)	Diff		RSCVnV* (*|MIR|NIR|Opt|RS*|V*)	0.35344 , 0.6081	12.32042
2013187240507011456	Oe (O)	Sin,Diff,SR,Hα emi		EmLine* (*|Em*)		
2027563492489195520	(B)	Outb	OMIT	Variable* (*|MIR|NIR|Opt|V*)	0.43859	224.60175*
2030965725082217088	Be (B)	Sin,SR,Hα emi				
2054338249889402880	G (G)		YSO	EmLine* (*|Em*)		
2060841448854265216	? (B)	ECL,Hα emi	ECL (EW)	EclBin (*|EB*|MIR|NIR|Opt|V)	60.54500	60.70308
2061252975440642816	Be (G)	ECL,Hα abs	ECL (EA)	** (*|***|MIR|NIR|Opt|V*)	2.13918	2.13901
2074690306197535232	(G)	Sin,Diff,SR,Hα emi		EmLine* (*|Em*)	0.11275, 0.1328	
2083649030845658624	? (B)	Hard X-ray				
2164630463117114496	Be (B)	Sin,Diff,SR,Hα emi		EmLine* (*|Em*)	0.59151	
2166378312964576256	(O)	ECL,Hα emi	ECL (EA)	EclBin (*|EB*|Em*|MIR|NIR|Opt|V*)	72.42781	72.56402
2169083000847538560	(O)					
2173852964799716480	? (B)	Outb		EmLine* (*|Em*)		
2175699216614191360	Be (O)	Sin,Diff,Hα emi	OMIT	LPV*_Candidate (*|Em*|LP?|NIR)	4.76998	61.82265*
2200433413577635840	(K)	Sin,Outb,2p Hα emi	L	LPV* (*|LP|NIR|V*|V*?)	0.55163	
2934216142176785920	(K)	Diff		Star (*|NIR|UV)	0.30354	
3355776901779440384	Be (B)	Sin,SR,2p Hα emi	OMIT	Star (*|NIR)	0.47404	554.63452*
3369399099232812160	Be (B)	Sin,2p Hα emi		EmLine* (*|Em*|NIR|Opt)	0.52602	
3444168325163139840	Oe (O)	Sin,Outb,2p Hα emi		EmLine* (*|Em*|NIR|SB*)	0.56017	
4054010697162430592	(B)	Sin,SR,Hard X-ray				
4076568861833452160	F (F)	Sin,Hα abs	CEP (CW)	LPV*_Candidate (*|IR|LP?|MIR|NIR)	28.45017	28.45893
4094491141885400576	(O)		GCAS			
4096527235637366912	(B)			Star (*|NIR)		
4260141158544875008	(B)					
4263591911398361472	(K)	Sin,SR	CST	Be* (*|Be*|Em*|NIR|UV)	1.43548	
4271992661242707200	(O)		LPV (SR)	LPV* (LP*)		60.73273*
4272583356022299520	(K)					
4281886474885416064	(K)	Diff	LPV (SR)	LPV* (LP*)		46.35025

Table A.3: continued.

