Spectroscopic follow-up of a subset of the *Gaia*/IPHAS catalogue of Hα-excess sources

M. Fratta,[1]★ S. Scaringi [ID],[1]★ M. Monguió,[2,3,4] A. F. Pala,[5] J. E. Drew [ID],[6] C. Knigge [ID],[7] K. A. Iłkiewicz[1] and P. Gandhi [ID][7]

[1]*Centre for Extragalactic Astronomy, Department of Physics, University of Durham, South Road, Durham, DH1 3LE, UK*
[2]*Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona (UB), Martí i Franquès 1, E-08028 Barcelona, Spain*
[3]*Departament de Física Quàntica i Astrofísica (FQA), Universitat de Barcelona (UB), Martí i Franquès 1, E-08028 Barcelona, Spain*
[4]*Institut d'Estudis Espacials de Catalunya, Institute of Cosmos Sciences, ICC (IEEC-UB), c. Gran Capità, 2-4, E-08034 Barcelona, Spain*
[5]*European Space Agency, European Space Astronomy Centre, Camino Bajo del Castillo s/n, E-28692 Villanueva de la Cañada, Madrid, Spain*
[6]*Department of Physics and Astronomy, Faculty of Maths and Physical Sciences, University College London, Gower Street, London, WC1E 6BT, UK*
[7]*School of Physics and Astronomy, University of Southampton, University Road, Southampton, SO17 1BJ, UK*

ABSTRACT

State-of-the-art techniques to identify Hα emission-line sources in narrow-band photometric surveys consist of searching for Hα excess with reference to nearby objects in the sky (position-based selection). However, while this approach usually yields very few spurious detections, it may fail to select intrinsically faint and/or rare Hα-excess sources. In order to obtain a more complete representation of the heterogeneous emission-line populations, we recently developed a technique to find outliers relative to nearby objects in the colour–magnitude diagram (CMD-based selection). By combining position-based and CMD-based selections, we built an updated catalogue of Hα-excess candidates in the Northern Galactic Plane. Here, we present spectroscopic follow-up observations and classification of 114 objects from this catalogue that enables us to test our novel selection method. Out of the 70 spectroscopically confirmed Hα-emitters in our sample, 15 were identified only by the CMD-based selection, and would have been thus missed by the classic position-based technique. In addition, we explore the distribution of our spectroscopically confirmed emitters in the *Gaia* CMD. This information can support the classification of emission-line sources in large surveys such as the upcoming WEAVE and *4-m Multi-Object Spectroscopic Telescope*, especially if augmented with the introduction of other colours.

Key words: techniques: photometric – techniques: spectroscopic – stars: emission-line, Be.

1 INTRODUCTION

Unresolved Hα emission can be associated with very heterogeneous populations of both single and binary stars, spanning varied evolutionary stages. This diversity makes the study of emission-line sources particularly important to understand the composition and evolution of our Galaxy. While some of these objects are frequently observed, others might be intrinsically very faint and/or belong to the rarest types of stellar populations. Isolated Hα-emitting sources include young stellar objects (YSOs), coronally active M-dwarfs, luminous blue variables, classical Be stars, and Wolf–Rayet stars. Accretion-driven Hα emission can also be observed from interacting binaries such as cataclysmic variables (CVs), symbiotic stars, and accreting neutron stars or black holes. These types of sources are not only rare, but also difficult to both identify and precisely classify.[1]

Even more challenging is the classification of objects that may not display clear Hα emission-line properties, but rather display Hα excess relative to their parent population. Even though these source types do not present evidence of Hα emission, their intensities in said band might still significantly exceed the ones of the objects they are compared to. An example is provided by the elusive population of low mass transfer rate CVs that have evolved beyond the so-called period minimum (Podsiadlowski, Han & Rappaport 2003; Breedt et al. 2012; Pala et al. 2020). Among the several ways to identify and classify stellar objects, the most telling is a moderate-resolution spectroscopy. However, spectra collection tends to be generally expensive, in particular when attempting to classify large ensembles of targets. Similar issues also affect classification studies based on light-curve analysis.

In order to alleviate the burden, several pre-selection methods have been developed and applied over the years. A popular approach to pre-select targets is based on optical photometry; it deploys specific colour and magnitude cuts to select targets for spectroscopy. Specifically related to Hα emission-line sources, Witham et al. (2008) leveraged photometric measurements from the *Isaac Newton Telescope* Photometric Hα Survey of the Northern Galactic Plane

---

★ E-mail: matteo.fratta@durham.ac.uk (MF); simone.scaringi@durham.ac.uk (SS)
[1]For instance, the list of CVs produced by Ritter & Kolb (2003) includes 1429 objects.

(IPHAS; Drew et al. 2005) to produce a list of ∼5000 Hα-excess candidates. These objects were selected by isolating Hα-bright targets in the r–Hα versus r–i colour–colour diagram (CCD), on a field-by-field basis. More recently, Monguió et al. (2020) produced the IGAPS catalogue, thus expanding and refining the work of Witham et al. (2008). By testing ∼53 000 000 targets for Hα excess, they identified ∼21 000 excess-line candidates relative to spatially nearby sources in the sky, from IPHAS and UV-Excess survey of the Northern Galactic Plane (UVEX; Groot et al. 2009) surveys. These 'position-based' selections are generally rather conservative, and the amount of bogus detections is usually low. While Monguió et al. (2020) did not pursue a spectroscopic confirmation of their selection, Witham et al. (2008) observed the spectra of ∼300 of their Hα-excess candidates. Out of them, 97 per cent were thus confirmed as emission-line sources. However, position-based approaches, due to their conservative nature, may fail to identify some outliers.

In order to include these objects, Fratta et al. (2021; henceforth FR21) developed a novel technique that prioritizes a better representation of the potential Hα-emitting populations. This improvement can sometimes be achieved at the expense of a higher false-positive rate. Taking advantage of the *Gaia* DR2 parallaxes, their method consisted in augmenting the classic position-based selection by identifying Hα-excess candidates relative to nearby sources in the M_G versus G_{BP}–G_{RP} colour–absolute[2] magnitude diagram (CMD)-based selection. However, it is relevant to point out that the absolute magnitudes in the *G* band, as well as the colour indices, were calculated with the assumption of zero extinction, and therefore they should be treated as upper limits. This contributes to increase the effects of population mixing in this parameter space. The authors assigned two *significance values* (or simply '*significances*', n_σ)[3] of Hα excess to each object in their meta-catalogue, one for each partitioning criterion (position-based or CMD-based): These defined the distance from the main stellar locus in IPHAS r–Hα versus r–i parameter space, expressed in units of standard deviation of the r–Hα distribution. The objects with at least one significance higher than three were identified as (3σ) Hα-excess candidates. Out of ∼7.5 million targets, this selection criterion yielded ∼28 000 Hα-excess candidates. To test the statistics of their selection, FR21 visually inspected ∼2000 randomly selected spectra from LAMOST DR5 data base (Cui et al. 2012), relative to a subset of their 3σ outliers. Among these spectra, 49 per cent showed Hα emission; a more conservative 5σ threshold yielded ∼7000 outliers, and the *purity* of this cut was 82 per cent. However, we point out that LAMOST observation strategy privileges relatively blue, bright objects in the Galaxy (Carlin et al. 2012). Therefore, an unconstrained cross-match with these data does not necessarily result in a homogeneous representation of the objects in the *Gaia* CMD.

This study is part of a pilot program to validate and test the CMD-based technique described in FR21, while simultaneously identifying and classifying new Hα-excess sources. To this end, a set of spectroscopic follow-up observations of the brightest 5σ Hα-excess candidates from FR21 is presented. The targets examined here are chosen in an attempt to homogeneously sample the *Gaia* CMD. In addition, we discuss in Appendix A how the distribution of our spectroscopically confirmed Hα-emitters across the *Gaia* CMD, supplemented by UVEX and Wide-field Infrared Survey Explorer

(WISE) parameter spaces, can be used to efficiently separate the different populations. This is essential to ease the task of target selection for large surveys such as the forthcoming WEAVE multi-object fibers facility (Dalton et al. 2012, 2018) or *4-m Multi-Object Spectroscopic Telescope* (*4MOST*; de Jong et al. 2022). With its ∼1000 fibers, its large integral field unit (IFU), and 20 mini IFUs, WEAVE will provide spectroscopic follow-up observations for *Gaia* and for the Low Frequency Array (Kassim et al. 2000) surveys, in the 370–1000 nm wavelength range, with intermediate ($R \sim 5000$) and high resolution ($R \sim 20\,000$). The *4MOST* survey will be able to simultaneously obtain spectra for ∼2400 objects, with a resolution range comparable to WEAVE.

