Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **465,** 3011–3038 (2017)
Advance Access publication 2016 November 2

doi:10.1093/mnras/stw2801

A population of eruptive variable protostars in VVV

C. Contreras Peña,[1,2,3★] P. W. Lucas,[2★] D. Minniti,[1,4] R. Kurtev,[5,3] W. Stimson,[2] C. Navarro Molina,[3,5] J. Borissova,[5,3] M. S. N. Kumar,[2,6] M. A. Thompson,[2] T. Gledhill,[2] R. Terzi,[2] D. Froebrich[7] and A. Caratti o Garatti[8]

[1]*Departamento de Ciencias Fisicas, Universidad Andres Bello, Republica 220, Santiago, Chile*
[2]*Centre for Astrophysics Research, University of Hertfordshire, Hatfield AL10 9AB, UK*
[3]*Millennium Institute of Astrophysics, Av. Vicuna Mackenna 4860, 782-0436, Macul, Santiago, Chile*
[4]*Vatican Observatory, V-00120 Vatican City State, Italy*
[5]*Instituto de Física y Astronomía, Universidad de Valparaíso, ave. Gran Bretaña, 1111, Casilla 5030, Valparaíso, Chile*
[6]*Instituto de Astrofisica e Ciencias do Espaco, Universidade do Porto, CAUP, Rua das Estrelas, P-4150-762 Porto, Portugal*
[7]*Centre for Astrophysics and Planetary Science, University of Kent, Canterbury CT2 7NH, UK*
[8]*Dublin Institute for Advanced Studies, School of Cosmic Physics, Astronomy and Astrophysics Section, 31 Fitzwilliam Place, Dublin 2, Ireland*

Accepted 2016 October 28. Received 2016 October 28; in original form 2016 February 16

ABSTRACT

We present the discovery of 816 high-amplitude infrared variable stars ($\Delta K_s > 1$ mag) in 119 deg^2 of the Galactic mid-plane covered by the VISTA Variables in the Via Lactea (VVV) survey. Almost all are new discoveries and about 50 per cent are young stellar objects (YSOs). This provides further evidence that YSOs are the commonest high-amplitude infrared variable stars in the Galactic plane. In the 2010–2014 time series of likely YSOs, we find that the amplitude of variability increases towards younger evolutionary classes (class I and flat-spectrum sources) except on short time-scales (<25 d) where this trend is reversed. Dividing the likely YSOs by light-curve morphology, we find 106 with eruptive light curves, 45 dippers, 39 faders, 24 eclipsing binaries, 65 long-term periodic variables ($P > 100$ d) and 162 short-term variables. Eruptive YSOs and faders tend to have the highest amplitudes and eruptive systems have the reddest spectral energy distribution (SEDs). Follow-up spectroscopy in a companion paper verifies high accretion rates in the eruptive systems. Variable extinction is disfavoured by the two epochs of colour data. These discoveries increase the number of eruptive variable YSOs by a factor of at least 5, most being at earlier stages of evolution than the known FUor and EXor types. We find that eruptive variability is at least an order of magnitude more common in class I YSOs than class II YSOs. Typical outburst durations are 1–4 yr, between those of EXors and FUors. They occur in 3–6 per cent of class I YSOs over a 4 yr time span.

Key words: stars: AGB and post-AGB – stars: low-mass – stars: pre-main-sequence – stars: protostars – stars: variables: T Tauri, Herbig Ae/Be – infrared: stars.

1 INTRODUCTION

The VISTA Variables in the Via Lactea (VVV; Minniti et al. 2010) survey has mapped a 560 deg^2 area containing $\sim 3 \times 10^8$ point sources with multi-epoch near-infrared (IR) photometry. The surveyed area includes the Milky Way bulge and an adjacent section of the mid-plane. The survey has already produced a deep near-IR Atlas in five bandpasses (Z, Y, J, H, K_s) and the final product will include a second epoch of the multifilter data and a catalogue of more than 10^6 variable sources.

One of the main scientific goals expected to arise from the final product of VVV is the finding of rare variable sources such as Cataclysmic Variables and RS CVn stars, among others (see Catelan et al. 2013, for a discussion on classes of near-IR variable stars that are being studied with VVV). One of the most important outcomes is the possibility of finding eruptive variable young stellar objects (YSOs) undergoing unstable accretion. Such objects are usually assigned to one of two subclasses: FUors, named after FU Orionis, have long duration outbursts (from tens to hundreds of years); EXors, named for EX Lupi, have outbursts of much shorter duration (from few weeks to several months). Both categories were optically defined in the first instance and fewer than 20 are known in total (see e.g. Reipurth & Aspin 2010; Scholz, Froebrich & Wood 2013;

★ E-mail: cecontrep@gmail.com (CCP); p.w.lucas@herts.ac.uk (PWL)

Audard et al. 2014), very likely because YSOs with high accretion rates tend to suffer high optical extinction by circum-stellar matter. For thorough reviews of the theory and observations in this subject, see Hartmann & Kenyon (1996), Reipurth & Aspin (2010) and Audard et al. (2014). Given that VVV is the first near-IR time domain survey of a large portion of the Galaxy, it is reasonable to hope for a substantial yield of new eruptive variable YSOs in the data set. In particular, we would expect the survey to probe for high-amplitude variability that occurs on typical time-scales of up to a few years, which corresponds more to EXors (or their younger, more obscured counterparts) than to FUors. Eruptive variable YSOs are important because it is thought that highly variable accretion may be common amongst protostars, though rarely observed owing to a duty cycle consisting of long periods of slow accretion and much shorter periods of unstable accretion at a much higher rate. If this is true, it might explain both the observed under-luminosity of low-mass, class I YSOs (the 'Luminosity problem'; see e.g. Kenyon et al. 1990; Evans et al. 2009; Caratti o Garatti et al. 2012) and the wide scatter seen in the Hertzsprung–Russell diagrams of pre-main-sequence (PMS) clusters (Baraffe, Chabrier & Gallardo 2009; Baraffe, Vorobyov & Chabrier 2012).

In the search for this rare class of eruptive variable stars, Contreras Peña et al. (2014) studied near-IR high-amplitude variability in the Galactic plane using the two epochs of UKIDSS Galactic Plane Survey (UGPS) *K*-band data (Lawrence et al. 2007; Lucas et al. 2008). Contreras Peña et al. found that ∼66 per cent of high-amplitude variable stars selected from UGPS data releases DR5 and DR7 are located in star-forming regions (SFRs) and have characteristics of YSOs. They concluded that YSOs likely dominate the Galactic disc population of high-amplitude variable stars in the near-IR. Spectroscopic follow-up confirmed four new additions to the eruptive variable class. These objects showed a mixture of the characteristics of the optically defined EXor and FUor subclasses. Two of them were deeply embedded sources with very steep 1–5 μm spectral energy distributions (SEDs), though showing 'flat-spectrum' SEDs at longer wavelengths. Such deeply embedded eruptive variables are regarded as a potentially distinct additional subclass, though only a few had been detected previously: OO Ser, V2775 Ori, HOPS 383 and GM Cha (see Hodapp et al. 1996; Kóspál et al. 2007; Persi et al. 2007; Caratti o Garatti et al. 2011; Safron et al. 2015). With the aims of determining the incidence of eruptive variability among YSOs and characterizing the phenomenon, we have undertaken a search of the multi-epoch VVV data set. In contrast to UGPS, the ongoing VVV survey offers several dozen epochs of K_s data over a time baseline of a few years. We expect that the VVV survey will also be used to identify YSOs by the common low-amplitude variability seen in nearly all such objects (e.g. Rice, Wolk & Aspin 2012). This will complement studies in nearby star formation regions and in external galaxies, such as the *Spitzer* YSOVAR programme (e.g. Wolk et al. 2015) and a two-epoch study of the Large Magellanic Cloud with *Spitzer* SAGE survey data (Vijh et al. 2009).

We have divided the results of this work in two publications. In this first study, we present the method of the search and a general discussion on the photometric characteristics of the whole sample of high-amplitude variables in the near-IR. We present the follow-up and spectroscopic characteristics of a large subsample of candidate eruptive variable stars in a companion publication (Contreras Peña et al. 2016, hereafter Paper II).

In Section 2 of this work, we describe the VVV survey, the data and the method used to select high-amplitude IR variables. Section 3 describes the make up and general properties of the sample, the evidence for clustering and the apparent association with SFRs. In this section, we also classify the light curves of variables found outside SFRs and use this information to estimate the contamination of our high-amplitude YSO sample by other types of variable star. We then estimate the high-amplitude YSO source density from our sample and compare the average space density with those of other high-amplitude IR variables. In Section 4, we discuss the physical mechanisms that drive variability in YSOs and classify our YSOs via light-curve morphology. This yields some ideas concerning which of the known mechanisms might be responsible for the observed variability. We test these mechanisms using two-epoch JHK_s data. Then we discuss the trends in the likely YSOs as a function of evolutionary status based on their SED. Finally, we discuss the large sample of likely eruptive variables. Section 6 presents a summary of our results.

2 VVV

The regions covered by the VVV survey comprise the bulge region within $-10^\circ < l < +10^\circ$ and $-10^\circ < b < +5^\circ$ and the disc region in $295^\circ < l < 350^\circ$ and $-2^\circ < b < +2^\circ$ (see e.g. Minniti et al. 2010).

The data are collected by the Visible and Infrared Survey Telescope for Astronomy (VISTA). The 4 m telescope is located at Cerro Paranal Observatory in Chile and is equipped with a near-IR camera (VIRCAM) consisting of an array of 16 2048 × 2048 pixels detectors, with a typical pixel scale of 0.339 arcsec, with each detector covering 694 × 694 arcsec^2. The detectors are set in a 4 × 4 array and have large spacing along the *x*- and *y*-axis. Therefore, a single pointing, called a 'pawprint', covers 0.59 giving partial coverage of a particular field of view. A continuous coverage of a particular field is achieved by combining six single pointing with appropriate offsets. This combined image is called a tile. The VVV survey uses the five broad-band filters available in VIRCAM, $Z(\lambda_{\rm eff} = 0.87\,\mu{\rm m})$, $Y(\lambda_{\rm eff} = 1.02\,\mu{\rm m})$, $J(\lambda_{\rm eff} = 1.25\,\mu{\rm m})$, $H(\lambda_{\rm eff} = 1.64\,\mu{\rm m})$ and $K_s(\lambda_{\rm eff} = 2.14\,\mu{\rm m})$.

The VVV survey area is comprised of 348 tiles, 196 in the bulge and 152 in the disc area. Each tile was observed in a single near-contemporaneous multi-filter ($ZYJHK_s$) epoch at the beginning of the campaign, with an exposure time of 80 s per filter. A second epoch of contemporaneous JHK_s was observed in 2015. The variability monitoring was performed only in K_s with an exposure time of 16 s.

The images are combined and processed at the Cambridge Astronomical Survey Unit (CASU). The tile catalogues are produced from the image resulting from combining six pawprints. The catalogues provide parameters such as positions and fluxes from different aperture sizes. A flag indicating the most probable morphological classification is also provided, with '−1' indicating stellar sources, '−2' borderline stellar, '1' non-stellar, '0' noise, '−7' indicating sources containing bad pixels and finally class = −9 related to saturation (for more details on all of the above, see Saito et al. 2012).

Quality control (QC) grades are also given by the European Southern Observatory (ESO) according to requirements provided by the observer. The constraints for VVV K_s variability data are: seeing <2 arcsec, sky transparency defined as 'thin cirrus' or better. The 'master epoch' of multifilter data taken for each tile in a contemporaneous JHK_s observing block and a separate *ZY* observing block have more stringent constraints: seeing <1.0, 1.0, 0.9, 0.9, 0.8 in *Z*, *Y*, *J*, *H*, K_s, respectively, and sky transparency of 'clear' or better. According to whether observations fulfil the constraints established by the observer, these are classified as fully satisfied

(QC A), almost satisfied where, for example, some of the parameters are outside the specified constraints by <10 per cent (QC B) and finally not satisfied (QC C).

2.1 Selection method

In order to search for variable stars, we used the multi-epoch data base of VVV comprising the observations of disc tiles with $|b| \leq 1^{\circ}$ taken between 2010 and 2012. We added the 2013–2015[1] data later to assist our analysis but they were not used in the selection. The catalogues were requested and downloaded from the CASU. We used catalogues of observations with QC grades A, B or C. Catalogues with QC grades C are still considered in order to increase the number of epochs. Some of them were still useful for our purposes. However, a small number of catalogues still presented some issues (e.g. zero-point errors, bad seeing) making them useless and as such were eliminated from the analysis. The number of catalogues in each tile varied from 14 to 23 epochs, with a median of 17 epochs per tile. When the 2013–2015 data were added, the number of epochs available for the light curves rose to between 44 and 59. K_s photometry is derived from *apermag3* aperture fluxes (2 arcsec diameter aperture).

For each tile, the individual catalogues are merged into a single master catalogue. The first catalogue to be used as a reference was selected as the catalogue with the highest number of sources on it. In every case, this corresponded to the catalogue from the deep K_s observation (80 s on source), which was taken contemporaneously with the *J*- and *H*-band data (in 2010). For all other epochs, the time on source was 16 s.