Gaia DR3 ID	Spec. class (Gaia)	Properties	Gaia-Xmatch Class (Subclass)	Simbad Class (Subclass)	Period (this work) (days)	Period (Gaia-Xmatch) (days)
4299904519833646080	A-F, shell? (F)	Hα abs	OMIT	Star (*|NIR)		
4321276689423536384		Diff,SR		LPV*_Candidate (*|LP?|NIR)		
4515124540765147776		Diff		LPV*_Candidate (*|C*?|LP?|NIR)		
4519475166529738112	(K)		OMIT	LPV*_Candidate (*|LP?|NIR)		13.7453
5311969857556479616	(B)	ECL	ECL (EB)	EclBin (*|EB*)	13.74443	
5323384162646755712		ECL			2.79201	
5326288831829996416	(K)					
5328449200388495616	(G)					
5338183383022960512	(G)					
5350869719969619840	(B)	Sin		EmLine* (*|Em*|NIR)	0.56011	
5524022735225482624	(O)	ECL	OMIT	EclBin (*|EB*)	7.98856	
5593263630487373696	(F)		OMIT	Variable* (*|MIR|NIR|Opt|V*)		
5599309216965305728	(B)	Sin	OMIT	Star (*|NIR)	0.47592	
5617186348318629248	(G)	Sin			3.87180	
5866345647515558400	(K)	Diff,Hard X-ray				
5866474526572151936	(O)	Outb	OMIT			
5868425648663616768	(K)					
5880159842877908352	(B)	Sin			31.77066	
5882737819707242240	(B)	Outb	OMIT	Star (*|NIR)		
5962956195185292288	M (M)	Sin,Diff	OMIT	LPV* (*|C*?|LP*|NIR)	106.75990	
5965503866703572480	(O)	Sin			0.15252	
5966574133883852416	(K)	Diff				