The paper is organized as follows: Section 2 introduces the initial selection cuts employed to define our target list. *Gemini* spectroscopic observations and data reduction are also discussed. In Section 3, our source classification is presented. In Section 4, our spectroscopic classification is used to test the CMD selection by FR21. In Appendix A, additional photometric parameter spaces for stellar population discrimination purposes are explored. In Section 5, our conclusions are drawn.

2 OBSERVATIONS AND DATA REDUCTION

The ∼28 000 3σ Hα-excess candidates identified by FR21 are located in the Northern Galactic Plane, within the $|b| \leq 5°$ and $29° \leq l \leq 215°$ coordinate ranges, and inside a 1.5-kpc radius relative to the Sun.[4] The distances were calculated with the parallax-inversion method; the '$\pi/\sigma_\pi > 5$' constraint (with π representing the parallax, and σ_π the corresponding uncertainty) for parallax measurements minimized the difference between this estimator and the distances obtained with a probabilistic approach (Astraatmadja & Bailer-Jones 2016). The availability of *Gaia* DR2 parallaxes enabled FR21 to locate the objects in the CMD, and thus complement the classic position-based Hα-excess candidates selection with the novel CMD-based technique. The results described in FR21 can be updated with the use of more precise *Gaia* DR3 parallaxes. However, the average difference between parallaxes from *Gaia* DR2 and EDR3 is 0.06 mas, for all the targets inspected by FR21.

The Hα-excess candidates from FR21 are represented by the grey dots in Fig. 1. From their distribution in the *Gaia* CMD (top panel), they can be tentatively visually divided into five groups. The regions of this parameter space occupied by each group are labelled with capital letters. Region A is usually associated to bright and relatively blue Be/Ae stars; CVs mostly populate region B, i.e. the area between the zero-age main sequence (ZAMS) and the white dwarf (WD) tracks; the targets in region C are mainly pre-main-sequence (MS) objects; D area is a predominantly populated by reddened red giants (RGs), however, reddened Be stars and bright YSOs can also be found; and E region is commonly associated to faint M-dwarfs. However, a classification based on this criterion is indicative at best, as it is strongly affected by population mixing effects, and by the zero-extinction assumption.

2.1 Sample selection

Our spectroscopic data were obtained at the *Gemini* (North and South) telescopes, via a poor-weather queue. In order to reach a compromise between a minimum required signal-to-noise ratio of

---

[2]To simplify the notation, in this paper the adjective 'absolute' is implicit.

[3]In FR21, the authors refer to the *significance* just as σ. As σ is commonly associated to the standard deviation of a distribution, a different notation is preferred in this paper.

[4]However, we point out that this threshold does not constitute a hard cut, and the distance of some objects from the Sun exceeds 1.5 kpc.
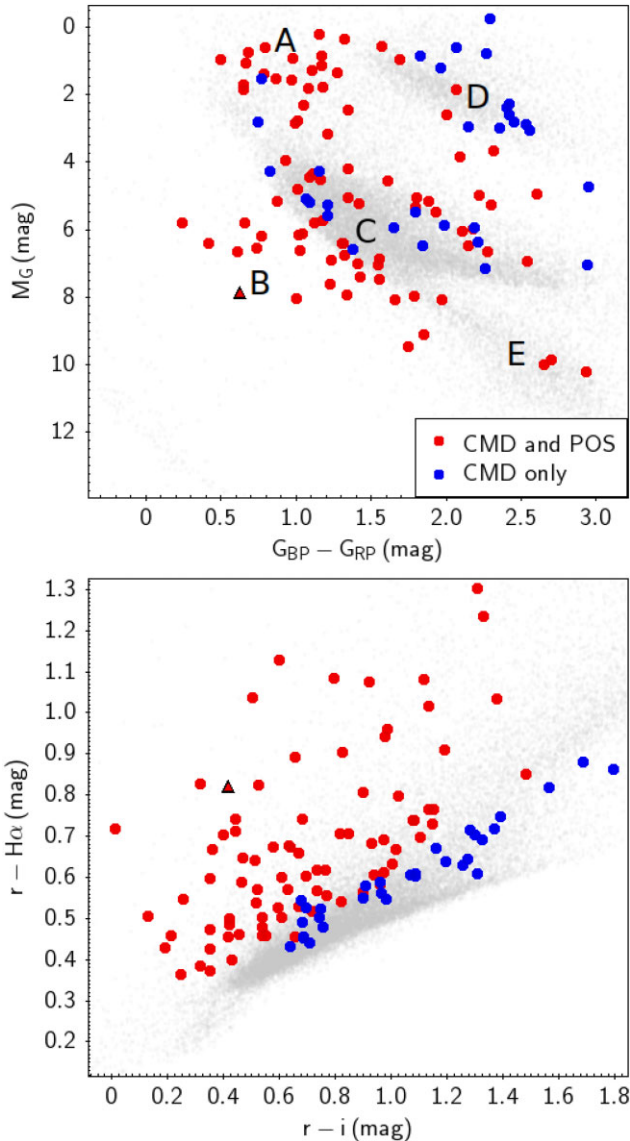
Figure 1. Locations in the *Gaia* CMD and in the IPHAS CCD of the 114 sources in our data set (top and bottom panels, respectively). While red colour refers to the Hα-excess candidates selected in FR21 from both partition types, the targets depicted in blue were identified from CMD-partitions only. No objects in our sample were identified in positional-partitions only. The grey dots plotted in the background as a reference represent all the 3σ outliers selected by FR21. The black-red triangle refers to the object *Gaia* DR2 259018688664493312, mentioned in the text.

10 (even for the faintest objects) and the necessary observing time, only objects with $r < 17.5$ mag were included in our follow-up. Only targets with a significance of being Hα-excess sources higher than or equal to five were considered. In an attempt to obtain a homogeneous sample across the tentative clusters A–E, our strategy has been to request follow-up observations of up to 2230 targets, sampled as evenly as possible across the *Gaia* CMD. Out of these 2230 spectra, 114 were obtained (67 from *Gemini North* and 47 from *Gemini South*). These are displayed as the red and blue dots in the *Gaia* CMD and in the IPHAS CCD (top and bottom panels, respectively). The red candidates were identified by both position-based and CMD-based outliers selection; however, the blue ones were detected only by the CMD-based selection. None of these outliers
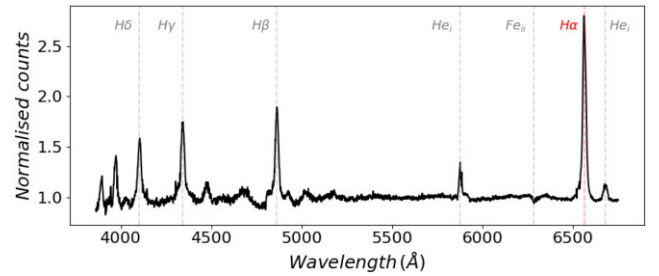

Figure 2. Example of reduced *Gemini* spectrum, belonging to the object *Gaia* DR2 259018688664493312. The most prominent emission/absorption lines are highlighted by dashed vertical lines.

resulted only from the position-based selection. As it can be seen in the top panel, the minimum-brightness constraint excludes, from our spectral follow-up, all the Hα-excess candidates that crowd the lowest stripe of the *Gaia* CMD. Among them we can list, for instance, all the objects that lie on (or around) the WD track. The bottom panel in the same figure brings out the value of the CMD-based selection technique. In fact, a non-negligible amount of outliers from FR21 (mainly blue dots) blends with the main stellar locus, in the IPHAS parameter space; without the prior population discrimination, these objects would have been hidden.