Fig. 1 shows the typical scatter shown by stars at different magnitudes across the analysed range. Here we can see considerable scatter at bright magnitudes, due to effects of saturation and the faint end of the distribution, which is dominated by photon noise. High-amplitude variable star candidates are selected from the master catalogue from stars which fulfilled the following criteria in the 2010–2012 data:

(i) Detection with a stellar morphological classification (class $= -1$) in every available epoch.
(ii) Ellipticity with ell<0.3 in every epoch.
(iii) The absolute difference (ΔK_s) between the brightest ($K_{s,max}$) and faintest point ($K_{s,min}$) in the light curve of the source to be larger than 1 mag (similar to the analysis in Contreras Peña et al. 2014).

The requirement for a detection at every available epoch in the 2010–2012 interval was designed to exclude most transient objects such as novae, as well as reducing the number of false positives. This was the initial classification scheme. However, we observed that for each tile we were selecting a large number of sources as variable star candidates. Fig. 2 shows the average K_s magnitude versus ΔK_s for variable stars in one of the VVV tiles. The figure shows that the majority of stars selected in the original classification scheme are located at the bright and faint ends of the distribution. The latter arise due to unreliable photometry at this faintest part. The VISTA detectors also become increasingly non-linear when reaching the saturation level. This non-linearity is corrected for in the creation

[1] We included a single K_s data point from 2015 observations, corresponding to the epoch with contemporaneous JHK_s photometry. Note that our analysis of the light-curve morphologies and periods is based on the 2010–2014 data only (Section 4.1). The 2015 data became available only after that was complete but they were used in the colour variability analysis (see Section 4.2).

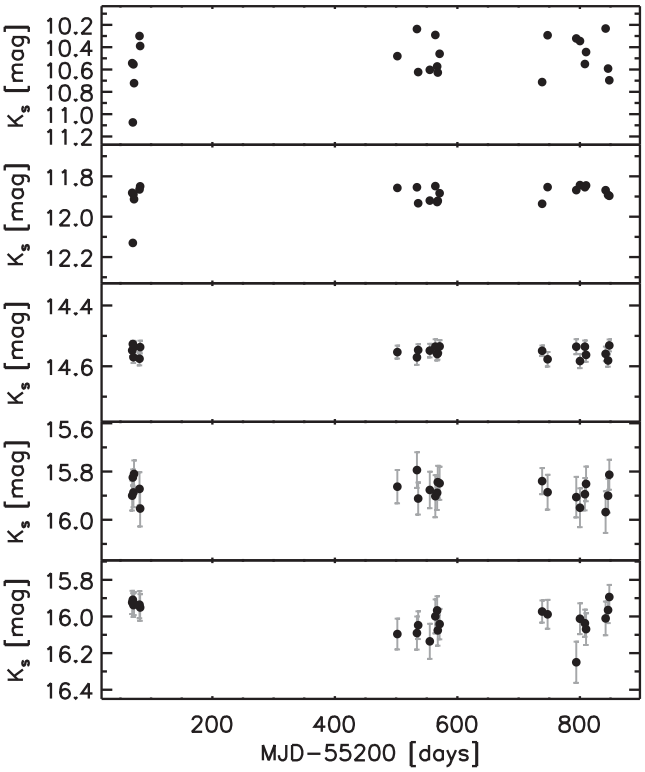

Figure 1. 2010–2012 light curves of 'non-variable' VVV objects (i.e. not classified as high-amplitude variables in our analysis). These are presented to show the typical scatter in magnitude across the analysed magnitude range. We note that photometry for the brightest star is the standard CASU pipeline photometry.

of the catalogues, but differences between the magnitudes of the same object can still be observed, even for objects classified as stellar sources (Saito et al. 2012). Fig. 10 in Saito et al. (2012) shows that when comparing the K_s magnitudes of stellar sources found in overlapping regions of adjacent disc tiles, stars found at the brighter end show an increasing difference in magnitude (an effect also observed in Cioni et al. 2011; Gonzalez et al. 2011). This effect would explain the large differences observed at the brighter end of Fig. 2. This part of the distribution also shows marked 'finger-like' sequences. Each of the sequences can be explained by the fact that the VISTA detectors have different saturation levels. In order to minimize these effects, we applied an additional cut.

(iv) We separated the average K_s distribution of Fig. 2 into bins of 0.5 mag and derived the mean and standard deviation, σ, on ΔK_s for each bin. In order to select an object as a candidate variable star, we required its ΔK_s to be 3σ above the mean ΔK_s at the corresponding magnitude level. This 3σ line is shown in red in Fig. 2 where we can see that it is able to account for the non-linearity effects at the bright end of the distribution.

This additional constraint reduced the number of variable star candidates by a large factor. The initial requirements yield 158 478 stars; the additional cut reduced this to 5085 stars. After the catalogue-based selection, we constructed 1 arcmin × 1 arcmin cut-out images around each candidate for every available epoch. Variable stars were confirmed as real through visual inspection of the individual images (an example is shown in Fig. 3). In some cases, we performed manual photometry with IRAF in order to confirm the variability of the star. The most common causes for the appearance of false positives were bad pixels in the images, saturation of bright

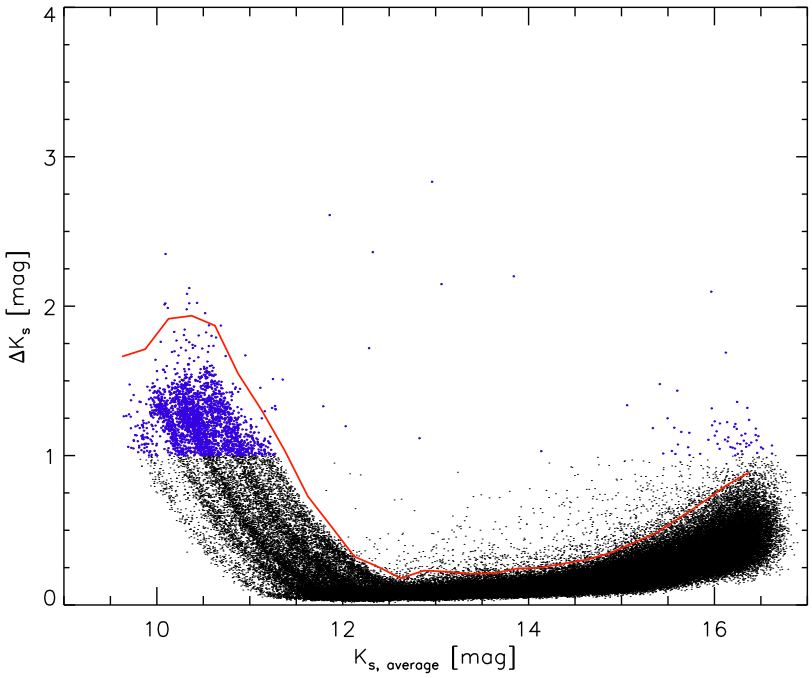

Figure 2. K_s versus ΔK_s for one of the VVV tiles studied in this work, showing stars with class $=-1$ and ellipticity <0.3 in every epoch (black circles). Variable star candidates which fulfil the condition $\Delta K_s > 1$ mag are shown as blue circles. The red solid line marks the additional 3σ cut applied to the objects as explained in the text. Stars above this line are selected for subsequent visual inspection.

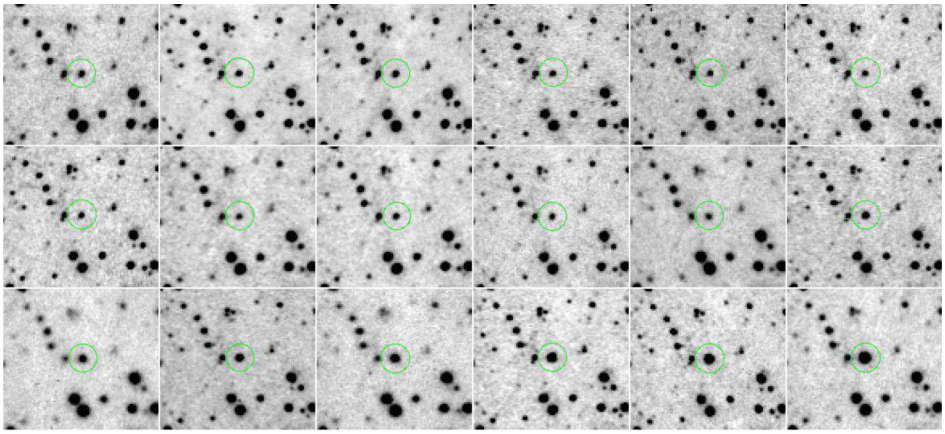

Figure 3. Example of the images used to visually inspect variable star candidates. In this case, we show the images, taken between 2010 and 2012, of variable star VVVv322. Each image has a size of 1 arcmin $\times$ 1 arcmin. The star gets brighter towards the end of the sequence.

sources, diffraction spikes and stars that were found on the edge of tiles. In the case of saturation, if this effect was present, it was quite evident in individual images. In most cases, saturation was observed in every single epoch, thus the variability observed in the light curve plots was not real and the source was marked as such.

This selection method yielded a total of 840 real variable stars. However, 25 of them are found twice as they are covered by adjacent tiles in VVV. The final list of VVV high-amplitude IR variables consists of 816 stars. This includes one variable star, VVVv815, that showed large variability in 2010 but did not meet all the selection criteria. The average magnitude for objects in the selected sample was found in the range $10.3 < K_s < 16.9$ mag.

Our requirement for a high-quality detection at every epoch between 2010 and 2012 (see items i and ii) is bound to cause us to miss some real variables, very likely including some of the faintest or highest amplitude variables if they dropped below $K_s \sim 16$ during that time or if they became saturated and were therefore no longer classified as point sources. A significant fraction of all VVV sources are blended with adjacent stars and they can fail to pass our cuts on the morphological class and ellipticity at one or more epochs in consequence. The same can be true for YSOs with extended reflection nebulae or strong H_2 jets, as they might have slightly extended morphologies and fail to be classified as point sources (see e.g. Chen et al. 2009; Ward et al. 2016). However, sources that pass these quality cuts are likely to be unblended and therefore to have reliable photometry (photometry from the VISTA pipeline is not always reliable for faint stars in crowded fields, see e.g. Saito et al. 2012). In order to check the reliability of the pipeline photometry, especially for faint sources with $K_s = 15$–16 mag, we obtained

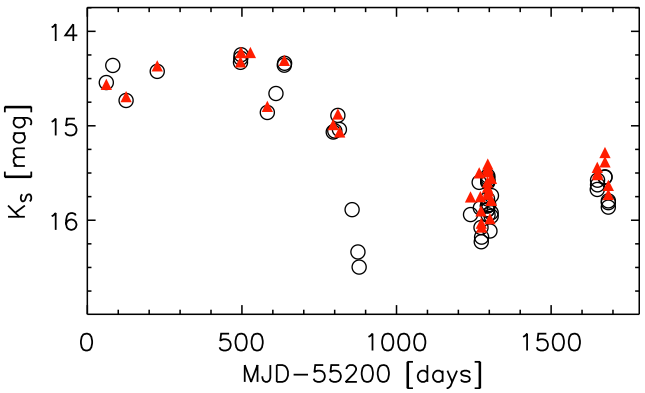

Figure 4. PSF (red triangles) versus aperture photometry (open circles) of star VVVv316. PSF photometry is shown only for data points classified as 'good' or 'good but faint' by DOPHOT.

point spread function (PSF) fitting photometry of all stars in tile d069 with DOPHOT (Schechter, Mateo & Saha 1993). The results confirmed that the variables found by our selection have reliable pipeline photometry. This is illustrated in Fig. 4 for variable star VVVv316, where the comparison of DOPHOT and VISTA pipeline photometry shows close agreement.

We investigated the incompleteness of our selection by examining two widely separated VVV disc tiles (d064 and d083), in which we removed our class and ellipticity cuts and required a minimum of only one detection in each year from 2010 to 2012 (with a stellar profile classification). This continues to select against transients and perhaps the most extreme variable YSOs but it allows us to assess incompleteness due to blending, which can cause sources to be absent or to have different profile classifications at different epochs. We found that this more relaxed selection added over 400 additional candidates in the two tiles down to (mean) K_s=15.5, an increase of more than a factor of 10. Following visual inspection, we found that the number of real high-amplitude variables was increased by a factor of ∼2, up to a limit of K_s= 15.5. At fainter mean magnitudes the completeness of our selection with criteria (i)–(iv) falls more steeply because most high-amplitude variables will not satisfy the quality cuts at every epoch as the sensitivity limit is approached.

A case of this selection effect is found in a variable star VVVv815 mentioned above. It showed a large variation ($\Delta K_s > 1$ mag) in the analysis of an early release of 2010 data. However, the star does not show up as a variable star candidate in the analysis described above. Inspection of the master catalogue for the respective tile shows that the star has a classification different from stellar in 3 out of 18 epochs available for tile d090 in the 2010–2012 period. This star is included in our final list of VVV high-amplitude variables because it is also part of the sample that has follow-up spectroscopic observations.

The number of stars in the analysed VVV area that fulfil criteria (i) and (ii) above is 12 789 000 stars. Considering the number of real variable stars, we see that high-amplitude IR variability was observed in approximately 1 out of 15 000 stars in the Galactic mid-plane at $295^\circ < l < 350^\circ$.