6052463412400724992	(K)	Diff				
6053890788968694656	(O)	Sin			0.22931	
6123873398383875456	F (F)	Sin,Hα abs	RR (RRC)	Star (*|NIR)	0.39674	0.39674
6228686549971375616	(F)	ECL	ECL	EclBin (*|EB*|MIR|NIR|Opt|V*)	1.27398	1.27391

Notes. Apart from our visual spectral classification, we also report the spectral classes in the *Gaia* DR3 (in parenthesis), and the classes given in the *Gaia* cross-match table and SIMBAD. The SIMBAD class definitions can be found in the *SIMBAD: Object types* documentation (https://simbad.cds.unistra.fr/guide/otypes.htx). An asterisk (*) next to the periods from the *Gaia* cross-match indicates those that could not be reproduced with our data.
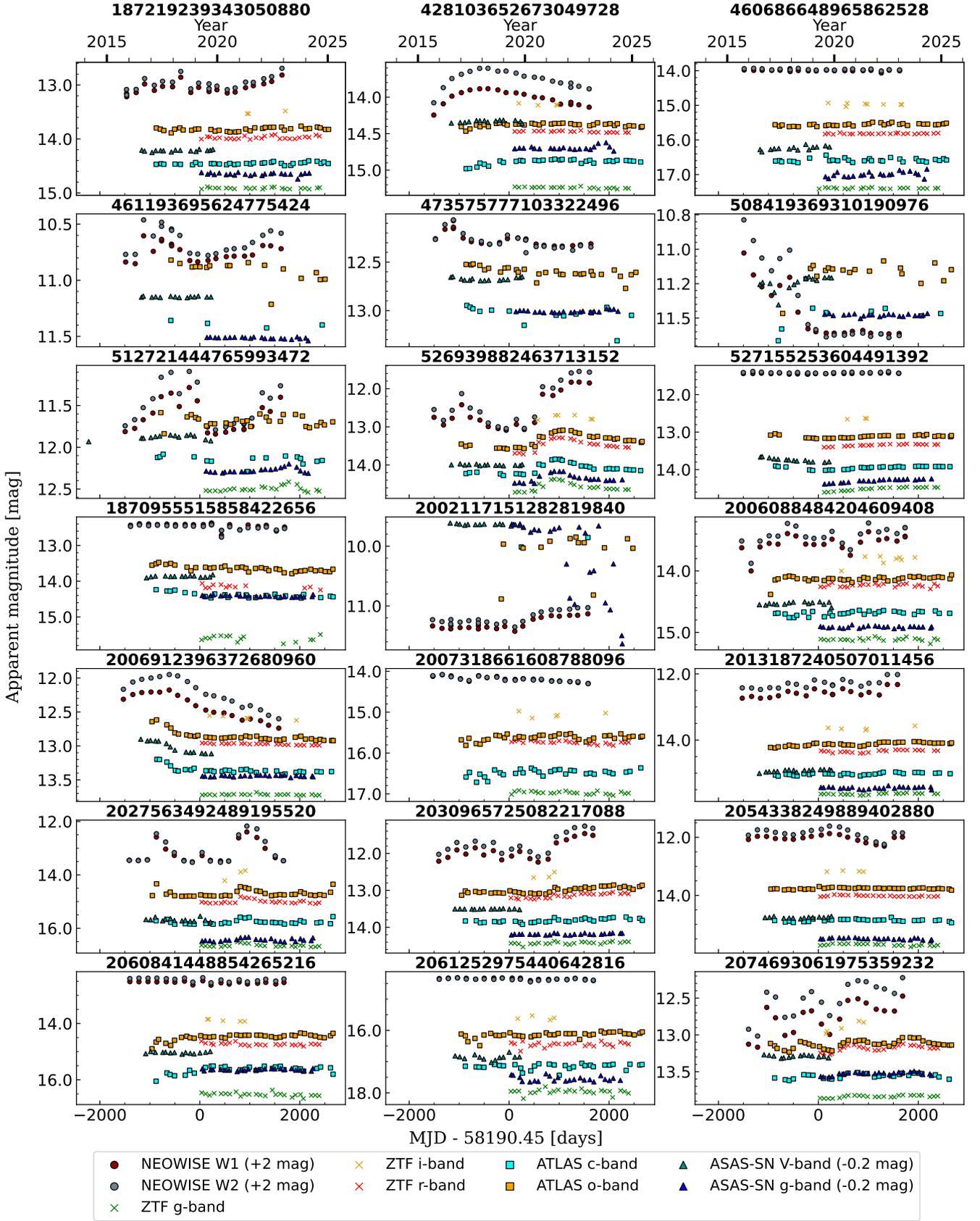
Appendix B: Additional figures

Fig. B.1: Optical and infrared light curves for the sources in our sample, binned in 50-day intervals. The reference Modified Julian Date (MJD) corresponds to the start of observations by ZTF. The number shown above each panel corresponds to the *Gaia* DR3 source ID.
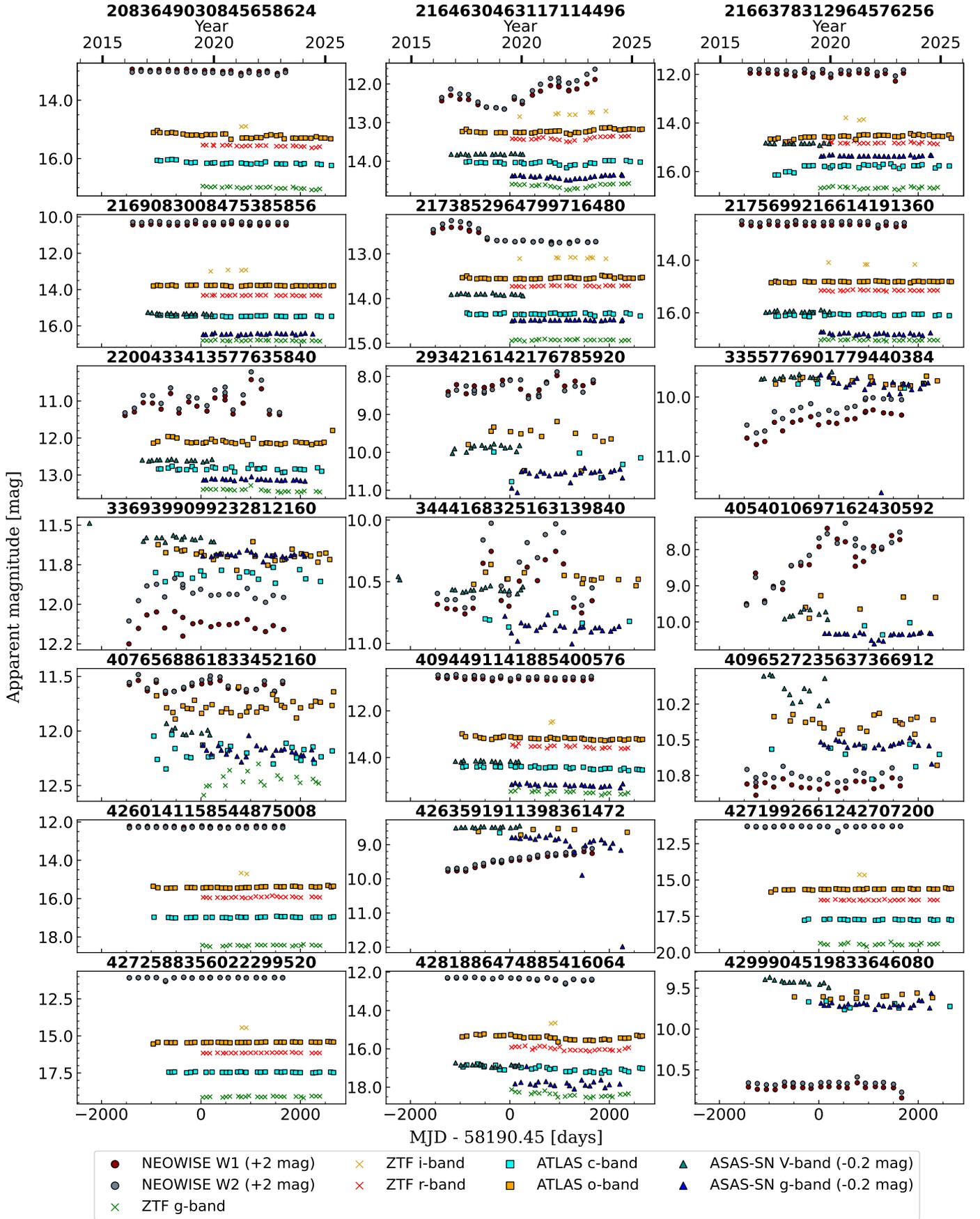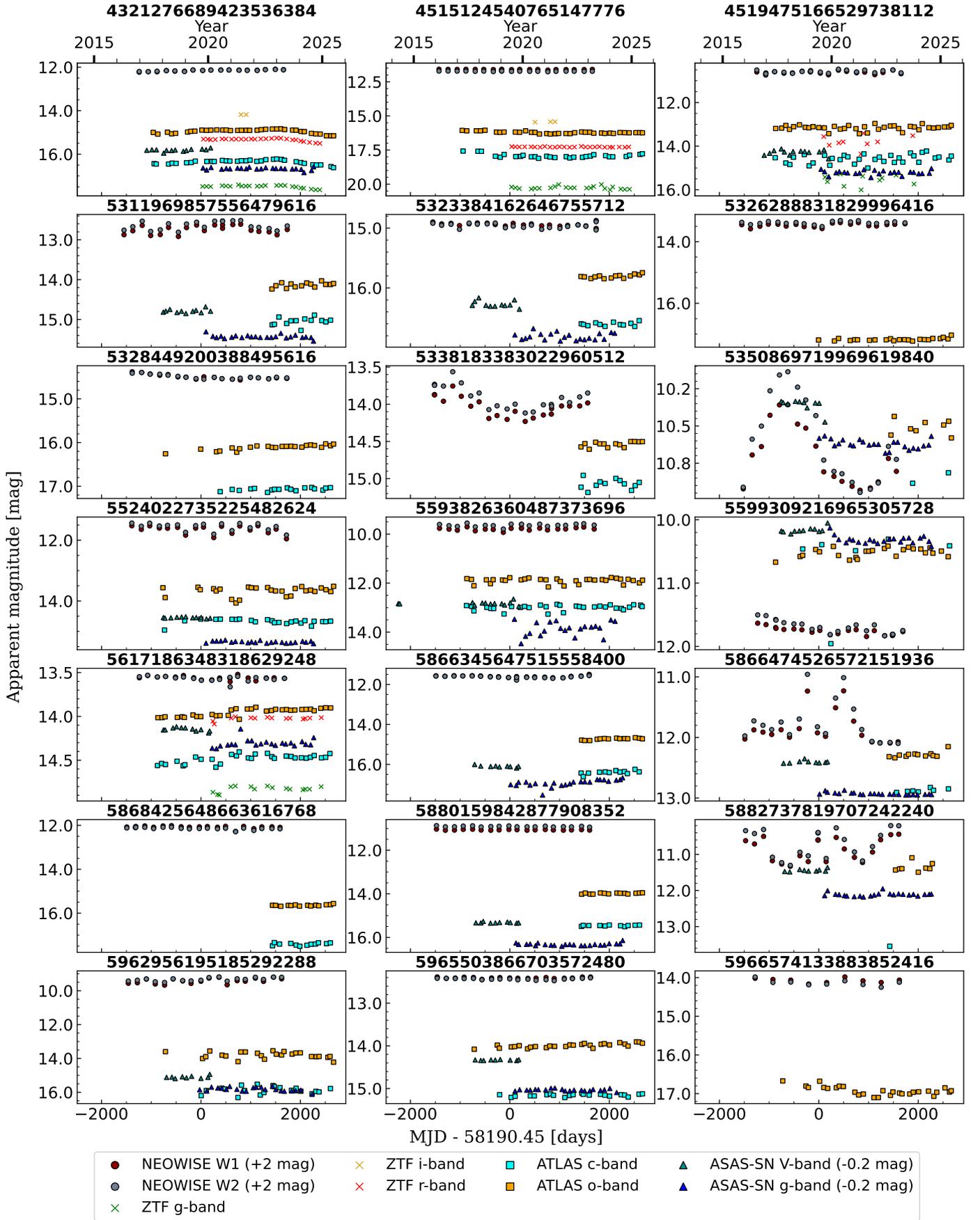
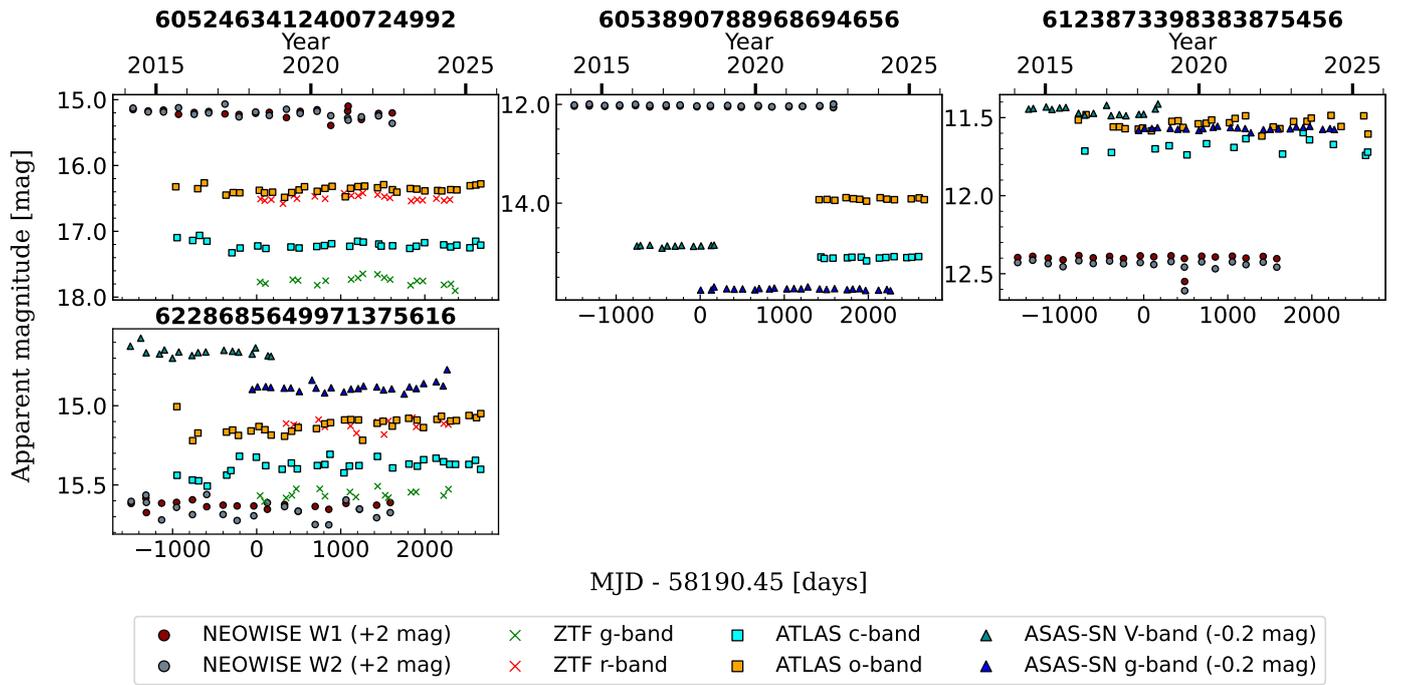Fig. B.1: continued.

Fig. B.1: continued.

Fig. B.1: continued.