2.2 Gemini spectra

The spectra analysed in this work were acquired with the Gemini Multi-Object Spectrograph (Allington-Smith et al. 2002; Hook et al. 2004), with individual exposure times of 450 s. Gemini observatory consists of two 8.1-m telescopes, based in Mauna Kea (Hawaii; *Gemini North*) and in Cerro Pachon (Chile; *Gemini South*), respectively. The observations were carried out between 2019 August and the end of 2020 January. The use of the *R*400 grating yielded a spectral resolution at the blaze wavelength (764 nm) of $R \sim 1900$. The obtained wavelength resolution is 1 Å.

The set-up of the detector caused systematic gaps in the data. To eliminate this effect, two spectra with different central wavelengths (more specifically, 525 and 535 nm, respectively) were consecutively collected for each target.[5] These two spectra were then averaged, in order to associate one single spectrum to each source in our list. *Gemini* raw data were reduced with the Image Reduction and Analysis Facility Software System (Tody 1986) and the SAOImage DS9 tool (Smithsonian Astrophysical Observatory 2000). The process involved a cosmic rays subtraction, as well as bias and flat-field corrections. CuAr arcs were used during wavelength calibration step. Flux calibration could not be performed, because the poor-weather program did not include the observation of any flux standard objects. A spline function was instead fitted to the non-calibrated counts. During this procedure, the spectral lines were masked-out. The flux was then divided by the spline. Fig. 2 displays an example of a reduced spectrum. It belongs to the object *Gaia* DR2 259018688664493312, the position of which is highlighted with a red triangle in Fig. 1.

3 SOURCES CLASSIFICATION

While testing the selection technique presented in FR21, our spectral analysis also aims to break the degeneracy that emerges from

---

[5]For a total observation time of 900 s per target.

tentatively characterizing the sources on the basis of their location in the *Gaia* CMD. Our classification is made more solid with the support of previous designations found in SIMBAD (Wenger et al. 2000) data base (when available). Out of 114 targets, 32 are already labelled in SIMBAD. Our classifications confirm the ones found in said data base for almost half of the matches (15/32). In the remaining 17 cases, SIMBAD does not provide unambiguous classifications: However, 12 of these targets are classified as 'emission-line stars', four are labelled as 'variable stars', and one simply as 'star'.

The most common spectral features that distinguish the stellar populations mentioned in Section 2 are the following. Given the distribution of our targets in the *Gaia* CMD, MS stars other than classical Be stars and active M-dwarfs are also listed here:

(1) **MS stars (except for classical Be stars and active M-dwarfs)**: Their spectra are characterized by absorption in the main Balmer and He lines (the latter ones are shown mainly in the hottest objects of this category). Among them, γ-Doradus are typically A–F spectral types stars, characterized by line-profile and radial velocities variability (predominantly produced by g-mode oscillations; Kaye et al. 1999);

(2) **Coronally-active M-dwarfs**: Molecular (TiO and CaH) absorption is one of the most prominent spectral features of (mainly early type) M-dwarfs (Rajpurohit et al. 2018) due to the low atmosphere temperature. This population usually shows relatively narrow Hα (and sometimes Ca) emission lines due to coronal activity;

(3) **Be and Ae stars**: The spectra of both these object types are characterized by a blue continuum, with narrow Hα emission. Broad absorption lines might be found in correspondence to the other Balmer lines. Furthermore, Fe$_{II}$ and O$_I$ emission lines might be observed in the spectra of Be stars; however, Ae stars often show emission in some forbidden lines. One of the peculiarities of these classes of objects is the high variability of the strength of the emission lines (Porter & Rivinius 2003);

(4) **RGs**: as red giants, these objects are characterized by a red continuum emission. Narrow absorption is commonly observed in the Hα, He, and Fe$_{II}$ lines;C

(5) **YSOs**: Due to the diverse nature of the pre-MS objects, their spectra can show very different features. Nonetheless, they usually present an infrared (IR) excess, as well as the Hα line in emission. Additionally, TiO and VO molecular absorption bands can be often spotted from T-Tauri stars and from Orion variables (Itoh et al. 2010). Sources in this latter group might also present Paschen emission lines in their spectra (Hillenbrand, Hoffer & Herczeg 2013);

(6) **CVs**: Generally characterized by a blue continuum, broad Balmer – and often He – emission lines originate from the hot accretion disc (Sheets et al. 2007; Han et al. 2020).

Target classification can be further supported by the analysis of the full width at half-maximum (FWHM) of the Hα line. Henceforth, the acronym 'FWHM' will refer to the full width at half-maximum of the Hα line, unless otherwise specified. The width of a given line can provide an indication for the presence of an accretion disc (in compact accretors): Due to Doppler broadening, disc-fed accretors are characterized by relatively broad spectral lines. Out of the inspected spectra, 69/114 belong to confirmed Hα-emitters, and the FWHM is measured for 68 of them. The peculiar shape of the Hα line of one of our sources does not allow the measurement of its FWHM, as testified by Fig. 3. The distribution of our targets in the *Gaia* CMD, colour-coded with respect to their FWHM, is displayed in Fig. 4. The Hα-emitters (the coloured dots), are evenly spread on the CMD. The objects with the largest FWHM can be found between the ZAMS and the WD tracks. Said region of the *Gaia* CMD is
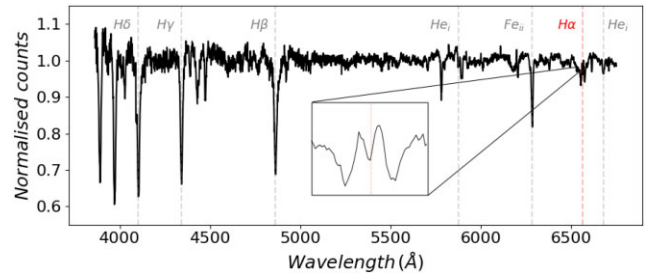

Figure 3. Spectrum of the object *Gaia* DR2 3455444577118544768. The shape in correspondence of the Hα wavelength does not allow the measurement of its FWHM.
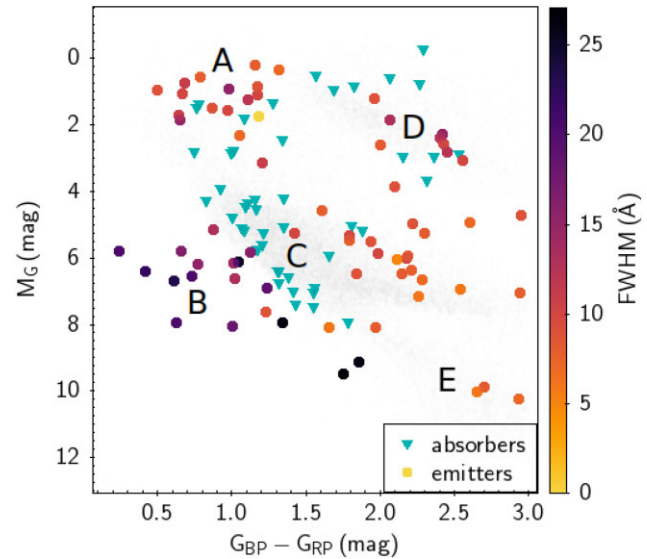

Figure 4. Positions of our targets in *Gaia* CMD, colour-coded according to the FWHM of the Hα line. The objects that show Hα absorption are depicted by the turquoise triangles. The grey dots represent all the 3σ outliers selected by FR21.

commonly associated to accreting WDs; both these source types are characterized by the presence of a disc surrounding the accreting object (Smak 1992; Szkody 1992; Beltrán 2020; Zhao et al. 2020). The remaining 45 objects, characterized by absorption in the Hα line, are depicted by the turquoise triangles. These objects are discussed later in the text.