2.2 Issues with saturation

The aforementioned problems of saturation still affect a small number of stars in our sample. This effect can become important when individual epochs of stars in our sample are brighter than $K_s = 12$ mag. Saturation will reduce the flux at the inner core of the star, thus the magnitude of the star derived by using smaller diameter aperture than the default 2 arcsec aperture will be fainter than the magnitude obtained from the default aperture.

In order to check whether the star is saturated, we first obtain the magnitudes from aperture photometry using the measure fluxed from five different diameters for the apertures. These diameters are 1 (Apermag1), 1.41 (Apermag2), 2 (Apermag3), 2.82 (Apermag4) and 4 arcsec (Apermag5). We find that saturated stars show relatively large differences between the magnitudes from the first three apertures and we set a threshold for saturation as stars having both Apermag1-Apermag3 > 0.05 mag and Apermag2-Apermag3 > 0.02 mag. Thus, any individual epoch of a star with Apermag3 < 12 mag (in the K_s passband) and having these differences is flagged as saturated.

In order to correct for saturation, we follow Irwin (2009) and defined a ring outside the saturated core to obtain a new flux estimate. We then determine an aperture correction for the ring from a set of bright, unsaturated stars found within 5 arcmin of our object of interest. In our analysis, we derived new fluxes using the ring defined by Apermag2 and Apermag4. Comparison with 2MASS K_s photometry indicates that this choice of apertures extends the dynamic range by 2.5 mag, relative to the pipeline photometry. We caution that comparison with 2MASS K_s photometry indicates that while this approach provides correct magnitudes, the uncertainties are large, typically 0.2 mag.

In Table 1, we present the 2010–2015 photometry for the 816 high-amplitude variable stars from VVV.

Table 1. K_s photometry of the 816 high-amplitude variable stars from VVV. The full version of the table is available online.

ID	MJD-55200 (d)	K_s (mag)	$K_{s,err}$ (mag)
VVVv1	69.1171875	14.400	0.016
VVVv1	70.1484375	14.436	0.010
VVVv1	71.1445312	14.381	0.016
VVVv1	72.1406250	14.225	0.014
VVVv1	81.0781250	14.606	0.022
VVVv1	82.1015625	14.680	0.024
VVVv1	501.9882812	14.297	0.017
VVVv1	533.9648438	15.271	0.044
VVVv1	535.9726562	15.133	0.033
VVVv1	554.9726562	14.853	0.029
VVVv1	564.0117188	14.223	0.019

3 HIGH-AMPLITUDE IR VARIABLES FROM VVV

3.1 General characteristics

The selection method of Section 2.1 yields 816 high-amplitude ($\Delta K_s > 1$ mag) IR variables. In order to study the properties of these stars, we searched for additional information in available public data bases. This search can be summarized as follows.

(i) SIMBAD: we query the SIMBAD data base (Wenger et al. 2000) for astronomical objects within a radius of 5 arcmin centred on the VVV object.

(ii) Vizier: additional information was provided with the use of the Vizier data base (Ochsenbein, Bauer & Marcout 2000). Here we queried whether the VVV object was found within 2 arcsec of

objects found in astronomical catalogues that were not available in SIMBAD.

(iii) The NASA/IPAC Infrared Science Archive (IRSA): here we queried for additional information in several near- and mid-IR public surveys, which include 2MASS (Skrutskie et al. 2006), DENIS (Epchtein et al. 1994), *Spitzer*/GLIMPSE surveys (see e.g. Benjamin et al. 2003), *WISE* (Wright et al. 2010), Akari (Murakami et al. 2007) and MSX6C (Price et al. 2001). The search was done automatically using the IDL scripts provided at the IRSA website. The catalogues were queried for objects found within a 10 arcsec radius of the VVV object. In most cases, several objects are found at these distances, then we only selected the nearest object to our star. In order to confirm whether these detections correspond to our variable star, 1 arcmin × 1 arcmin VVV images around the star were visually inspected.

In addition, we used the *WISE* image service within IRSA to inspect multicolour images of the areas around our variable stars, in order to establish a possible association of the VVV object with an SFR.

The general properties of the sample can be found in Table 2. Column 1 presents the original designation given to the sources. Column 2 corresponds to the full VVV designation for the source. Coordinates for the objects are given in Columns 3 and 4, with columns 5 and 6 presenting the Galactic coordinates of the sources. In columns 7–11, we present the nearly contemporaneous Z, Y, J, H, K_s photometry from VVV. Column 12 gives ΔK_s, the absolute value of the peak-to-trough difference from the 2010–2014 light curves from VVV. Column 13 presents $\alpha_{\rm class}$, the 2–23 μm spectral index parameter that relates to the evolutionary class of sources that appear to be associated with SFRs (the method and data used to estimate this parameter is explained in Section 4.3). In column 14, we present whether the object is likely associated with SFRs or not, whilst column 15 presents the classification of the object from its light curve. The latter is discussed throughout the text. Finally, in column 16, we present the approximate period for variable stars where we are able to measure this parameter.

Most of the variable stars are unknown from searches in SIMBAD and Vizier (∼98 per cent). Among the known variables, there are two novae, Nova Cen 2005 and Nova Cen 2008 (Hughes et al. 2010; Saito et al. 2013), two eclipsing binaries (EBs), EROS2-GSA J132758-631449 and PT Cen (Derue et al. 2002; Budding et al. 2004), one high-mass X-ray binary, BR Cir (see e.g. Tudose et al. 2008) and nine OH/IR stars. Among the objects not previously classified as variable stars, 159 are found in the *Spitzer*/GLIMPSE catalogue of intrinsically red sources from Robitaille et al. (2008), with the majority being classified as likely YSOs from their mid-IR colours and brightness.

3.2 YSO population

At this point, the reader should note that most of the variables are listed as spatially associated with SFRs (∼65 per cent, falling to ∼50 per cent after allowing for chance projection by non-YSOs, see Section 3.3, 3.4 and 4) and these stars have spectral indices that indicate a class I or flat-spectrum evolutionary status, therefore they are likely in an early evolutionary stage. They are usually sufficiently red to be undetected ($i > 21$ mag) in sensitive panoramic optical surveys such as VPHAS+ (Drew et al. 2014) unlike most of the known FUor and EXors. The spectral indices of the YSOs are discussed later in Section 4.3 following classification of the light curves and an attempt to decontaminate chance projections of other variables in SFRs.

3.3 Association with SFRs

Fig. 5 shows the distribution for the 816 VVV variables across the Galactic mid-plane. It can be seen that our objects appear to be highly clustered, with their distribution following that of the SFRs from the Avedisova et al. (2002) catalogue (red diamonds in the bottom plot of Fig. 5). To study the apparent clustering, we derive the two-point angular correlation function, $w(\theta)$, and the nearest neighbour distribution of the sample of high-amplitude variables. To derive $w(\theta)$, we follow Bate, Clarke & McCaughrean (1998) and first estimate the mean surface density of companions (MSDC). For each star, we compute the angular separation, θ, to all other stars, and bin the separation into annuli of separation θ to $\theta + \delta\theta$. The MSDC results from dividing the number of pairs found, $N_{\rm p}$, at a given separation by the area of the annulus and dividing by the total number of stars, N_*, or

$$\mathrm{MSDC} = \frac{N_{\rm P}}{2\pi\theta\delta\theta N_*}.$$

The latter relates to $w(\theta)$ as

$$w(\theta) = \mathrm{MSDC} \times \frac{A}{N_*} - 1,$$

where A is the area covered by the survey (Bate et al. 1998, and references therein). This correlation function is valid as long as the separations θ are smaller than the angular length of the sample. We show the two-point correlation function in Fig. 6. We can see that we do not find any pairs for separations $\theta < 20$ arcsec, hence $w(\theta) = -1$. For separations between 20 and 100 arcsec, $w(\theta)$ is larger than the values expected for random pairings ($w(\theta) = 0$) and it remains somewhat above zero for separations up to a few hundred arcseconds. This nearest neighbour distribution of Fig. 6 also shows an excess of close neighbours at distances $R < 200$ arcsec, compared to the expected number from a random (Poisson) distribution. Thus, we are confident that we are tracing clustering in the VVV sample, on a spatial scale similar to that of distant Galactic clusters and star formation regions.

As an illustration of how variable stars in VVV are preferentially located in areas of star formation, Fig. 7 shows the K_s image of the area covered in tile d065. 25 highly variable stars are found in this tile and it is clear that these are not evenly distributed along the area covered in d065 and instead are found clustered around an area of star formation, which is better appreciated in the cut-out image from *WISE* (Wright et al. 2010).

To establish a likely association with an SFR, we used the criteria established in the UGPS search (see Contreras Peña et al. 2014), which were based on entries in the SIMBAD data base and the Avedisova catalogue within a 5 arcmin radius of each high-amplitude variable. In addition, we also check *WISE* images for evidence of star formation in the area of the object, e.g. evidence of bright extended 12 μm emission near the location of the objects or the finding of several stars with red $W1-W2$ colours (sources appearing green, yellow or red in *WISE* colour images) around the VVV object. We find that 530 of our variable stars are spatially associated with SFRs, which represents 65 per cent of the sample, remarkably similar to the observed association in UGPS objects (Contreras Peña et al. 2014).

3.4 Contamination by non-YSOs

In Contreras Peña et al. (2014), we estimated that about 10 per cent of objects are probable chance associations with SFRs. This number is likely to be larger in our current analysis, given that: (i) we are sampling mid-plane sightlines across the Galactic disc; (ii) the

Table 2. Parameters of the high-amplitude variables from VVV. For the description of the columns, see Section 3.1. Here we show the first 40 sources in the list. The complete list is available online.