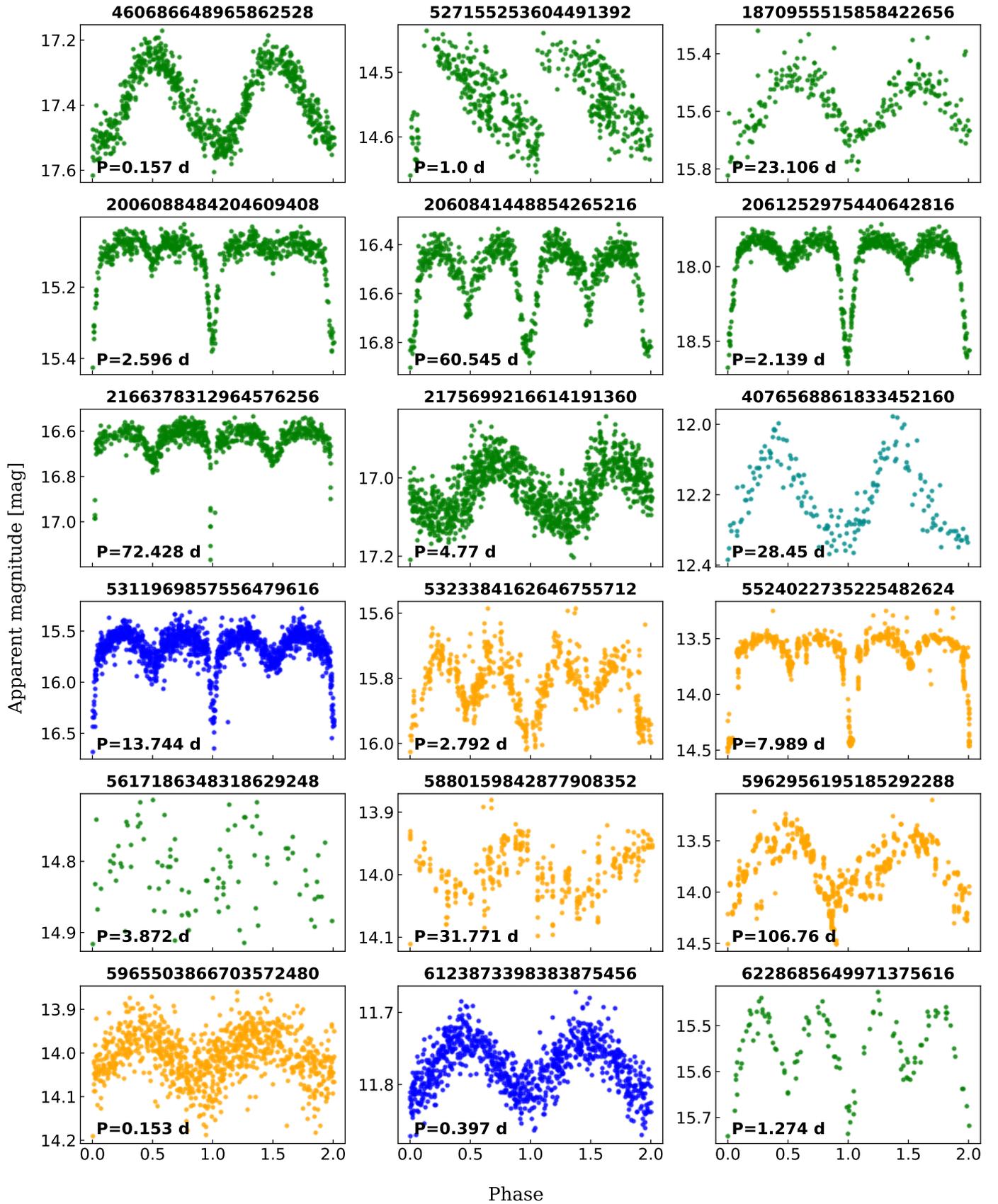
Fig. B.2: Folded light curves of the periodic variables in our sample. Colours represent different bands: ZTF g-band (green), ASAS-SN V-band (cyan) and g-band (blue), and ATLAS o-band (orange). The periods used to fold the light curves are shown in the bottom-left corner of each subplot.