As an outcome of our spectral classification, the examined data set is found to be a compound of: 35 MS stars (except for Be stars and M-dwarfs); 30 YSOs (while Marton et al. 2019 recently produced a catalogue of more than 1.1 million of YSO candidates, selected from *Gaia* DR2 and WISE surveys with machine learning techniques); 18 Be stars (whereas the all-sky Be Star Spectra data base described in Neiner et al. 2011 counts more than 2000 Be stars); 17 CVs (to be added to the 1429 CVs listed by Ritter & Kolb 2003); 10 RGs; and four active M-dwarfs (Lépine & Gaidos 2011 found ∼9000 bright active M-dwarfs from the SUPERBLINK survey. However, the authors estimate that their selection is only ∼75 per cent complete). Fig. 5 shows six exemplary spectra, one for each identified population. Our results are summarized in Table A1, in the appendix. The taxonomy of our sample is shown in Fig. 6, where the various populations are represented in the *Gaia* CMD with different colours and shapes. As can be noticed, targets labelled likewise generally tend to cluster in well-defined regions of the *Gaia*
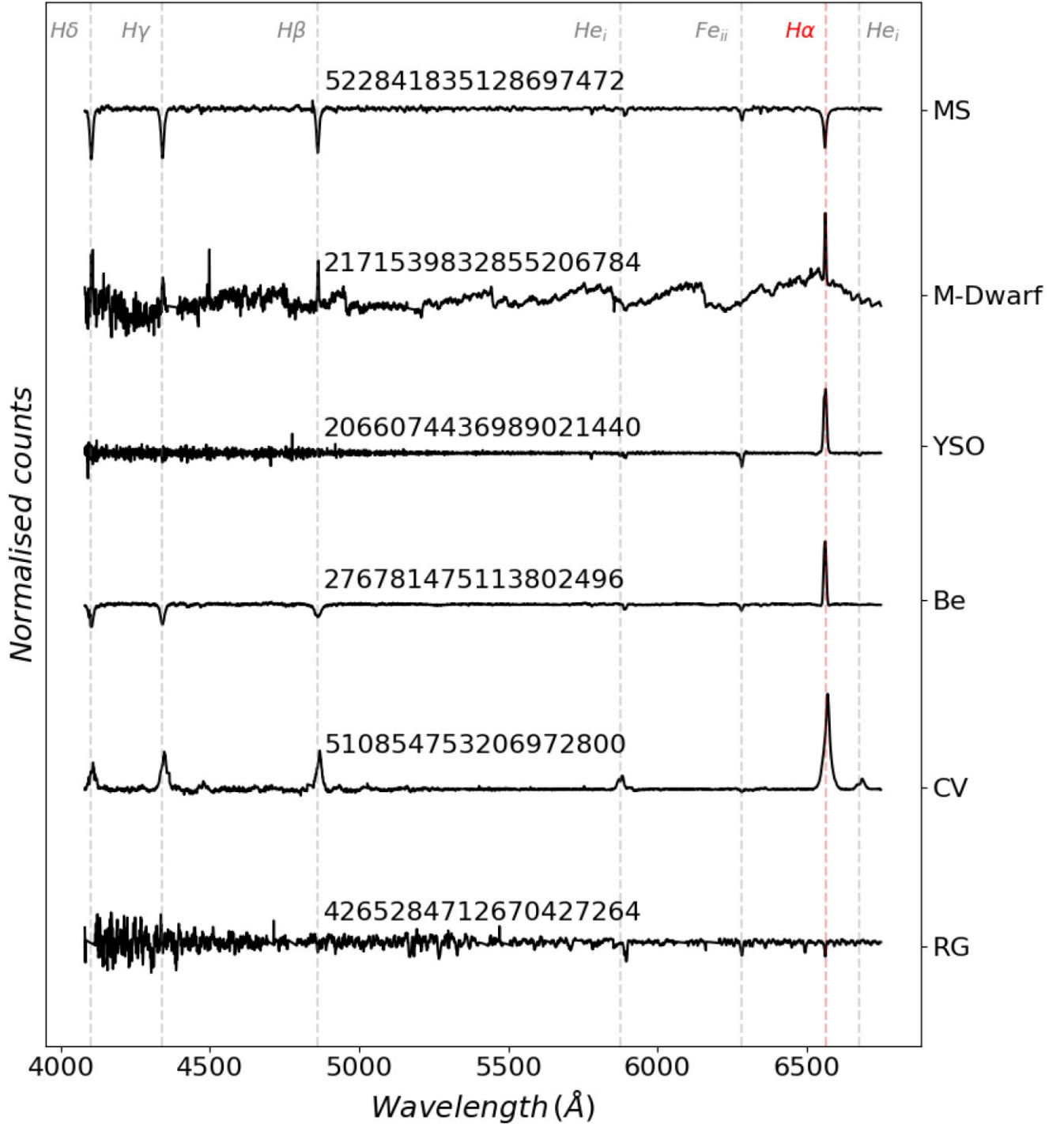
Figure 5. The spectra shown here (flux-normalized for visualization purposes) exemplify the six categories to which the objects in our sample are assigned. The labels on the right vertical axis associate each spectrum to its corresponding population. The respective targets are identified by their (*Gaia* DR2) Source ID. Some Balmer, He, and Fe lines are highlighted with dotted lines.

CMD – the only noticeable exception to this statement is provided by YSOs, which, for their manifold nature, are more widely spread in this parameter space. However, the overlaps of these regions clearly demonstrate that the location in the *Gaia* CMD does not provide enough information to unambiguously tie a source to a stellar population. For instance, a particularly dominant companion would make it an extremely difficult task to discern a CV from a MS star; moreover, the bright and red tail of the YSO distribution seems to blend with the RGs; furthermore, distinguishing line emitting Be stars from common, bright, and MS stars can be very tricky.

4 TESTING THE CMD-BASED IDENTIFICATION

The distribution of our sample in the *Gaia* CMD (top panel) and in the IPHAS CCD (bottom panel) is shown in Fig. 7. The colour-code is the
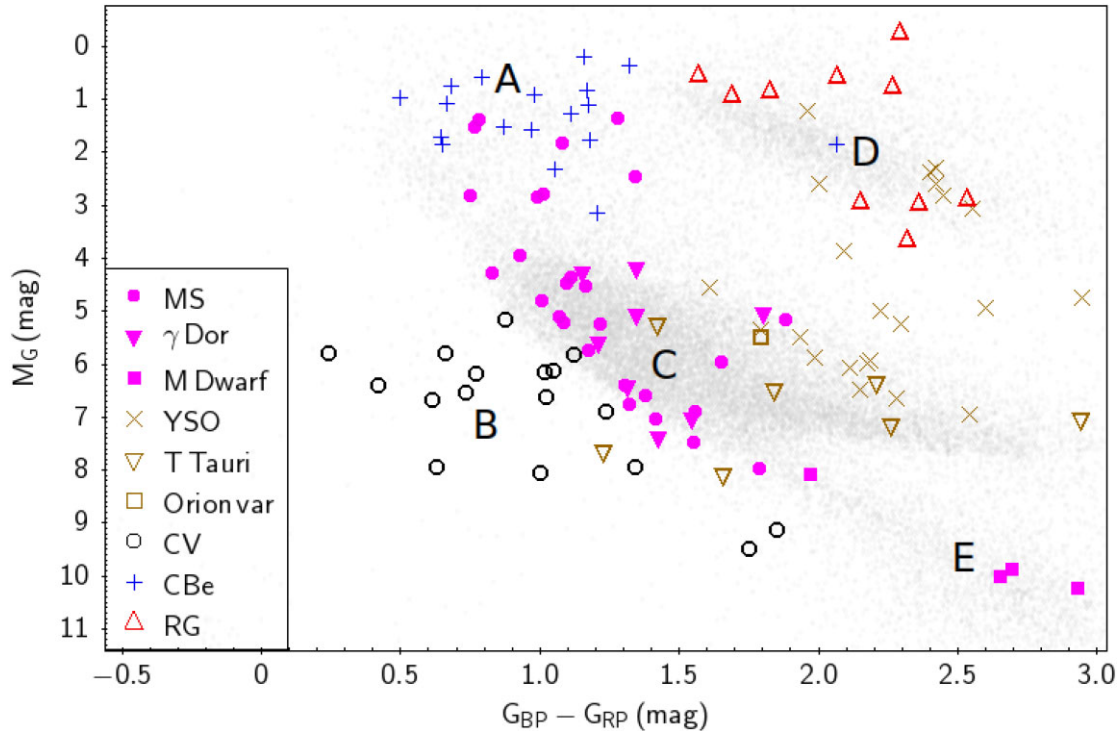
Figure 6. Location of our sources in the *Gaia* CMD, colour-coded with respect to their spectral classification. The grey dots represent the Hα-excess candidates selected in FR21.

same applied in Fig. 1. Here, the dots represent the spectroscopically confirmed Hα-emitters, while the triangles display the absorbers. Out of the 69 emitters, 55 were selected by both the position-based and CMD-based techniques. The remaining 14 emitters were obtained only through the CMD-based method (no outliers in our sample were identified by the position-based selection only), demonstrating the improvements introduced by this selection in terms of completeness. This is further shown by the fact that the CMD-significance is higher than the position-significance for 45/55 of the common emitters. All the 14 CMD-emitters are spectroscopically classified as YSOs. This confirms that the CMD-based method is more efficient than the position-based one in identifying emission-line sources within this population (as also stated in section 5 of FR21).