Object ID	VVV designation	α (J2000)	δ (J2000)	l (°)	b (°)	Z (mag)	$Z_{\rm err}$ (mag)	Y (mag)	$Y_{\rm err}$ (mag)	J (mag)	$J_{\rm err}$ (mag)	H (mag)	$H_{\rm err}$ (mag)	$K_{\rm s}$ (mag)	$K_{\rm s,err}$ (mag)	$\Delta K_{\rm s}$ (mag)	$\alpha_{\rm class}$	SFR	Class[a]	Period[b] (d)
VVVv1	VVV J114135.16-622055.51	11:41:35.16	−62:20:55.51	294.926 03	−0.567 70	–	–	–	–	17.99	0.06	15.95	0.02	14.44	0.01	1.34	−0.29	y	STV	72
VVVv2	VVV J114412.94-623449.09	11:44:12.94	−62:34:49.09	295.280 05	−0.711 46	–	–	–	–	–	–	18.78	0.23	15.71	0.03	2.52	1.22	y	STV	–
VVVv3	VVV J115113.03-623729.29	11:51:13.03	−62:37:29.29	296.071 99	−0.557 84	13.17	0.01	12.93	0.01	12.90	0.01	12.70	0.01	12.24	0.01	2.21	–	n	Known	–
VVVv4	VVV J115808.69-630708.60	11:58:08.69	−63:07:08.60	296.950 57	−0.867 85	–	–	–	–	18.23	0.08	16.62	0.04	15.32	0.02	1.10	−0.24	y	STV	–
VVVv5	VVV J115959.68-622613.20	11:59:59.68	−62:26:13.20	297.020 26	−0.157 16	17.69	0.02	16.62	0.01	15.87	0.01	15.25	0.01	13.53	0.01	1.30	–	n	LPV	–
VVVv6	VVV J115937.81-631109.77	11:59:37.81	−63:11:09.77	297.128 36	−0.899 53	19.02	0.05	18.08	0.04	16.80	0.02	15.95	0.02	15.50	0.02	1.02	–	n	EB	1.64
VVVv7	VVV J120202.67-623615.60	12:02:02.67	−62:36:15.60	297.284 72	−0.275 38	–	–	–	–	–	–	–	–	17.22	0.12	1.60	2.39	y	Eruptive	–
VVVv8	VVV J120059.11-631636.18	12:00:59.11	−63:16:36.18	297.295 82	−0.958 38	–	–	–	–	–	–	–	–	16.86	0.09	1.38	0.64	y	Eruptive	–
VVVv9	VVV J120217.23-623647.83	12:02:17.23	−62:36:47.83	297.313 81	−0.278 88	–	–	–	–	18.29	0.08	16.33	0.03	14.64	0.01	2.78	−0.39	y	Dipper	–
VVVv10	VVV J120250.85-622437.62	12:02:50.85	−62:24:37.62	297.339 12	−0.067 49	18.57	0.04	18.23	0.05	16.97	0.03	16.35	0.03	16.07	0.04	1.19	–	n	STV	–
VVVv11	VVV J120436.62-625704.60	12:04:36.62	−62:57:04.60	297.637 41	−0.561 88	20.20	0.19	19.01	0.13	17.87	0.07	16.79	0.05	16.08	0.08	1.10	–	n	STV	–
VVVv12	VVV J121033.19-630755.71	12:10:33.19	−63:07:55.71	298.331 85	−0.626 11	–	–	–	–	–	–	16.30	0.03	15.04	0.03	1.74	0.28	y	Fader	–
VVVv13	VVV J121216.83-624838.32	12:12:16.83	−62:48:38.32	298.476 03	−0.278 14	–	–	–	–	–	–	–	–	16.72	0.14	1.40	1.39	y	STV	–
VVVv14	VVV J121218.13-624904.48	12:12:18.13	−62:49:04.48	298.479 58	−0.284 95	19.48	0.10	18.88	0.12	17.84	0.06	16.74	0.05	15.56	0.05	1.29	0.88	y	LPV-YSO	124
VVVv15	VVV J121226.09-624416.97	12:12:26.09	−62:44:16.97	298.482 52	−0.203 71	19.00	0.07	17.72	0.04	16.57	0.02	15.60	0.02	15.04	0.03	1.09	–	y	EB	2.27
VVVv16	VVV J121329.76-624107.74	12:13:29.76	−62:41:07.74	298.594 98	−0.133 64	18.01	0.03	17.13	0.02	16.06	0.01	14.72	0.01	13.66	0.01	1.29	0.92	y	Eruptive	–
VVVv17	VVV J121352.08-625549.90	12:13:52.08	−62:55:49.90	298.672 78	−0.369 86	–	–	–	–	–	–	17.79	0.12	16.41	0.10	1.22	0.44	y	STV	–
VVVv18	VVV J121950.31-632142.24	12:19:50.31	−63:21:42.24	299.398 68	−0.706 95	17.82	0.02	17.29	0.02	16.20	0.01	15.52	0.01	15.32	0.02	1.04	–	n	STV	–
VVVv19	VVV J122255.30-632352.56	12:22:55.30	−63:23:52.56	299.745 94	−0.702 70	19.56	0.08	18.67	0.06	17.55	0.04	16.40	0.03	15.61	0.03	1.82	–	n	EB	16.88
VVVv20	VVV J122827.97-625713.97	12:28:27.97	−62:57:13.97	300.324 02	−0.198 49	–	–	–	–	17.38	0.04	14.10	0.01	11.70	0.01	1.71	0.60	y	Eruptive	–
VVVv21	VVV J122902.24-625234.10	12:29:02.24	−62:52:34.10	300.381 93	−0.115 33	–	–	–	–	–	–	17.12	0.05	15.80	0.03	1.79	0.86	y	LPV-YSO	603
VVVv22	VVV J123105.60-624457.34	12:31:05.60	−62:44:57.34	300.605 47	0.030 57	–	–	–	–	18.81	0.14	16.94	0.05	15.55	0.03	1.73	−0.34	y	STV	–
VVVv23	VVV J123128.53-624433.10	12:31:28.53	−62:44:33.10	300.648 55	0.040 70	19.44	0.07	18.37	0.05	17.14	0.03	15.57	0.02	14.40	0.01	1.51	−0.20	y	Fader	–
VVVv24	VVV J123235.68-634319.61	12:32:35.68	−63:43:19.61	300.847 94	−0.926 62	17.17	0.01	16.13	0.01	14.05	0.01	12.95	0.01	12.23	0.01	1.20	–	n	LPV	430
VVVv25	VVV J123514.37-624715.63	12:35:14.37	−62:47:15.63	301.081 29	0.025 87	–	–	–	–	–	–	16.01	0.02	12.34	0.01	1.68	0.22	y	Eruptive	–
VVVv26	VVV J123845.66-631136.03	12:38:45.66	−63:11:36.03	301.503 20	−0.356 74	–	–	–	–	19.67	0.29	16.70	0.04	14.71	0.01	2.45	1.07	y	Eruptive	–
VVVv27	VVV J123848.33-633939.15	12:38:48.33	−63:39:39.15	301.531 14	−0.823 47	–	–	–	–	–	–	–	–	12.04	0.01	2.73	–	n	LPV	1329
VVVv28	VVV J123911.54-630524.76	12:39:11.54	−63:05:24.76	301.546 88	−0.251 38	–	–	–	–	–	–	18.91	0.32	16.78	0.09	1.39	0.54	y	STV	3.68
VVVv29	VVV J123931.48-630720.38	12:39:31.48	−63:07:20.38	301.585 93	−0.281 70	–	–	–	–	–	–	–	–	16.94	0.10	2.13	1.32	y	Eruptive	–
VVVv30	VVV J124140.56-635033.57	12:41:40.56	−63:50:33.57	301.856 16	−0.991 28	17.92	0.02	17.22	0.02	16.40	0.02	15.66	0.02	15.23	0.02	1.22	–	n	EB	1.91
VVVv31	VVV J124140.15-635918.05	12:41:40.15	−63:59:18.05	301.860 93	−1.136 89	13.91	0.01	13.27	0.01	12.46	0.01	12.08	0.01	11.69	0.01	1.15	–	n	LPV	760
VVVv32	VVV J124357.15-625445.09	12:43:57.15	−62:54:45.09	302.079 91	−0.053 14	19.58	0.07	18.08	0.04	16.07	0.01	14.07	0.01	12.45	0.01	2.49	0.33	y	LPV-YSO	–
VVVv33	VVV J124425.05-631355.76	12:44:25.05	−63:13:55.76	302.141 53	−0.371 16	–	–	–	–	–	–	18.10	0.16	16.43	0.07	1.37	–	n	STV	–
VVVv34	VVV J125029.87-625124.93	12:50:29.87	−62:51:24.93	302.824 70	0.014 65	19.22	0.05	18.32	0.05	17.86	0.06	16.70	0.04	15.71	0.03	1.30	−0.34	y	STV	–
VVVv35	VVV J125206.52-635711.52	12:52:06.52	−63:57:11.52	303.005 57	−1.081 52	17.62	0.01	16.85	0.01	17.13	0.03	16.21	0.03	15.97	0.04	1.24	–	n	EB	1.12
VVVv36	VVV J125917.72-633008.44	12:59:17.72	−63:30:08.44	303.808 25	−0.643 94	14.38	0.01	13.84	0.01	13.27	0.01	12.56	0.01	11.81	0.01	1.03	−0.11	y	Eruptive	–
VVVv37	VVV J130243.05-631130.00	13:02:43.05	−63:11:30.00	304.203 31	−0.347 74	18.80	0.03	17.75	0.03	16.74	0.02	15.89	0.02	15.38	0.03	1.34	–	n	STV	–
VVVv38	VVV J130311.38-631439.09	13:03:11.38	−63:14:39.09	304.254 11	−0.402 59	–	–	–	–	–	–	17.79	0.13	16.32	0.07	1.41	1.20	y	STV	48.96
VVVv39	VVV J130440.98-635313.45	13:04:40.98	−63:53:13.45	304.388 93	−1.052 77	18.67	0.03	17.89	0.03	17.57	0.05	16.38	0.04	15.49	0.03	1.51	–	n	STV	–
VVVv40	VVV J130600.43-630144.40	13:06:00.43	−63:01:44.40	304.582 98	−0.203 94	20.71	0.20	19.49	0.14	16.52	0.02	15.42	0.02	14.69	0.01	1.24	–	n	EB	10.29

Notes. [a]LPV:long-period variable; LPV-Mira:long-period variable found in SFR but with Mira-like characteristics; LPV-YSO:long-period variable found in SFR with YSO-like characteristics. STV:short-term variable; EB:eclipsing binary

[b]We note that in some objects classified as long-period variables (LPV, LPV-MIRA and LPV-YSO), the light curve shows a clear periodic behaviour over long time-scales (>100 d). However, we are unable to measure an exact period. In those cases this column is left blank.

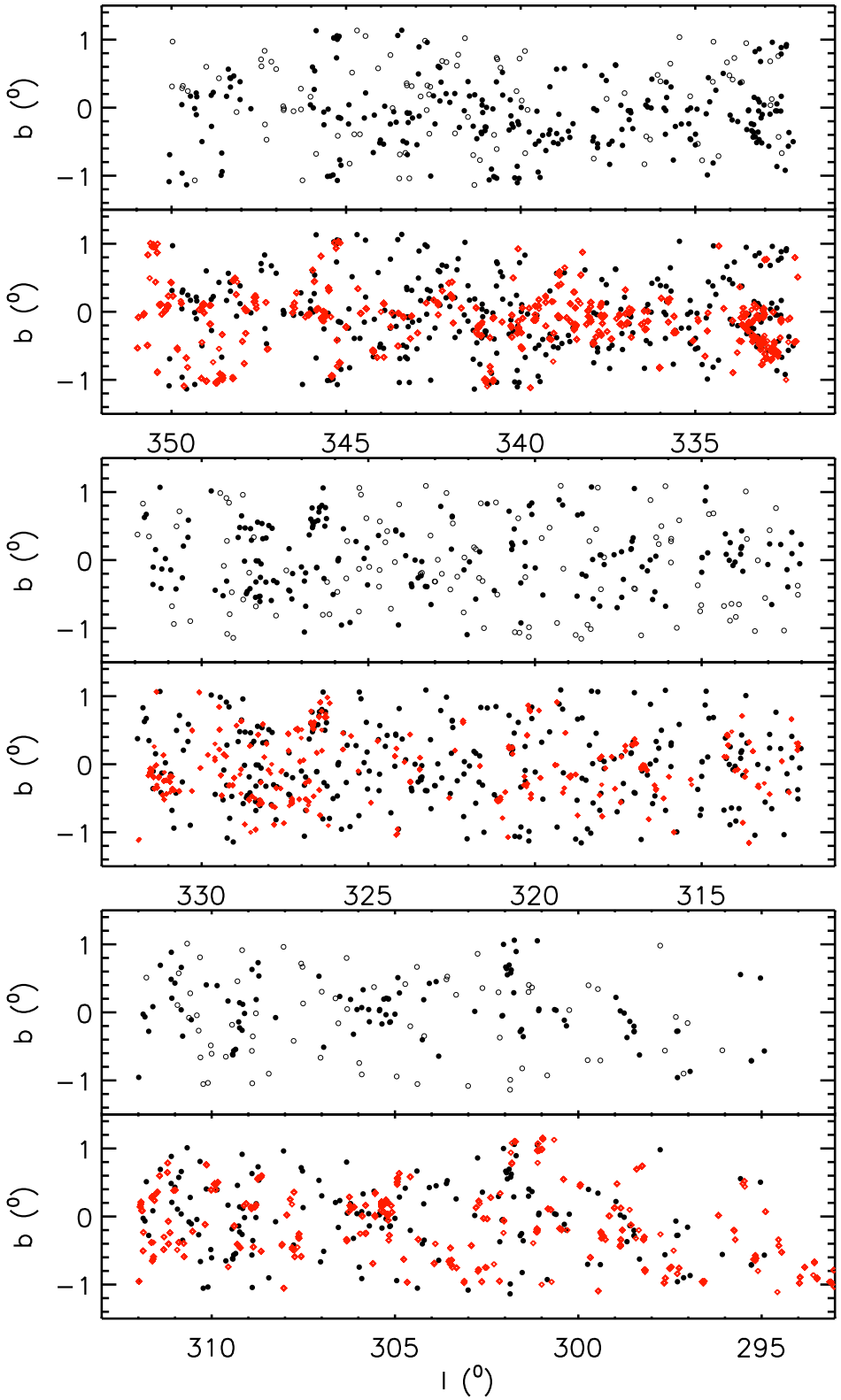

Figure 5. (top) Galactic distribution of high-amplitude variable stars selected from VVV. These are divided into objects likely associated with SFRs (black filled circles) and those that are found outside these regions (black open circles). The bottom graph shows the same distribution for the 816 high-amplitude variables (black circles), but this time including the areas of star formation from the Avedisova (2002) catalogue (red diamonds).

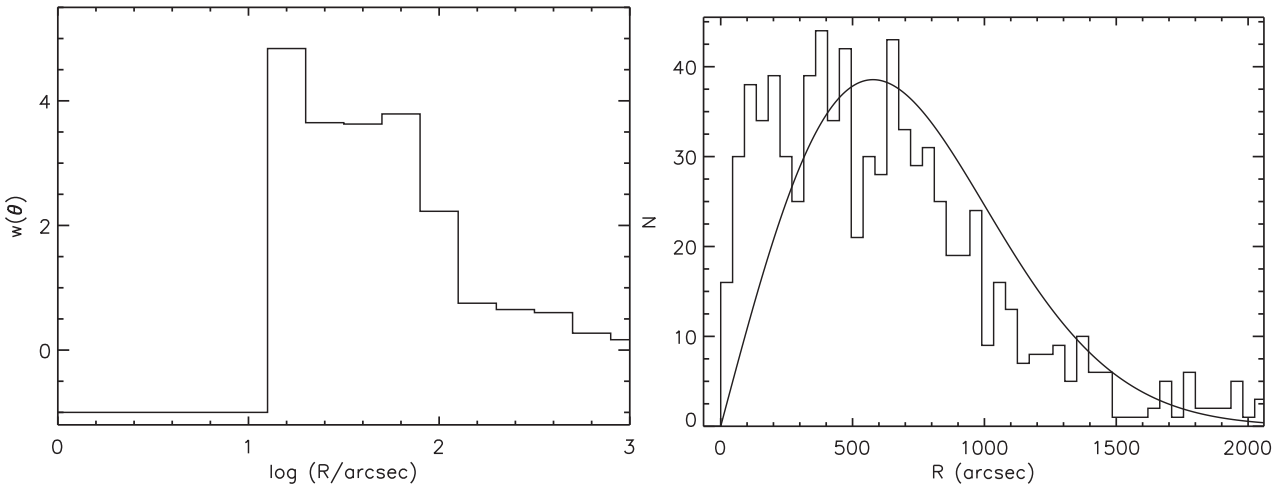

Figure 6. (left) Two-point angular correlation function of the sample of VVV high-amplitude variables. (right) Nearest neighbour distribution for the same sample. The smooth curve represents the expected distribution of a random (Poisson) distribution.

higher extinction in the Galactic mid-plane and the brighter saturation limit of VVV than UGPS allows for a larger number of bright evolved variable stars to show up in our results. To determine the percentage of objects that might be catalogued as likely associated with SFRs by chance, we used the following method.