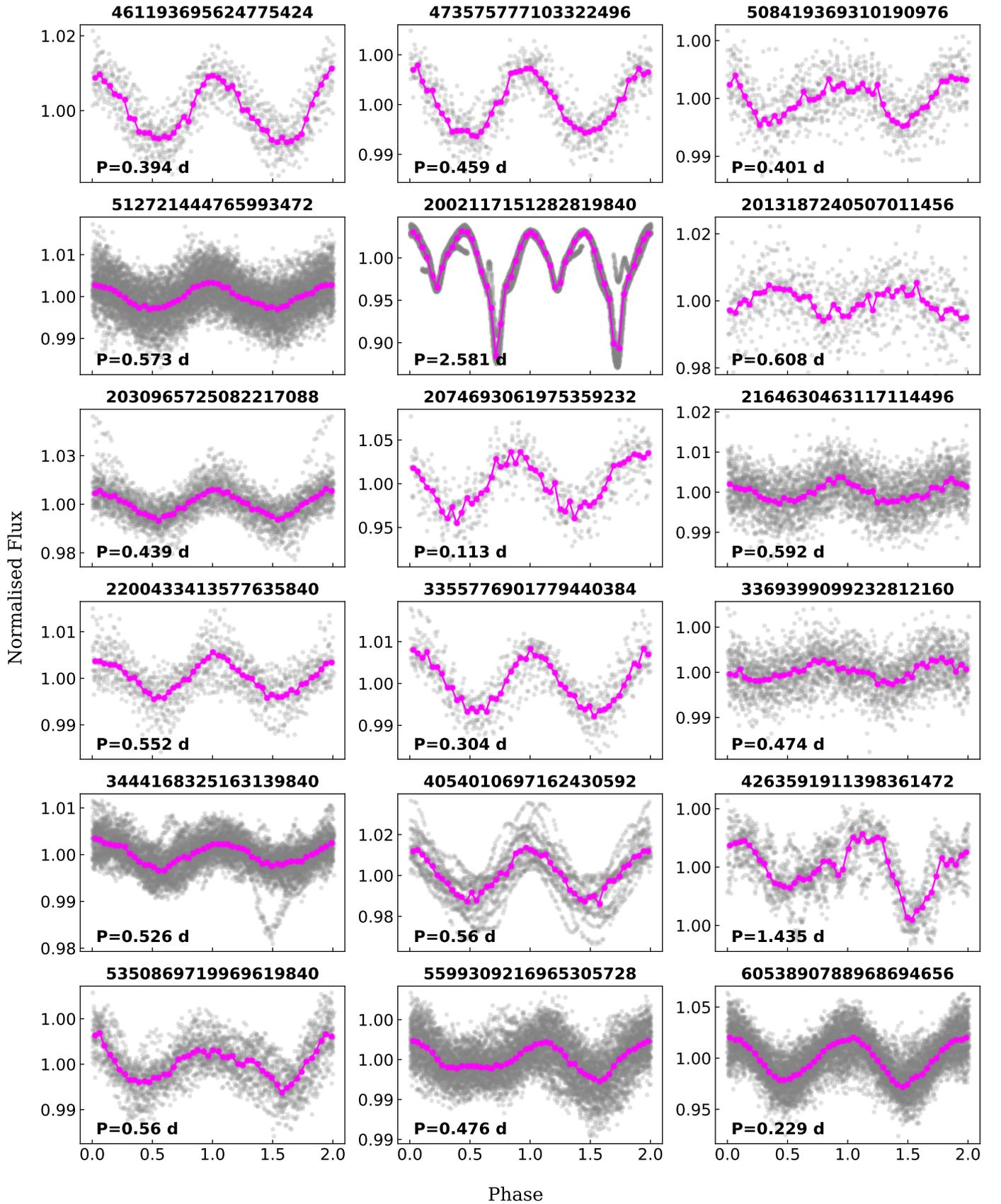
Fig. B.3: Same figure as Fig. B.2, but with the light curve data from TESS. Grey points represent the original unbinned data, while pink points show the data binned into 50 bins, each with a width of 0.04 in phase space.