Out of the 45 absorbers, 28 (62 per cent) were selected by both position-based and CMD-based selections. The remaining 17 absorbers were identified only by the CMD-based method. The 45 absorbers include 35 MS stars (other than classical Be stars and M-dwarfs) and all the 10 RGs. The variability time-scale typical of these stellar populations is significantly longer than the ∼10 yr that separate IPHAS and LAMOST observations. For this reason, the photometric identification of these absorbers as Hα-excess candidates cannot be justified solely on the basis of temporal variability. However, it is relevant to stress that the algorithm presented in FR21 does not aim at identifying Hα-*emitters*, but rather objects that present an *excess* in the Hα flux, with reference to nearby targets in the *Gaia* CMD. Fig. 17 in FR21 shows two examples to visually explain this phenomenon. Out of the 10 RGs in our sample, seven were selected only by the CMD-based method. To test the reliability of this selection criterion in region D of the CMD, a group of outliers that lie in that area is visually isolated. Considering that the vast majority of them was identified only by the CMD-based method, the purity yielded by this selection criterion in region D of the *Gaia* CMD seems to be significantly lower than the one from the position-based

selection. However, half of the YSOs identified only by the CMD-based method also lie in that area. The position-based selection alone would have missed these outliers. A similar test is performed for the MS stars population: The majority of the visually selected outliers along the ZAMS were identified only by the CMD-based technique. However, 25/35 of the MS stars in our sample were identified by both selections. Therefore, the position-based one does not seem to provide advantages in terms of purity, along the ZAMS. On the other hand, one YSO that lies in the region was selected only by the CMD-based technique.

Table 1 summarizes how many members of each population in our sample were identified as Hα-excess candidates by the CMD-based selection and how many by the position-based one. The number of spectroscopically identified Hα-absorbers in each population is provided within the parentheses.

5 CONCLUSIONS

FR21 developed a method to identify Hα-excess candidates with reference to nearby sources in the *Gaia* CMD. In this study, part of a pilot-program to validate this CMD-based selection technique is presented. More specifically, its performances in terms of *completeness* are compared to those of classical position-based methods. Our analysis involves the spectral observation and classification of 114 bright ($r \leq 17.5$ mag) Hα-excess candidates. These are drawn from a group of 2230 5σ outliers that homogeneously sample the *Gaia* CMD. Their spectra were obtained by the two *Gemini* telescopes. Table 1 shows how many Hα-excess candidates were identified by each selection criterion, on a population-by-population basis. The amount of Hα-absorbers in each group of objects is specified within parentheses. Since the spectra that present the Hα line in absorption belong to either RGs or MS stars (other than classical Be stars or active M-dwarfs), Hα emission variability seems to be discarded as
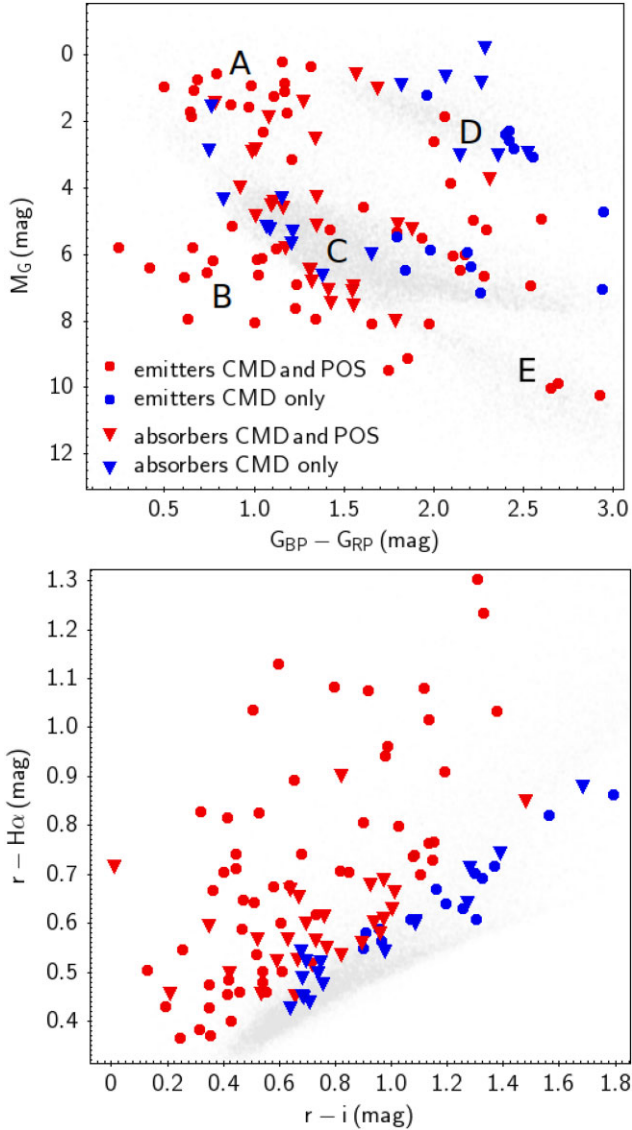
Figure 7. Locations in the *Gaia* CMD and in the IPHAS CCD of the 114 sources in our data set (top and bottom panels, respectively). The colour-code is the same of Fig. 1. While the dots display the Hα line in emission, the triangles show it in absorption.

Table 1. Number of members from each population in our sample detected by the CMD and/or position-based selections, respectively. The amount of absorbers from each population is included inside the parentheses.

Population	CMD-based selection (abs)	Position-based selection (abs)
YSO	30 (0)	16 (0)
Classical Be	18 (0)	18 (0)
CV	17 (0)	17 (0)
RG	10 (10)	3 (3)
Active M-dwarf	4 (0)	4 (0)
Other MS star	35 (35)	25 (25)
Total	114 (45)	83 (28)

a possible explanation for their detection as outliers. Supplementary analyses to justify the selection of these objects as Hα-excess sources are required.

We further show that the positions of our spectroscopically classified objects in the *Gaia* CMD, UVEX CMD, and WISE CCD can be used to better discern the populations of Hα emission-line sources. A pre-selection based on this criterion can be used to improve the identification and classification of emission-lines sources in large data bases. This is a particularly relevant task, specially with the upcoming arrival of large spectroscopic surveys such as WEAVE and *4MOST*.

DATA AVAILABILITY

The 114 analysed *Gemini* spectra can be found in VizieR as 'Spectroscopic follow-up of Hα-excess sources in the Northern Galactic Plane'.

REFERENCES

Allington-Smith J. et al., 2002, PASP, 114, 892
Astraatmadja T. L., Bailer-Jones C. A. L., 2016, ApJ, 833, 119
Beltrán M., 2020, preprint (arXiv:2005.06912)
Breedt E., Gänsicke B. T., Marsh T. R., Steeghs D., Drake A. J., Copperwheat C. M., 2012, MNRAS, 425, 2548
Carlin J. L. et al., 2012, Res. Astron. Astrophys., 12, 755
Cui X.-Q. et al., 2012, Res. Astron. Astrophys., 12, 1197
Dalton G. et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, Proc. SPIE Conf. Ser. Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV. SPIE, Bellingham, p. 84460P

Dalton G. et al., 2018, in Evans C. J., Simard L., Takami H., eds, Proc. SPIE Conf. Ser. Vol. 10702, Ground-based and Airborne Instrumentation for Astronomy VII. SPIE, Bellingham, p. 107021B

de Jong R. S. et al., 2022, Proc. SPIE 12184, Ground-based and Airborne Instrumentation for Astronomy IX. SPIE, Bellingham

Drew J. E. et al., 2005, MNRAS, 362, 753

Fratta M. et al., 2021, MNRAS, 505, 1135(FR21)