(1) Create a master catalogue of objects in the 76 tiles that were classified as stars in each of the epochs from the 2010–2012 analysis.

(2) Select 816 stars randomly from this catalogue and query SIMBAD for objects found in a 5 arcmin radius.

(3) Count the number of objects within this radius that were classified in any of the categories that could relate to star formation. This categories included T Tauri and Herbig Ae/Be stars, H II regions, Dark clouds, dense cores, millimeter and submillimeter sources, FU Orionis stars, among others.

(4) Repeat items (2) through (3) 40 times.

Fig. 8 shows the number of stars found within 5 arcmin of the VVV object and that were classified in the categories shown above, $N_{\rm simbad}$ versus the percentage of VVV objects with this number. It is already apparent that the percentage of chance selection will be higher than that estimated from GPS. However, we note that in order for an object to be flagged as associated with an SFR in Table 2, we needed at least four SIMBAD objects to appear within the 5 arcmin radius, thus giving us an estimate that ∼30 per cent of the non-YSO population is spatially associated with an SFR by chance. Inspection of *WISE* images of 100 randomly selected sources yields a similar fraction of chance associations but most of these were also identified as SFR-associated from the SIMBAD results, so the *WISE* data only slightly increased the chance association fraction. The Avedisova catalogue added an even smaller fraction of chance associations not indicated by the SIMBAD and *WISE* data, so the final chance association probability for non-YSOs with star formation regions is 35 per cent.

The number of non-YSOs in the SFR-associated sample is less than 35 per cent because non-YSOs do not dominate the full high-amplitude sample but constitute about half of it. We found 286/816 (35 per cent) of variables outside SFRs, i.e. in 65 per cent of the area, suggesting that 54 per cent (35/0.65) of the sample is composed of objects other than YSOs but this neglects the fact that some YSOs will be members of SFRs that are not known in the literature nor visible in *WISE* images (see Section 3.5). Consequently, random addition of 35 per cent of half of the total sample of 816 variables to the SFR-associated sample would be expected to cause only 27 per cent contamination of the SFR-associated subsample by non-YSOs. This conclusion that the SFR-associated population of variables is dominated by bona fide YSOs is supported by the two colour diagrams (Figs 10 and 22) and light curves of the population (see Sections 3.5 and 4.1), which differ from those outside SFRs. We note that the results of spectroscopic follow-up of a subsample of VVV objects associated with SFRs (Paper II) show a figure of 25 per cent, a consistent figure despite some additional selection effects in that subsample.

3.5 Properties of variables outside SFRs

To establish the nature of the objects that could be contaminating our sample of likely YSOs and may also be interesting variable stars, we study the properties of objects found outside areas of star formation. Many of these are listed in SIMBAD as IR sources (from the IRAS and MSX6C catalogues) and associated with OH masers, as well as being catalogued as evolved stars in previous surveys. Visual inspection of their light curves also shows that a large percentage of objects have periodic variability, with $P > 100$ d, whilst the remainder of the sample shows variability over shorter time-scales of the order of $P < 20$ d. We use the phase dispersion minimization (PDM; Stellingwerf 1978) algorithm found in the NOAO package in IRAF to search for a period in the light curve of these objects. This allows us to derive the periods or at least the approximate time-scale of the variability of objects found outside areas of star formation. To provide a comparison with the PDM results, we also used the GATSPY LOMBSCARGLEFAST implementation of the Lomb–Scargle algorithm, which benefits from the automatic optimization of the frequency grid so that significant periods are not missed. We found that PDM was generally better for the purpose of this initial investigation. The Lomb–Scargle algorithm is designed to detect sinusoidal variations, whereas PDM makes no assumptions about the form of the light curve and is therefore much more sensitive to the periods of EBs, for example. The Lomb–Scargle method did help with the classification of a small number of long period variables (LPVs).

Out of the 286 stars in this subsample, five objects correspond to known objects from the literature (novae, EBs and a high-mass

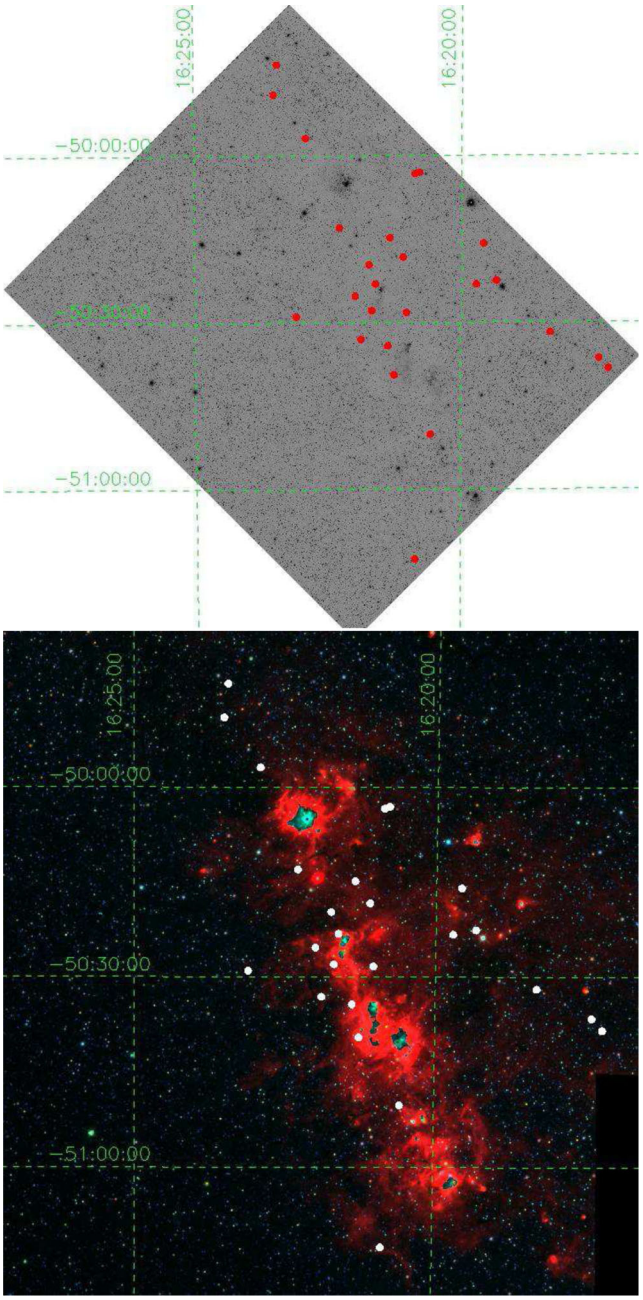

Figure 7. The top graph shows the K_s image of tile d065 along with the high-amplitude variable stars found in this region. The clustering of the variable stars is already apparent in this image. The bottom graph shows the *WISE* false colour image of the same region (blue = 3.5 μm, green = 4.6 μm, red = 12 μm). Here, the fact that variable stars preferentially locate around areas of star formation can be better appreciated.

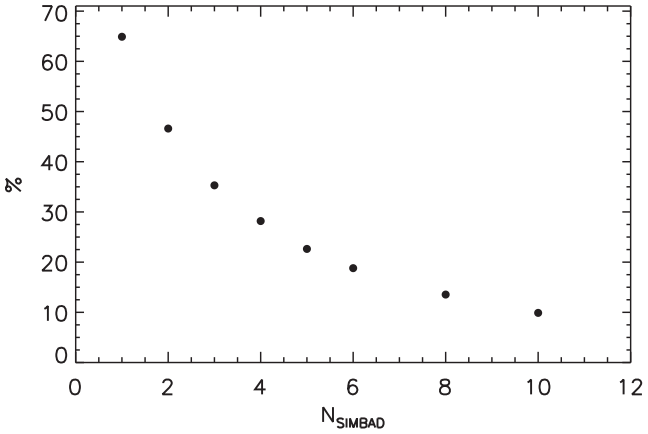

Figure 8. Percentage of objects flagged as likely associated with SFRs as a function of the number of objects classified in the categories that could relate to star formation found within a 5 arcmin radius query in SIMBAD.

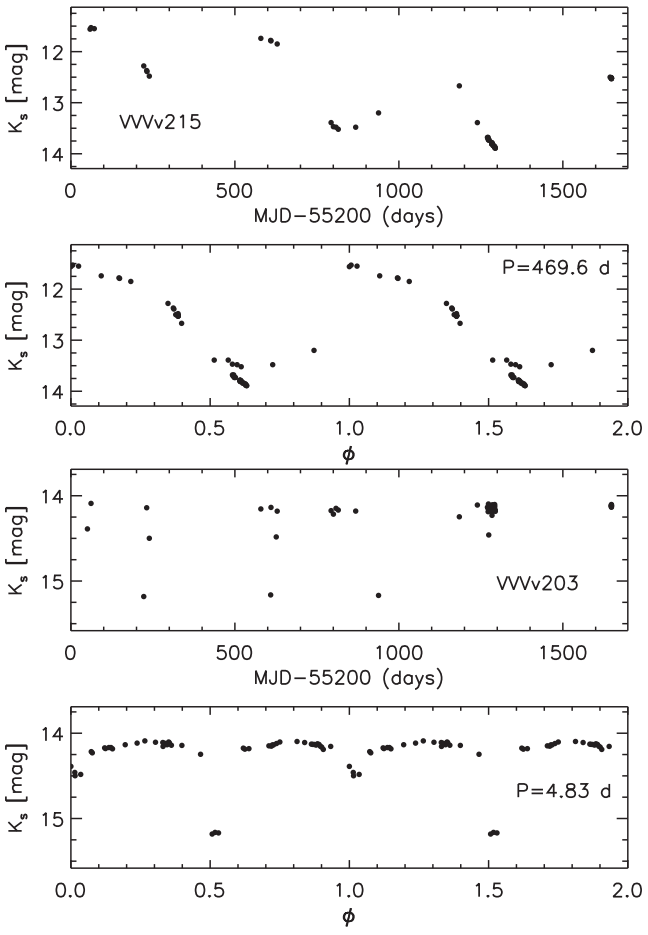

Figure 9. Examples of K_s light curves for objects not associated with SFRs. (top) the LPV VVVv215. (bottom) The EB VVVv203.

X-ray binary), 45 per cent of them are LPVs and 17 per cent are EBs where we are able to measure a period. In addition, 30 per cent of the sample is comprised of objects in which variability seems to occur on short time-scales. The light curves of many objects in the latter group resemble those of EBs with measured periods, but with only one or two dips sampled in the data set. We suspect that most of these could also be EBs but we are not able to establish the periods. Finally, we also find 18 objects (6 per cent) that do not appear to belong to any of the former classes. In Fig. 9, we show two examples of objects belonging to these different subclasses where a period could be derived. The LPV VVVv215 is typical of many of the dusty Mira variables in the data set that show long-term trends caused by changes in the extinction of the circum-stellar dust shell. These trends are superimposed on the pulsational, approximately sinusoidal variations, with the result that the K_s magnitude, at a given point in the phase curve, can differ between cycles.

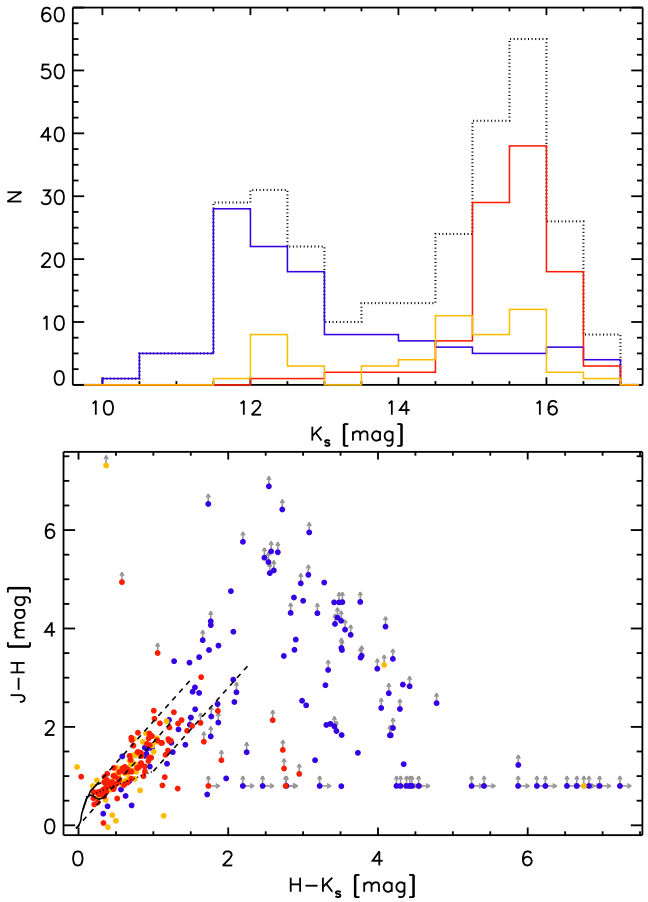

Figure 10. (top) Overall K_s distribution (from 2010 data) of objects not associated with SFRs (black dashed-dotted line) and the same distribution separated for LPVs (solid blue line), EBs and known objects (solid orange line) and other classes of variable stars (solid red line). (bottom) $(J - H)$, $(H - K_s)$ colour–colour diagram for LPVs (blue circles), EBs and known objects (orange circles) and other classes of variable stars (red circles). In the diagram, arrows mark lower limits. The solid curve at the lower left indicates the colours of main-sequence stars. The short dashed line is the CTTS locus of Meyer, Calvet & Hillenbrand (1997) and the long dashed lines are the reddening vectors.