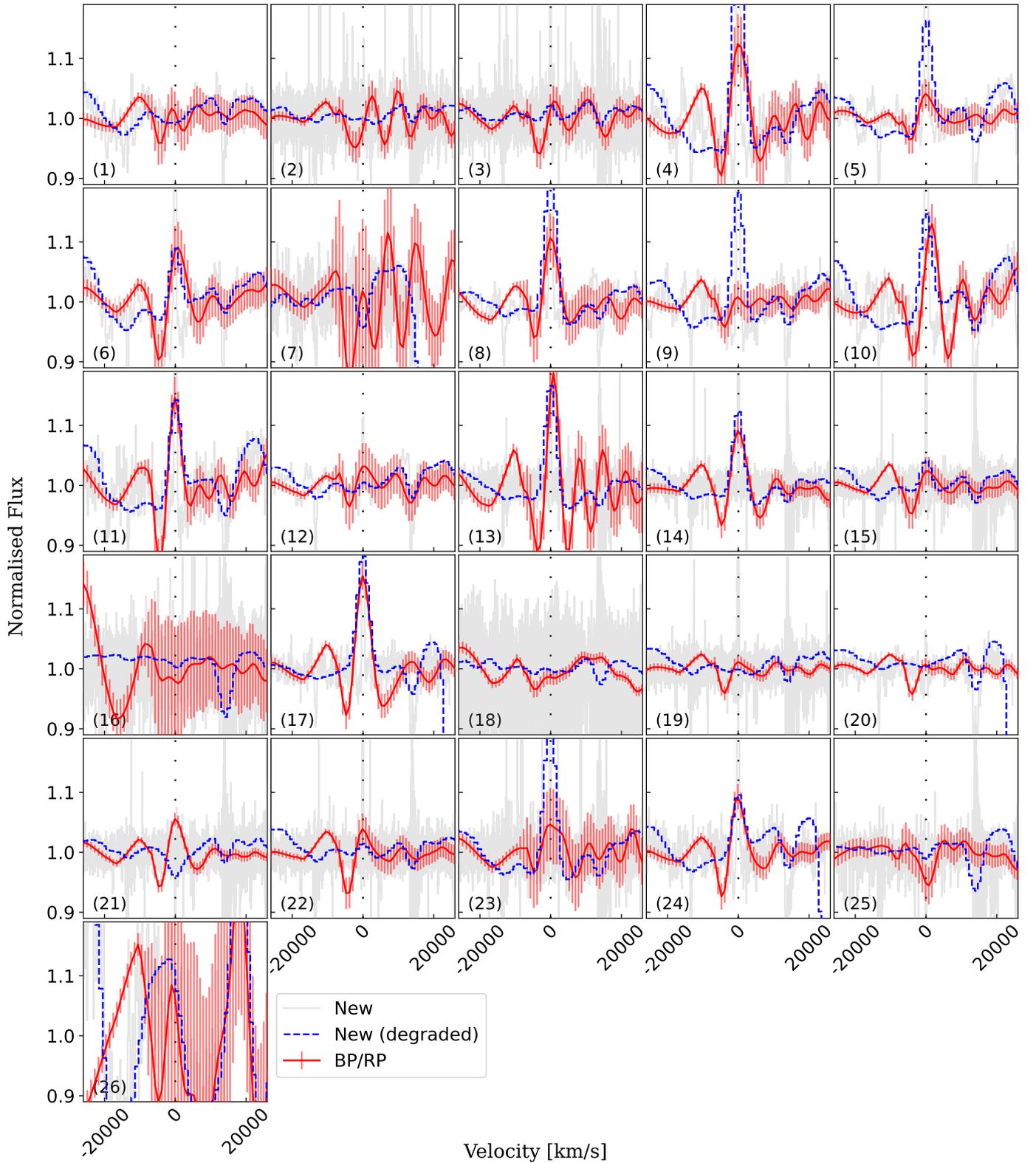
Fig. B.4: Flux normalised Hα velocity profiles for the subsample of observed sources (similar to Fig. 5). We compare the *Gaia* low-resolution BP/RP spectrum (red solid line) with our new follow-up spectrum, degraded to match the BP/RP resolution (blue dashed line). The original-resolution follow-up spectrum is plotted in grey for comparison. All sources are shown on the same scale. The index corresponding to each source (see Table A.2) is shown in the bottom left corner of each spectrum.