Groot P. J. et al., 2009, MNRAS, 399, 323

Han Z., Boonrucksar S., Qian S., Xiaohui F., Wang Q., Zhu L., Dong A., Zhi Q., 2020, PASJ, 72, 76

Hillenbrand L. A., Hoffer A. S., Herczeg G. J., 2013, AJ, 146, 85

Hook I. M., Jørgensen I., Allington-Smith J. R., Davies R. L., Metcalfe N., Murowinski R. G., Crampton D., 2004, PASP, 116, 425

Itoh Y., Gupta R., Oasa Y., Sen A. K., Tanaka M., Terai T., Nakaoka S., 2010, PASJ, 62, 1149

Kassim N. E., Lazio T. J. W., Erickson W. C., Crane P. C., Perley R. A., Hicks B., 2000, in Butcher H. R., ed., Proc. SPIE Conf. Ser. Vol. 4015, Radio Telescopes. SPIE, Bellingham, p. 328

Kaye A. B., Handler G., Krisciunas K., Poretti E., Zerbi F. M., 1999, PASP, 111, 840

Lépine S., Gaidos E., 2011, AJ, 142, 138

Marton G. et al., 2019, MNRAS, 487, 2522

Monguió M. et al., 2020, A&A, 638, A18

Neiner C., de Batz B., Cochard F., Floquet M., Mekkas A., Desnoux V., 2011, AJ, 142, 149

Neugebauer G. et al., 1984, ApJ, 278, L1

Pala A. F. et al., 2020, MNRAS, 494, 3799

Podsiadlowski P., Han Z., Rappaport S., 2003, MNRAS, 340, 1214

Porter J. M., Rivinius T., 2003, PASP, 115, 1153

Rajpurohit A. S., Allard F., Rajpurohit S., Sharma R., Teixeira G. D. C., Mousis O., Rajpurohit K., 2018, A&A, 620, A180

Ritter H., Kolb U., 2003, A&A, 404, 301

Sheets H. A., Thorstensen J. R., Peters C. J., Kapusta A. B., Taylor C. J., 2007, PASP, 119, 494

Smak J., 1992, Acta Astron., 42, 323

Smith B. J., Price S. D., Baker R. I., 2004, ApJS, 154, 673

Smithsonian Astrophysical Observatory, 2000, Astrophysics Source Code Library, record ascl:0003.002

Szkody P., 1992, in Vogt N., ed., ASP Conf. Ser. Vol. 29, Cataclysmic Variable Stars. Astron. Soc. Pac., San Francisco, p. 42

Tody D., 1986, in Crawford D. L., ed., Proc. SPIE Conf. Ser. Vol. 627, Instrumentation in Astronomy VI. SPIE, Bellingham, p. 733

Wenger M. et al., 2000, A&AS, 143, 9

Witham A. R., Knigge C., Drew J. E., Greimel R., Steeghs D., Gänsicke B. T., Groot P. J., Mampaso A., 2008, MNRAS, 384, 1277

Wright E. L. et al., 2010, AJ, 140, 1868

Zhao B. et al., 2020, Space Sci. Rev., 216, 43

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

spectra

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX: PHOTOMETRY-BASED PRE-SELECTION

The availability of large spectroscopic surveys makes the exploration of different pre-selection methods a crucial task. The last paragraph of Section 3 shows how the distribution of our sample in the *Gaia* CMD can be used as indication to classify Hα-excess sources. This

study shows how this information can be augmented with additional diagnostics, to ease a spectroscopic classification. The physical and chemical properties of diverse source types can affect wider different pass-bands out of the *Gaia* ones. Additional information from bluer/redder pass-bands can thus better inform on the exact class.

The optical photometry from *Gaia* is complemented here with ultraviolet and IR intensities. These are leveraged from UVEX and WISE surveys, respectively. The UVEX started in 2006, and it is based on the island of La Palma (Spain). With the same set-up as IPHAS, the Wide Field Camera mounted on the *Isaac Newton Telescope* provides U,[6] g, r, and He_{I5875} magnitudes for targets in $|b| \leq 5°$ (Groot et al. 2009; Monguió et al. 2020), up to 22 mag. Its main task consists in the identification of new Galactic stellar remnants. On the other hand, the WISE is an all-sky survey funded by NASA that was launched in 2009 December. In 6 months, it scrutinizes the IR properties of the entire sky in four bands ($W1$, $W2$, $W3$, and $W4$), centred at 3.4, 4.6, 12, and 22 μm (Wright et al. 2010). With its 4 million pixels, this survey achieves a remarkable sensitivity: more than 1000 times better than InfraRed Astronomical Satellite (Neugebauer et al. 1984) in the 12-μm band, and 50 000 times better than Diffuse InfraRed Background Experiment (Smith, Price & Baker 2004) in the two mid-IR bands. Thanks to its features, WISE is particularly efficient in detecting (even old, faint) brown dwarfs, ultra-luminous infrared galaxies, active galactic nuclei, and quasi-stellar objects.

Out of the 114 Hα-excess candidates in our dataset, 78 have an available measurement in UVEX U, g, and r bands, whilst 73 have associated $W1$, $W2$, and $W3$ WISE measurements.[7] A cross-match between these two groups yields 57 matches, i.e. half of our data set. The non-detection of a source carries valuable information for classification purposes, per se. In fact, objects that appear to be too faint (or too bright), in relation to the limiting magnitudes of the survey, are not detected. Among the 36 objects in our sample that are not included in UVEX survey, 27 are MS stars[8] (one of them being a Be star), and eight are RGs. These two populations are usually faint in UVEX bands, as confirmed by the fact that the MS stars and RGs included in UVEX are among the faintest objects in our sample, in the U band. The faintest target in our sample sets a lower limit of $U_{MS/RG} > 21.3$ mag on the apparent magnitude of the remaining MS stars and RGs in our sample. Similarly, 23 MS stars (including three Be stars and one M-dwarf) and 13 CVs have no detections in WISE bands. The sources in these categories with also a WISE counterpart are among the faintest in all WISE bands of interest. The $W1_{MS/CV} > 14.8$ mag, $W2_{MS/CV} > 14.8$ mag, and $W3_{MS/CV} > 12.7$ mag constraints for the remaining MS stars and CVs in our sample are set by the faintest objects with WISE counterparts.

Fig. A1 shows the locations in the M_U versus g–r CMD[9] (top panel) and in the $W2$–$W3$ versus $W1$–$W2$ CCD (bottom panel) of all the objects with valid measurements in the corresponding bands. Limiting the analysis to the Hα-emitting populations, a visual inspection of the three parameter spaces combined shows that Be stars and YSOs seem to be better separate in UVEX CMD than in the *Gaia* CMD. CVs are clearly bluer compared to the ZAMS in

---

[6]We point out that U band measurements in the IGAPS catalogue are not uniformly calibrated.

[7]We point out that the large sizes of WISE pixels might produce some spurious cross-matches, specially in crowded fields such as the one being studied here.

[8]However, it is worth noticing that all the four M-dwarfs in our sample are included in UVEX catalogue.

[9]The zero-extinction assumption is also applied to obtain the absolute magnitude in the UVEX U band (M_U).

Table A1. Results of our spectroscopic classification. For each target, the following entries are provided: *Gaia* DR2 Source ID (for almost all the sources in our data set, the Source ID did not change between DR2 and EDR3. The only exception is *Gaia* DR2 4306883326285609728, that became *Gaia* EDR3 4306883326286990464); the *Gaia* DR2 Barycentric Right Ascension and Declination, at epoch 2015.5; the distance, obtained inverting *Gaia* DR2 parallax; the apparent magnitude in IPHAS *r* band; the maximum significance (between the two provided by FR21) of being an Hα-excess source; the FWHM of the Hα line ('NA', if said line is seen in absorption); and our spectral classification.