The objects belonging to different classes show very different properties. Fig. 10 shows the K_s distribution for these objects, where it can be seen that the LPVs dominate the bright end of the distribution with a peak at $K_s \sim 11.8$ mag and showing a sharp drop at brighter magnitudes, probably due to the effects of saturation. EBs and other classes are usually found at fainter magnitudes. The near-IR colours of the two samples (bottom plot of Fig. 10) show that LPVs tend to be highly reddened objects or have larger $(H - K_s)$ colours than EBs and other classes, which usually have the colours of lightly reddened main sequence and giant stars. We will see in Section 4 that this low reddening and lack of K-band excess (in most cases) distinguishes the EBs and other shorter period variables from the sample spatially associated with SFRs, so contamination of the SFR sample by these shorter period variables should not be very significant.

The colour–colour diagram of Fig. 10 also supports the idea that this sample might contain some bona fide YSOs that are not revealed by our searches of SIMBAD and the *WISE* images, as mentioned in our discussion of contamination. In the figure, we observe objects (red circles) that show $(H - K_s)$ colour excesses and are neither known variables nor classified as LPVs or EBs. By simply selecting red circles located to the right of the reddening vector passing through the reddest main-sequence stars, we estimate that 44 objects have colours consistent with a YSO nature. This would represent 15 per cent of objects outside SFRs and 5 per cent of the full sample of 816 variables. A more detailed study would of course be needed to confirm their nature as YSOs. We also note that the lower left part of the classical T Tauri stars (CTTS) locus plotted in the figure extends into the region occupied by lightly reddened main-sequence stars so, in principle, some of these individual red circles can potentially be YSOs.

The bright LPVs are very likely pulsating asymptotic giant branch (AGB) stars. These type of stars are usually divided into Mira variables, which are characterized by displaying variability of $\Delta K > 0.4$ mag and with periods in the range $100 < P < 400$ d and dust-enshrouded AGB stars, which are heavily obscured in the optical due to the thick circum-stellar envelopes (CSE) developed by heavy mass-loss ($\sim 10^{-4}$ M$_\odot$ yr^{-1}). The latter group, comprised of carbon-rich and oxygen-rich stars (the latter often referred to as OH/IR stars if they display OH maser emission), show larger amplitudes in the K band (up to 4 mag) and have periods in the range $400 < P < 2000$ d (for the above, see e.g. Jiménez-Esteban et al. 2006b; Whitelock, Feast & van Leeuwen 2008).

AGB stars are bright objects and should be saturated at the magnitudes covered in VVV. However, due to the large extinctions towards the Galactic mid-plane, we are more likely to observe these type of objects in VVV compared to our previous UKIDSS study. We can estimate the apparent magnitude of Mira variables at the different Galactic longitudes covered in VVV and by assuming that these objects are located at the Galactic disc edge ($R_{GC} = 14$ kpc; Minniti et al. 2011) and then considering other Galactocentric radii. At a given longitude, l, we derive A_V as the mean value of the interstellar extinction found at latitudes b between -1° and 1°. The interstellar extinction is taken from the Schlegel, Finkbeiner & Davis (1998) reddening maps and corrected following Schlafly & Finkbeiner (2011), i.e. $E(B - V) = 0.86E(B - V)_{\rm Schlegel}$. We then assume that extinction increases linearly with distance, at a rate $A_V/D_{\rm edge}$ (mag kpc^{-1}), with $D_{\rm edge}$ the distance to the Galactic disc edge at the corresponding l. We finally take the absolute magnitude as $M_K = -7.25$ (Whitelock et al. 2008).

Fig. 11 shows the estimated apparent magnitudes of Mira variables at different l and at varying R_{GC}. In the figure, we also show the magnitude, $K_s = 11.5$, that marks the drop in the number of the detection of these objects, as observed in the histograms of the K_s distribution (Fig. 10). We can see as we move away from the Galactic Centre, a Mira variable would most likely saturate in VVV, especially at $l < 310^\circ$. This occurs due to the effects of having relatively larger extinctions towards the Galactic Centre (see bottom plot of Fig. 11) and that a star at $R_{GC} = 14$ kpc is located farther away from the observer as l approaches $l = 0^\circ$. We note that Ishihara et al. (2011) finds that most AGB stars are found at $R_{GC} < 10$ kpc, so if we place a Mira variable at smaller Galactic radii ($R_{GC} = 7$, 10 kpc), we see that it is less likely for such a star to show up in our sample. However, variable dust-enshrouded AGB stars, which undergo heavy mass-loss, suffer heavy extinction due to their thick CSE and thus are fainter than Mira variables (AGB stars with optically thick envelopes are found to be ~ 5 K_s magnitudes fainter than objects with optically thin envelopes in the work of Jiménez-Esteban et al. 2006a) and thus less likely to saturate in VVV, even at large distances. Then, most AGB stars in our sample are probably dust-enshrouded objects.

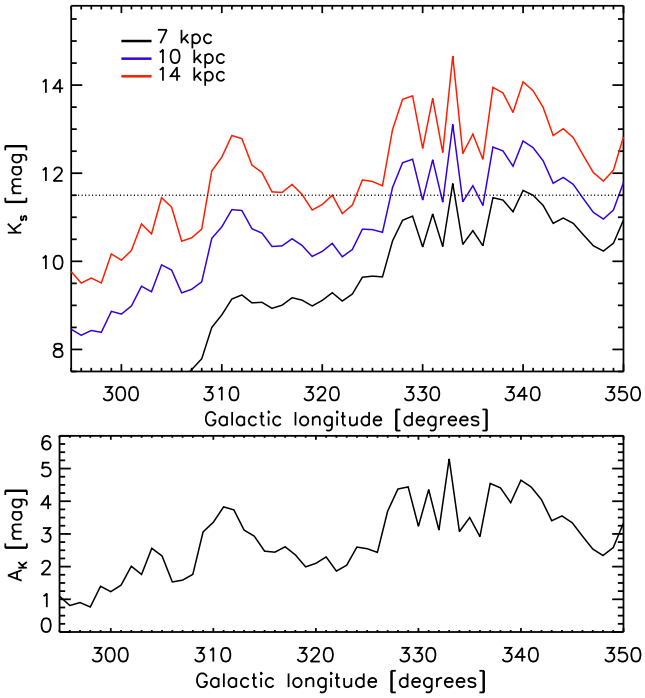

Figure 11. (top) Apparent K_s magnitude, derived as explained in the text, for a Mira variable located at the Galactic disc edge (solid red line). The same value is shown for a Mira located at lower Galactic radii $R_{GC} = 10$ (blue line) and 7 kpc (black line). The magnitude where the number of detections for LPVs drops ($K_s = 11.5$ mag) is marked by a dotted line. (bottom) K-band Galactic extinction column as a function of Galactic longitude.

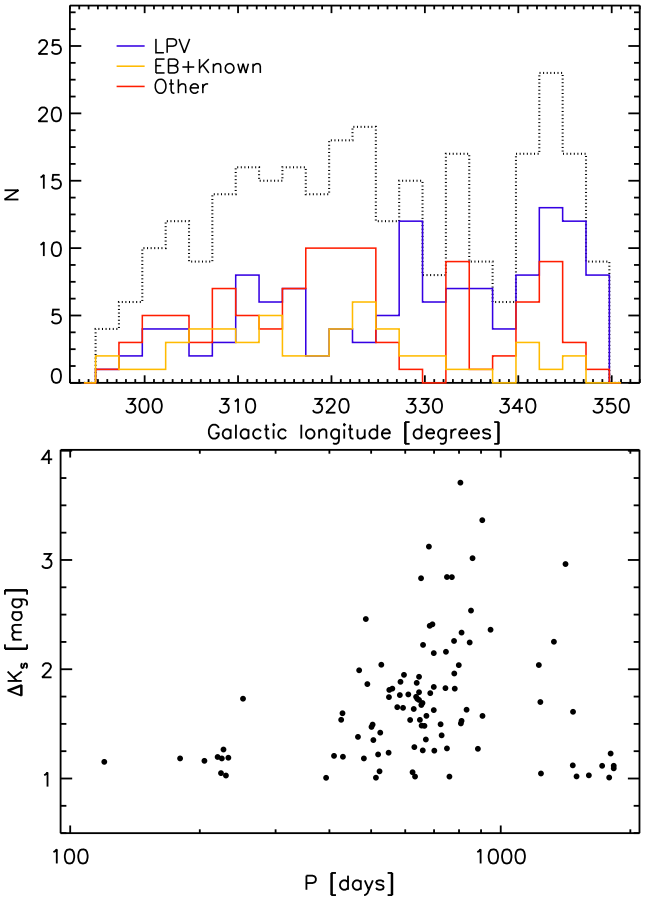

Figure 12. (top) Overall Galactic longitude distribution of objects outside SFRs (black dashed-dotted line). We also show the same distribution divided into LPVs (blue line), EBs and known objects (orange line) and other classes of variable stars (red line). (bottom) Period versus K_s amplitude for LPVs with a measured period.

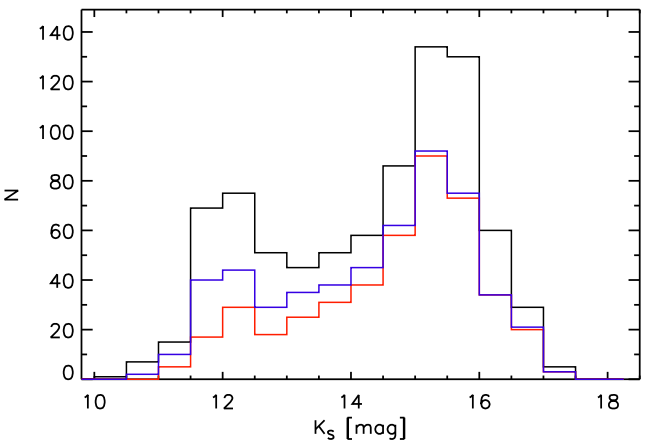

Figure 13. K_s distribution (from 2010 data) of the 816 VVV selected variable stars (solid black line). We show the same distribution for the overall sample of stars found to be likely associated with areas of star formation (blue line) and for the same sample but removing objects that show Mira-like light curves in Section 4.1 (red line).

The observations confirm the trend expected from the analysis above. Fig. 12 shows that the number of LPVs increases as we come closer to the Galactic Centre. In addition, when taking into account AGB stars with measured periods, we confirm that the majority of AGB stars show periods longer than 400 d and large amplitudes (see lower panel in Fig. 12), as expected in heavily obscured AGB stars. It is interesting to see in the same figure that variable objects with periods longer than 1500 d show lower amplitudes than expected for their long periods. This is similar to the observed trend in the variable OH/IR stars of Jiménez-Esteban et al. (2006b). According to the authors, these objects correspond to stars at the end of the AGB. We note that the apparent lack of high-amplitude objects at longer periods could relate to the fact that more luminous (longer period) objects have smaller amplitudes expressed in magnitudes (red supergiants often display $\Delta K < 1$ mag; see e.g. van Loon et al. 2008).

This population of bright pulsating AGB stars can also explain the observed bimodality of the K_s distribution for the full sample of VVV high-amplitude variable stars (see Fig. 13). The peaks of the distribution occur at $K_s \sim 11.8$ and 15.8. The peak at the bright end is at the same magnitude as the peak for LPVs.

When we only plot objects which are found to be likely associated with areas of star formation, the peak at the bright end becomes less evident (see blue histogram in Fig. 13). When we plot only SFR-associated sources that do not have LPV-like light curves (see Section 4.1) the bimodality almost disappears, as shown by the red histogram in Fig. 13. AGB stars are probably the main source of contamination in our search for eruptive YSOs in SFRs, especially dust-enshrouded AGB stars which can have IR colours resembling those of YSOs. Hence, it is fortunate that we can remove most of this contamination by selecting against LPV-like light curves.

3.6 YSO source density

Our finding in Section 3.3 that YSOs constitute about half of the detected population of high-amplitude variables in VVV disc fields indicates that they represent the largest single population of high-amplitude IR variables in the Galactic mid-plane, at least in the range $11 < K < 17$. We note that extragalactic studies of high-amplitude stellar variables are dominated by the more luminous AGB star population (see e.g. Javadi et al. 2015). Our analysis only considered the VVV disc tiles with $|b| < 1^{\circ}$, this amounts to 76 tiles, covering each 1.636 deg^2 of the sky. After allowing for the small overlaps between adjacent tiles, the total area covered in this part of the survey is 119.21 deg^2. Adopting a 50 per cent YSO fraction for the full sample of 816 variables implies a source density of 3.4 deg^{-2}. As noted in Section 2.1, the stringent data quality cuts in our selection procedure excluded ∼50 per cent of high-amplitude variables down to $K_s = 15.5$ and a high fraction at fainter magnitudes where completeness falls (see Fig. 13). The corrected source density is therefore ∼7 deg^{-2}.