Source ID (*Gaia* DR2)	RA (°)	Dec (°)	Distance (kpc)	r (mag)	$n_{\sigma_{max}[CMD,POS]}$	FWHM (Å)	Class
183806256466977920	79.54732	35.13356	0.91(5)	16.35	13.56	19(1)	CV
184238295817134208	82.51049	36.82756	0.62(2)	14.69	6.42	NA	MS
185881786526045440	74.66526	35.09533	1.2(1)	16.77	8.49	14.2(6)	CV
188400875040872832	76.02748	38.72527	2.8(3)	13.66	5.76	9.14(7)	Be
202583509889708800	77.95731	44.28475	0.58(3)	16.62	12.50	NA	MS
207006089252554240	75.83720	47.08507	1.3(2)	17.22	9.14	18.3(2)	CV
250899658585804544	59.41636	51.61552	0.506(7)	13.60	9.38	NA	MS
251693368542860288	59.30778	51.99074	1.19(6)	15.61	11.32	10.15(8)	YSO
258089193320478208	67.22134	48.52566	1.9(1)	15.44	7.74	11.17(4)	Be
259018688664493312	71.64038	48.96539	0.56(2)	16.56	15.95	20.30(1)	CV
276781475113802496	62.89925	55.87940	4.7(7)	14.82	7.7	10.15(7)	Be
422750061837839232	0.93751	58.46850	1.0(1)	17.29	6.67	NA	MS
424396924098269312	13.42935	57.96660	4.4(4)	13.86	10.47	NA	RG
425129194536945024	10.71283	58.32971	0.97(4)	16.51	14.89	NA	MS
425995304767655040	13.37703	59.19563	3.6(3)	13.83	13.05	NA	RG
427292453608805504	13.32775	60.77675	0.82(1)	14.02	12.10	NA	MS
427316058750043520	13.55209	60.94371	3.5(3)	13.73	7.32	NA	RG
427473838665897984	13.38334	61.84135	0.339(3)	14.24	7.17	NA	MS
428233081505186688	5.22420	59.16597	0.98(5)	16.91	6.30	NA	MS
463122170299415168	45.27165	60.53203	2.4(3)	15.84	10.45	9.14(1)	YSO
464931073088436096	41.95254	60.96403	2.3(1)	13.95	14.50	7.11(3)	Be
465712207381733888	41.97116	61.56827	6(1)	14.37	9.23	7.11(3)	Be
510854753206972800	22.63292	62.35899	0.79(5)	16.92	11.45	18.3(2)	CV
512016318518918400	26.73920	63.00894	6(1)	15.64	16.33	13.20(4)	Be
514896283072146560	30.58104	63.37740	0.277(7)	16.78	8.10	8.1(1)	M-dwarf
522539950472429312	17.01763	61.66252	0.478(4)	13.63	6.29	NA	MS
522542802330612608	16.68316	61.54304	6(1)	13.95	7.01	NA	RG
522841835128697472	16.42042	62.10566	3.0(2)	13.94	6.43	NA	MS
524238661574474624	13.60656	64.62303	2.10(7)	13.73	10.66	NA	MS
524301986569104768	14.93122	64.91167	0.76(4)	17.26	15.22	25.4(6)	CV
526118620296454528	13.55275	66.29011	0.657(8)	13.92	7.58	NA	MS
527026954341506688	8.95623	64.25796	3.7(6)	16.44	15.88	NA	RG
1823343535657240064	298.95041	20.02398	5.1(8)	14.30	7.58	NA	RG
1833363385113024896	300.48773	23.60346	1.8(2)	17.04	7.69	17.1(4)	CV
1970080784065021184	317.62957	43.13736	1.8(2)	16.69	7.79	NA	MS
1971264992446842368	321.13415	45.06151	1.2(1)	17.02	11.38	25.4(5)	CV
1971760631674787072	318.70753	44.74772	2.3(4)	16.72	7.63	14.2(4)	CV
1978346259653262976	322.95547	48.06728	0.80(3)	16.0	10.81	8.12(5)	YSO
1980077092801857664	331.73643	51.09846	1.09(9)	16.68	22.70	23.3(2)	CV
1997162958039726080	348.79460	55.95792	2.5(3)	15.79	8.97	NA	MS
2007611926287679616	337.32027	57.08722	4.4(4)	13.90	9.17	11.17(4)	Be
2015167735817519360	345.94170	61.81282	0.75(2)	14.67	13.17	8.12(3)	YSO
2017612358184192768	359.83326	66.38672	0.79(2)	14.63	15.15	7.11(1)	YSO
2018379134410850688	294.93998	22.08596	0.63(2)	16.64	6.96	8.0(1)	T-Tauri
2021418317687630080	293.75924	24.77452	0.417(9)	16.22	9.75	6.0(1)	M-dwarf
2022704849371681024	290.84948	23.73654	0.53(1)	15.56	6.63	NA	Gamma Dor
2024391951234118912	292.20999	24.98576	2.1(3)	16.75	8.75	NA	Gamma Dor
2030374978109714688	298.23984	30.08825	1.42(5)	15.41	12.01	NA	MS
2033218590053883648	294.44810	31.67799	5(1)	16.04	6.24	NA	MS
2034706057485253376	297.86111	32.94033	1.03(7)	17.20	6.80	NA	MS
2034718873668327680	297.07516	32.64354	6(1)	14.83	5.84	NA	RG
2046886550363859968	295.31150	32.87483	3.8(5)	16.08	8.31	NA	MS
2054712221272378496	304.89408	33.57262	1.19(7)	16.41	5.72	NA	MS
2061420410434776192	305.63634	39.28241	0.79(1)	13.81	11.67	NA	MS
2066074436989021440	311.37173	41.05052	3.0(3)	15.43	7.12	12.18(6)	YSO
2071104221657306496	309.20076	45.34096	0.93(3)	15.94	13.20	5.08(3)	YSO
2162336164614235648	316.96769	44.09499	0.475(5)	13.94	10.86	21.32(6)	CV

Table A1 – *continued*

Source ID (Gaia DR2)	RA (°)	Dec (°)	Distance (kpc)	r (mag)	$n_{\sigma_{max[CMD,POS]}}$	FWHM (Å)	Class
2162850736074155392	316.39512	46.37367	3.3(4)	15.62	7.37	NA	MS
2162947798009095808	312.86245	44.22094	0.79(4)	16.93	20.98	6.09(5)	YSO
2164154516334658560	318.81208	46.45168	3.9(3)	14.61	8.45	NA	MS
2165559073718735872	316.57904	48.68376	3.3(5)	15.92	13.61	NA	RG
2166302068698392960	311.57080	45.42623	3.7(6)	16.14	9.97	10.15(6)	YSO
2168759236665389312	315.91445	50.26468	0.59(1)	15.69	10.23	7.11(6)	YSO
2170925927411884672	323.05921	49.74432	1.6(2)	17.02	13.46	16.2(2)	CV
2171539832855206784	324.25968	51.53298	0.208(3)	16.85	6.96	7.1(1)	M-dwarf
2201164799270238976	334.85416	59.55555	3.4(4)	15.07	16.14	8.12(1)	YSO
2207626251145725440	345.36781	62.93661	0.89(2)	14.79	6.35	9.14(3)	YSO
2209703675287141504	356.09811	66.02327	1.40(8)	15.89	8.65	NA	MS
2210266522162289536	350.61836	66.02533	0.74(1)	14.77	21.19	9.14(2)	YSO
3105714495538124288	101.47439	− 2.90572	6(1)	14.87	12.27	12.05(3)	Be
3113703547027886720	104.11159	0.38287	1.4(2)	17.31	11.52	NA	MS
3119336516892417024	99.76717	− 0.80271	1.9(3)	16.07	9.99	8.03(3)	YSO
3125759898182046080	102.12067	1.51568	3.7(4)	13.84	7.67	10.04(2)	Be
3126136064298079616	102.23128	1.59818	4.8(8)	13.9	13.62	8.03(2)	Be
3126247218051915264	103.83627	2.31710	0.72(6)	16.52	6.70	NA	Gamma Dor
3126354699614424064	102.17797	2.06890	1.3(2)	16.77	6.91	NA	Gamma Dor
3129126499706430976	103.45549	4.99618	4.5(7)	14.06	4.97	9.03(2)	Be
3130722814493379968	100.68126	4.59081	0.275(9)	16.75	9.55	6.0(1)	M-dwarf
3132911705924538112	102.16806	6.64056	3.9(7)	14.76	8.94	15.06(6)	Be
3324947321590659456	94.65241	6.94508	0.65(3)	16.38	23.69	9.03(3)	T-Tauri
3326685615112100992	100.31993	9.45838	0.73(2)	14.38	33.26	10.04(3)	T-Tauri
3326734470363384320	100.68635	9.98419	0.76(6)	16.87	5.95	7.0(3)	T-Tauri
3326896854488117632	99.87232	9.72769	0.73(3)	15.92	6.40	7.0(1)	T-Tauri
3326904585429362560	100.10781	9.84931	0.73(6)	16.20	17.18	6.02(3)	T-Tauri
3326938262268528128	100.43208	10.13123	0.70(3)	14.68	6.46	8.0(2)	Orion Var
3349903246243514752	88.17739	17.18745	0.77(4)	15.64	6.07	8.03(4)	T-Tauri
3350768218295646336	100.70632	10.18121	0.469(9)	14.01	5.80	NA	Gamma Dor
3355257352475460480	98.03440	12.48440	6(1)	13.99	8.46	8.03(1)	Be
3356132937390349824	99.12068	14.54367	1.2(2)	16.09	7.86	27.1(3)	CV
3369781587542465280	96.72360	17.79049	5.1(8)	14.37	7.41	9.04(3)	Be
3372309639651922432	97.35764	19.57940	3.0(3)	15.33	7.34	NA	MS
3376955492893794048	93.22946	22.17039	0.78(4)	15.93	8.35	23.1(2)	CV
3424187076448918272	88.42817	21.87315	3.5(5)	14.41	8.30	12.05(5)	Be
3444277112392781440	86.47305	29.61278	4.0(4)	13.95	8.16	15.23(3)	Be
3448388529968087040	85.47474	32.60364	4.7(8)	14.43	9.99	10.15(5)	Be
3451057044687431296	89.29311	32.65179	2.7(3)	15.30	26.23	NA	MS
3455444577118544768	85.92677	34.93837	4.2(6)	14.75	6.41	Inv.[a]	Be
4259907860230511616	281.70494	− 1.61207	3.6(7)	14.06	6.52	9.03(3)	YSO
4260073676017333248	281.49998	− 1.21673	3.0(5)	15.12	8.02	11.04(3)	YSO
4261708203134890624	286.48896	− 2.63683	1.0(1)	16.34	9.33	19.1(4)	CV
4261769672710353920	285.84803	− 2.29662	0.227(3)	16.48	14.9	28.1(2)	CV
4265284712670427264	282.72070	− 1.15056	3.3(7)	16.78	13.38	NA	RG
4273688795363205760	277.83036	1.11545	0.452(7)	14.61	7.25	8.03(3)	YSO
4278629893245072384	282.28770	1.42557	0.58(1)	14.66	6.41	10.04(7)	YSO
4281419427268084480	286.68511	4.66668	0.71(3)	15.40	9.3	10.04(7)	YSO
4282917305693991936	284.19650	7.01528	0.40(2)	17.12	6.32	26.1(2)	CV
4283915662276175104	279.88333	4.99783	3.1(5)	15.65	10.57	NA	RG
4305857275780220032	286.25677	6.40762	2.3(4)	16.94	5.98	9.04(6)	YSO
4306883326285609728	284.75531	7.12050	1.12(6)	14.66	10.58	NA	Gamma Dor
4309797032104899328	288.96905	11.20355	1.9(1)	13.88	13.53	12.05(6)	YSO
4311951731334685440	283.65170	10.57241	0.98(3)	14.32	7.16	NA	MS
4311967365016126976	283.85071	10.91176	1.8(1)	15.52	6.72	NA	Gamma Dor
4316059300586640256	291.39037	12.68881	3.4(6)	15.05	6.23	16.06(9)	YSO
4514006371833731456	286.21437	17.35407	1.8(2)	16.62	9.68	NA	Gamma Dor

Note. [a]The spectrum of this object presents Hα-inversion, as shown in Fig. 3.

UVEX CMD than in the *Gaia* CMD; however, M-dwarfs appear to be equally well identifiable in UVEX CMD or the *Gaia* CMD. Future studies may be able to further leverage multiband photometry to quantify the class separations of Hα-emitting sources. This may be particularly useful to optimize target selection in large spectroscopic surveys.
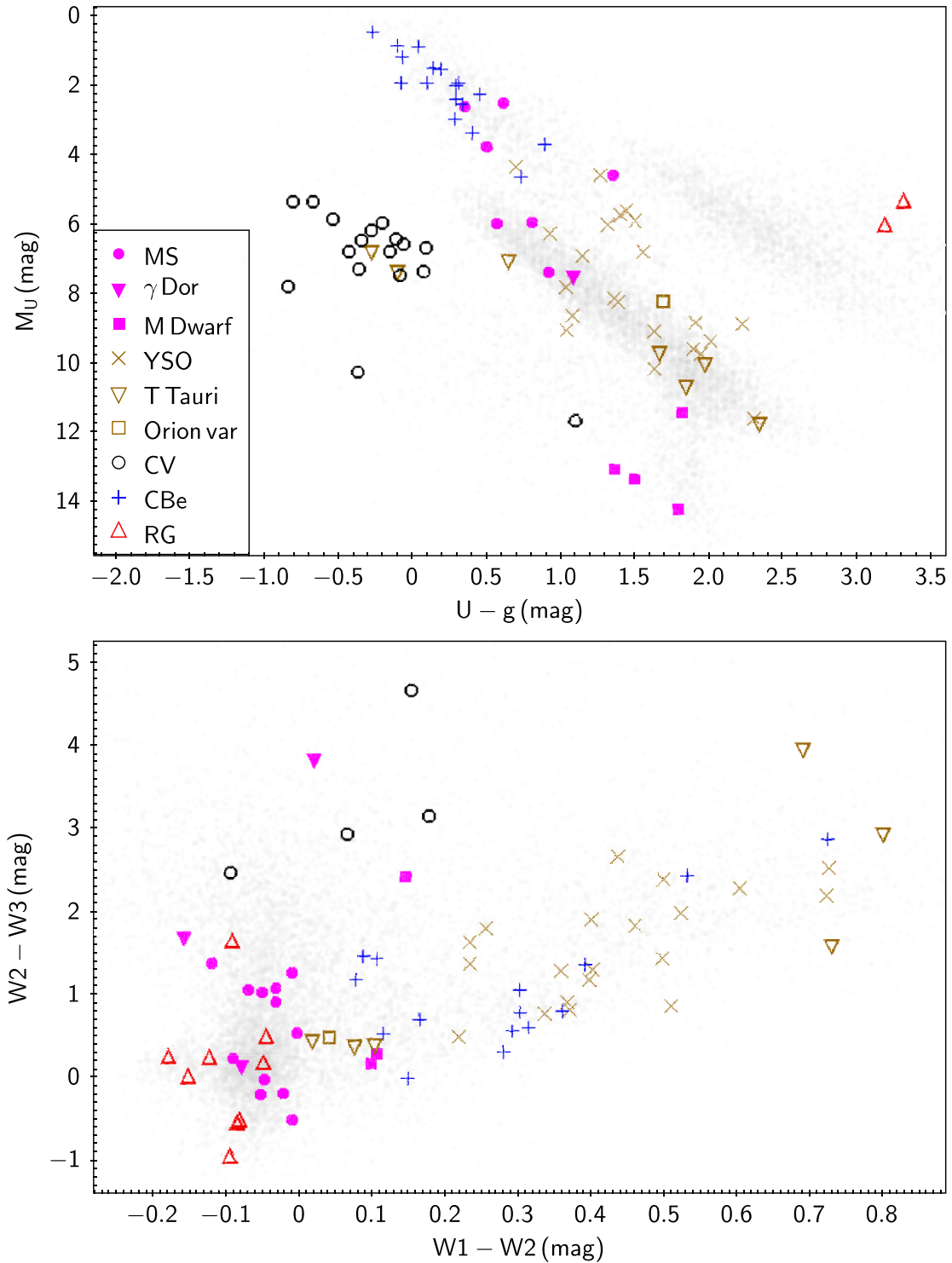
Figure A1. Location in UVEX CMD and WISE CCD of our subsamples, with available photometric measurements in the corresponding bands. The colour (and shape)-code applied here is the same of Fig. 6. The grey dots in the background depict the Hα-excess candidates found by FR21, included in UVEX or in WISE surveys.

This paper has been typeset from a TEX/LATEX file prepared by the author.