When considering the source density of high-amplitude IR variables in the UGPS, Contreras Peña et al. (2014) argued that the observed source density underestimates the actual source density due to three main effects: (1) with only two epochs of K-band data, most high-amplitude variables will be missed; (2) the source density rises towards the magnitude cut of $K < 16$, indicating that many low-luminosity PMS variables, which are detected at distances of 1.4–2 kpc, would be missed at larger distances and (3) the data set used in the UGPS search of Contreras Peña et al. excludes the mid-plane and is therefore strongly biased against SFRs. The UGPS YSO source density is estimated to reach 12.7 deg^{-2} when correcting for these three factors. In the case of VVV, given the higher number of epochs obtained from this survey and that this analysis is not biased against areas of star formation, the source density is only likely to be affected by item (2) of the UGPS analysis. Fig. 13 shows the magnitude distribution of the VVV variables associated with SFRs, where we can see a similar behaviour to the UGPS results, with the density of sources rising steeply towards faint magnitudes. Contrary to the UGPS search, we do not have a hard magnitude cut in the VVV sample, which includes sources as faint as $K_s \sim 17$. However, the number of sources decreases at $K_s > 16$, so we estimate an effective magnitude detection limit of $K_s = 16.25$ mag. This implies that if typical sources from VVV have similar characteristics to UGPS objects in Cygnus and Serpens ($K = 14.8$, $d = 1.4$–2 kpc), then we would not detect them at distances $d > 3.32$ kpc. The complete sample of star-forming complexes from Russeil (2003) shows that 83 per cent of them are located beyond these distances. Correcting for this factor, we then estimate a true source density of 41 deg^{-2}, though this figure does not include YSOs with low mass and luminosity that are too faint to be sampled by VVV due to the absence of nearby SFRs in the survey area. This figure of 41 deg^{-2} is three times larger than the one estimated from the UGPS analysis of Contreras Peña et al. (2014) (12.7 deg^{-2}).

Two effects can account for the larger source density in VVV than UGPS. (1) High-luminosity YSOs are less common, but they can be observed at larger distances. The UGPS study would not find such objects at large distances because the available data set did not cover the mid-plane of the Galactic disc, in which all distant SFRs are located due to their small scaleheight. Since VVV does cover the mid-plane, we are able to detect these rare higher luminosity YSOs. This seems to be supported by the slightly larger distances established for members of the spectroscopic subsample in Paper II. (2) In the UGPS study, most (23/29) of the variables in SFRs were located in just two large SFRs: the Serpens OB2 association and portions of Cygnus X. The much smaller size of the UGPS sample (in number of SFRs and number of variables) meant that there was considerable statistical uncertainty in the area-averaged source density. Moreover, the incidence of high-amplitude variability is greater at the earlier stages of YSO evolution (see Section 4.3), so the numbers in the UGPS study may have been reduced by a relative lack of YSOs at these stages in the two large SFRs surveyed.

The estimated highly variable YSO source density remains much larger than that estimated for Mira variables in Contreras Peña et al. (2014), indicating a higher average space density for the variable YSOs. The observed variables in SFRs also outnumber the EBs and unclassified variables in the magnitude range of this study. However, we are likely to miss a large part of the population of high-amplitude EBs due to the sparse time sampling of VVV.

In Appendix A, we attempt to calculate the source density and space density of high-amplitude EBs from the OGLE-III Galactic disc sample of Pietrukowicz et al. (2013). In this, we are aided by a recent analysis of the physical properties of the large sample of *Kepler* EBs (Armstrong et al. 2014), which indicates that EBs with high amplitudes in the VVV K_s and OGLE I passbands are dominated by systems with F- to G-type primaries. We use simple calculations to show that while EBs can have very high amplitudes at optical wavelengths, the eclipse depth should not exceed 1.6 mag in K_s. Similarly, we find that eclipse depths should not exceed 3 mag in I. These results are supported by the VVV and OGLE-III data sets (Pietrukowicz et al. 2013) in which the distribution of EB amplitudes falls to zero by these limits. YSOs with $\Delta K_s > 1.6$ mag are very numerous in our sample so we conclude that high-amplitude YSOs greatly outnumber EBs above this limit. Below $\Delta K_s = 1.6$ mag, it is harder to reach a firm conclusion (see Appendix A). The space densities of EBs and those YSOs massive enough to be sampled by VVV may be comparable at $1 < \Delta K_s < 1.6$ mag. High-amplitude YSOs are likely to be more numerous if the variability extends down to the peak of the initial mass function at low masses, given that high-amplitude EB systems contain a giant with mass of order 1 $M_{\odot}$.

4 ANALYSIS OF VARIABLES IN STAR FORMATION REGIONS

The results presented here concern YSO variability in the K_s band-pass. At these wavelengths, variability in typical YSOs is produced by physical mechanisms (or a combination of them) affecting the stellar photosphere, the star–disc interface, the inner edge of the dust disc as well as spatial scales beyond 1 au (see e.g. Rice et al. 2015). These mechanisms include cold or hot spots on the stellar photosphere (e.g. Scholz 2012), changes in disc parameters such as the location of the inner disc boundary, variable disc inclination and changes in the accretion rate (as shown by Meyer et al. 1997). Variable extinction along the line of sight can also be responsible for the observed changes in the brightness of YSOs. Dust clumps that screen the stellar light have been invoked to explain the variability observed in Herbig Ae/Be stars and early-type CTTS (group also known as UX Ori stars; see e.g. Herbst & Shevchenko 1999; Eiroa et al. 2002). In other scenarios, variable extinction can be produced by a warped inner disc dust that is being uplifted at larger radii by a centrifugally driven wind, azimuthal disc asymmetry produced by the interaction of a planetary mass companion embedded within the disc or by occultations in a binary system with a circum-binary disc (see e.g. Romanova et al. 2013; Bans & Königl 2012; Bouvier

et al. 2013; Windemuth & Herbst 2014, and references therein). Finally, sudden and abrupt increases in the accretion rate (of up to three orders of magnitude) explain the large changes observed in eruptive variable YSOs. The variability in these systems traces processes occurring at the inner disc (in EXors; see Lorenzetti et al. 2012) or at larger spatial scales beyond 1 au, such as instabilities leading to outbursts events (in FUors; see e.g. Audard et al. 2014).

The amplitude of the variability induced by most of these mechanisms is not expected to be larger than $\Delta K \sim 1$ mag. Table 6 of Wolk, Rice & Aspin (2013) shows the expected amplitude of the K-band variability that would be produced by these different mechanisms. Cold spots and hot spots and changes in the size of the inner disc hole are not expected to show ΔK larger than 0.75 mag. We do note that the variability produced by hot spots from accretion depends on the temperature of the spot and the percentage of the photosphere that is covered by such spots, thus sufficiently hot spots can produce larger changes in the magnitude of the system. The range in ΔK from variable extinction is effectively limitless as it depends on the amount of dust that obscures the star. Large changes ($\Delta K > 1$ mag) have been observed from variable extinction in YSOs, e.g. AA Tau, V582 Mon (Bouvier et al. 2013; Windemuth & Herbst 2014). Nevertheless, variable extinction can be inferred from colour variability (see e.g. Section 4.2). Wolk et al. (2013) also estimate that a change in the accretion rate of a class II object of $\log \dot{M}(M_{\odot}\ yr^{-1})$ from -8.5 to -7 yields $\Delta K \sim 0.75$ mag. Thus, larger changes as observed in eruptive variables will produce large amplitudes.

Given all of the above, it is reasonable to expect variability in our YSO sample to be dominated by accretion-related variability and/or events of obscuration by circum-stellar dust.

4.1 Light-curve morphologies

We have visually inspected the light curves of our 530 SFR-associated variables in order to gain insight into the physical mechanism causing the brightness variations. In addition, we used PDM in IRAF and LOMBSCARGLEFAST in GATSPY to search for periodicity in the light curves of our objects. We stress that this is a simple and preliminary classification that is highly influenced by the sparse sampling of VVV. A more detailed study is planned in future, with improved precision by applying the differential photometry method of Huckvale, Kerins & Sale (2014) to the VVV images. We have divided the morphologies in the following classifications.

(i) Long-term periodic variables: defined as objects showing periodic variability with $P > 100$ d. This limit is adopted for consistency with the limit used in the analysis of objects outside of SFRs, with the benefit that contamination by long-period AGB stars will be confined to this group. We measure periods for most of these objects, albeit with some difficulty in phase-folding the data in many of them. In this subsample, we find 154 stars, representing 29 per cent of objects spatially associated with SFRs. In Section 3.4, we contended that field high-amplitude IR variables with periodic light curves ($P > 100$ d) are very likely dust-enshrouded AGB stars, these being identifiable by their smooth, approximately sinusoidal light curves. We estimated that ∼27 per cent of the 530 SFR-associated variables would be non-YSOs and up to 45 per cent of these would be LPVs, implying that this subsample may contain ∼64 dusty Mira variables. Visual inspection of the 154 light curves indicates that while some have a smooth sinusoidal morphology (after allowing for long-term trends due to variable extinction in the expanding circum-stellar dust shell), others display short time-scale scatter superimposed on the high-amplitude long-term periodicity. In Fig. 14, we show the examples of objects VVVv309 and VVVv411. The short time-scale variability in their light curves is definitely not consistent with the typical light curves of Mira variables and their periods of 143.95 and 190.6 d, respectively, are shorter than those of the dusty Miras detected outside SFRs (see Fig. 12). This short time-scale scatter is typical of the scatter observed in normal YSOs due to a combination of hot spots, cold spots and small variations in accretion rate or extinction, so it is reasonable to think that most of the LPVs with short time-scale scatter are in fact YSOs.

To support this interpretation, Fig. 15 shows the period versus ΔK_s and period versus K_s distributions, for objects where we are able to measure a period. The period versus ΔK_s distribution is similar to that observed in LPVs outside SFRs (Fig. 12) except that there is a larger number of 'long-term' periodic variables found with periods, $100 < P < 350$ d. The 65 blue points are the objects with short time-scale scatter and the red points are the remaining 89, categorized by careful inspection of the light curves of the 154 long-term periodic variables in SFRs. The blue points clearly dominate the group with $P < 350$ d and they also have a distinctly fainter distribution of K_s magnitudes, similar to that shown in the red histogram in Fig. 13, which represents all objects in SFRs except those with Mira-like light curves. As expected, the red points with Mira-like light curves typically have $K_s \sim 12$, similar to the LPV distribution plotted in Fig. 10. We conclude that inspection of the light curves can separate the evolved star population of LPVs from the YSOs in SFRs with fair success, though we caution that this is an imperfect and somewhat subjective process that can be influenced by outlying data points and our limited knowledge of the time domain behaviour of circum-stellar extinction in dusty Mira systems. The limitations are demonstrated by the presence of a number of blue points with $K_s \sim 12$ and $P > 350$ d in the lower panel of Fig. 15 and a hint of bimodality even in the 'decontaminated' magnitude distribution in Fig. 13 (red histogram). In the subsequent discussion of YSOs from our sample, we only include the 65 objects with short time-scale scatter (called LPV-yso) and assume that the other 89 sources are dusty AGB stars or other types of evolved star (or LPV-Mira).

This decontamination of AGB stars reduces the SFR-associated sample to 441 objects. The long-term periodic YSOs represent 15 per cent of this sample.

Periods, $P > 15$ d, are longer than the stellar rotation period of YSOs or the orbital period of their inner discs (Rice et al. 2015). Some YSOs have been observed to show variability with periods longer than even 100 d. WL 4 in ρ Oph shows periodic variability with $P = 130.87$ d (Plavchan et al. 2008), which can be explained by obscuration of the components of a binary system by a circum-binary disc. The K-band amplitude of the variability in that system is somewhat less than 1 mag. However, it is possible to think that a similar mechanism might be responsible for the variations in some of our objects. Hodapp et al. (2012) show that variable star V371 Ser, a class I object driving an H_2 outflow, has a periodic light curve with $P = 543$ d. The authors suggest that variability arises from variable accretion modulated by a binary companion. In view of this, the variability in some of the long-term periodic variables might be driven by accretion and we discuss this in Paper II, based on spectroscopic evidence for a subsample of them.

(ii) Short-term variability: this group comprises objects that either have periodic variability and measured periods, $P < 100$ d, (75 objects) or else have light curves that appear to vary continuously over short time-scales ($t < 100$ d) but not with an apparent period (87 objects). Their light curves do not resemble those of detached

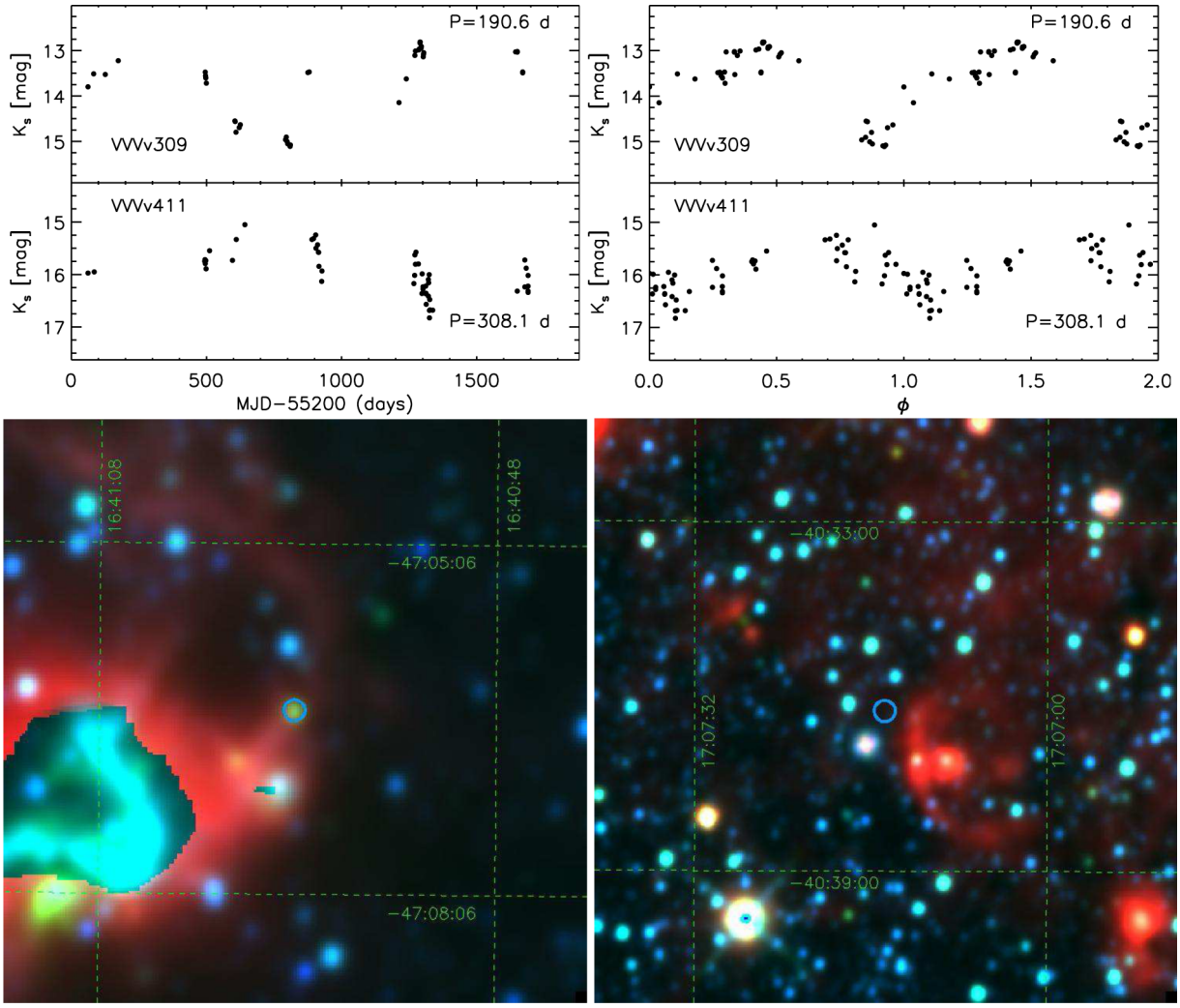

Figure 14. (top left) Examples of K_s light curves of the LPVs VVVv309 and VVVv411, which are found in areas of star formation. (top right) Phased light curves for the same objects. (bottom) 10 arcmin × 10 arcmin *WISE* false colour images (blue = 3.5 μm, green = 4.6 μm, red = 12 μm) centred on VVVv309 (left) and VVVv411 (right). In both images, the location of the variable star is marked by a ring around the location of the object. VVVv309 is 114 arcsec from H II region GRS G337.90 −00.50 (see e.g. Culverhouse et al. 2011). The 12 μm *WISE* image of VVVv309 saturates at the centre of the H II region creating the blue/green 'inset' in the false colour image. VVVv411 is located 104 arcsec from the IR bubble [CPA2006] S10 (Simpson et al. 2012) as well as other indicators of ongoing star formation.

EBs because they vary continuously and cannot be contact binaries (W UMa variables) as their periods are typically longer than 1 d. For objects in this classification that have measured periods, we observe a broad distribution from 1 to 100 d and the amplitudes are in the range $\Delta K_s = 1$–2 mag (see Fig. 16). If we join together the long-term periodic variables and the short-term variables (STVs) with measured periods, we find that sources with periods $P > 100$ d show higher amplitudes, on average, and sources with $P > 600$ d have redder SEDs (larger values of the spectral index α). There are no clear gaps in the period distribution, so the 100 d division between the two groups that we adopted to aid decontamination is arbitrary. We find 162 stars in the STV group, which represents 37 per cent of the decontaminated SFR-associated sample.

High-amplitude periodic variability has been observed in YSOs over a wide range of periods. RWA1 and RWA26 in Cygnus OB7 (Wolk et al. 2013) vary with periods of 9.11 and 5.8 d, respectively. The variability has been explained as arising from extinction and inner disc changes. As mentioned before, variability with $P > 15$ d is not expected to arise from the stellar photosphere or changes in the inner disc of YSOs. This instead could be related to obscuration events from a circum-binary disc, such as in V582 Mon (Windemuth & Herbst 2014) and YSOs ONCvar 149 and 479 in Rice et al. (2015). Variable accretion has been invoked to explain the observed periodic variability ($P \sim 30$ d) of L1634 IRS7 (Hodapp & Chini 2015). The shorter periods within this group may indicate rotational modulation by spots in objects with amplitudes not far above 1 mag.

(iii) Aperiodic long-term variability: this category can be divided into three different subclasses: (a) Faders: here the light curves show a continuous decline in brightness or show a constant magnitude for the first epochs followed by a sudden drop in brightness that lasts for a long time (≥1 yr), continuing until the end of the time series in 2014. This type of object might be related to either stars going back to quiescent states after an outburst or objects dominated by long-term extinction events similar to the long-lasting fading event in AA Tau (Bouvier et al. 2013) or some of the faders in Findeisen et al. (2013). (b) Objects that show long-lasting fading events and then return to their normal brightness (such as VVVv504 in Fig. 17), which we refer to as dippers. These might also be related

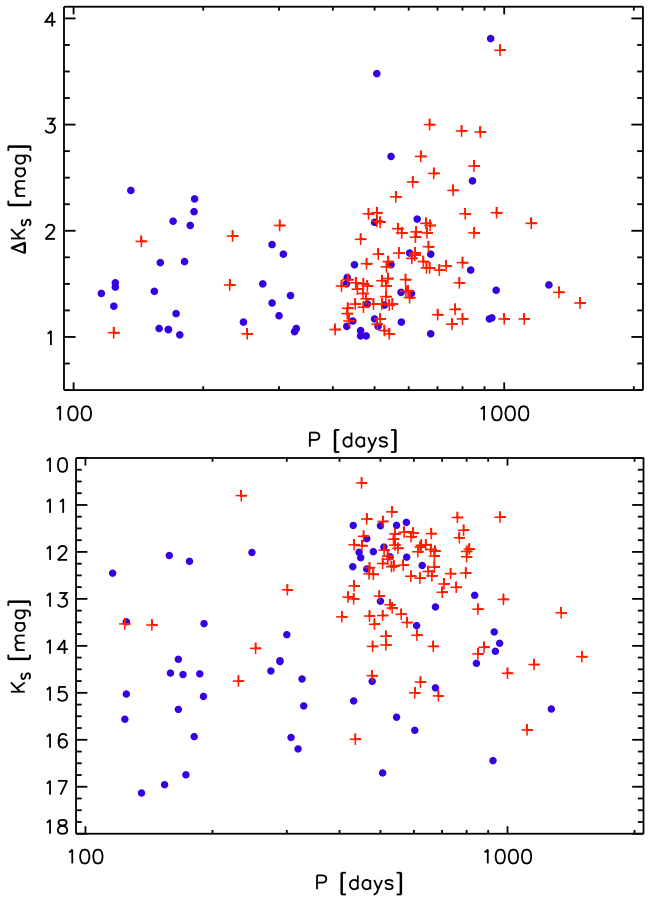

Figure 15. (top) K_s amplitude versus period for stars in SFRs with long-term periodic variability. LPVs that show Mira-like light curves are shown in red plus signs while other sources are shown as blue circles. (bottom) K_s magnitude (2010) versus period for the same sample of stars.

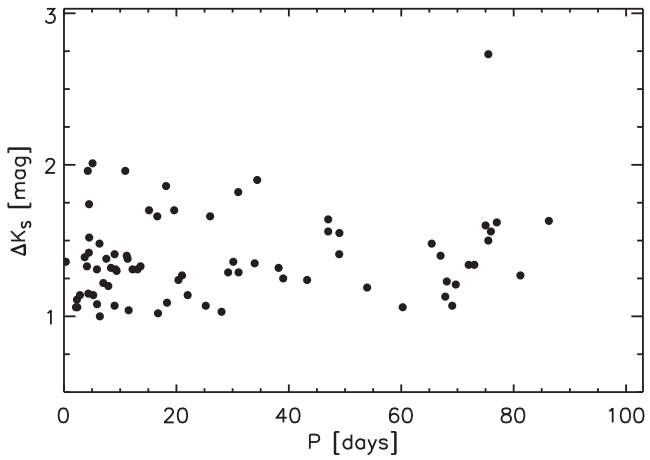

Figure 16. K_s amplitude versus period for stars with short-term variability and with a measured period.

to extinction events. Examples of objects in groups (a) and (b) can be seen in Fig. 17.

Group (c) contains sources with outbursts, typically of long duration ($\geq$1 yr). In a very small number of objects, the outburst duration appears to be much shorter, of the order of weeks. The increases in brightness are also unique or happen no more than twice during

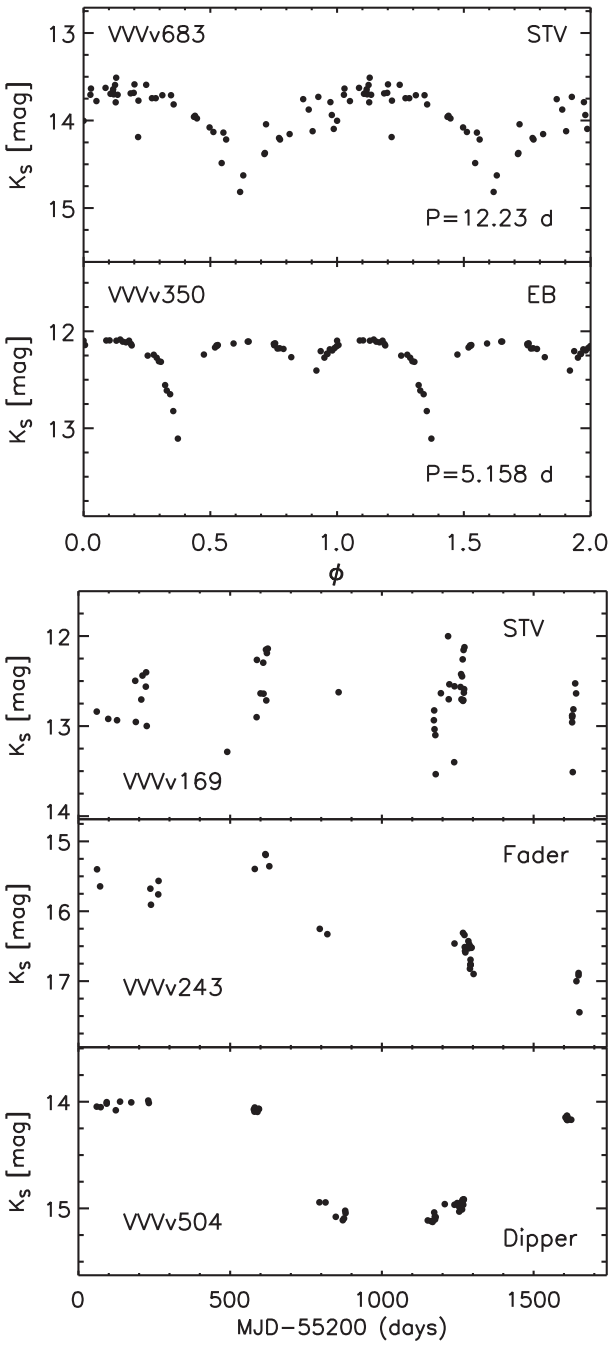

Figure 17. Examples of K_s light curves for the different classifications as explained in the text. (top) Phased light curves of STV star with a measured period, VVVv683 and EB VVVv350. (bottom) Light curves of STV star, without a measured period, VVVv169, the fader VVVv243 and the dipper VVVv504.

the light curve of the object, thus not resembling the light curves of objects in the STV category. An exception is VVVv118, which shows four brief rises on time-scales of weeks.

The light curves in this category typically have a monotonic rise of 1 mag or more, though sometimes a lower level scatter is present atop the rising trend. In a small number of cases, the rise in the light curve is poorly sampled, starting at or before the beginning of the time series, but the subsequent drop exceeds 1 mag. Fig. 18 shows four examples of objects falling in the eruptive classification. The examples have been selected in order to illustrate the different

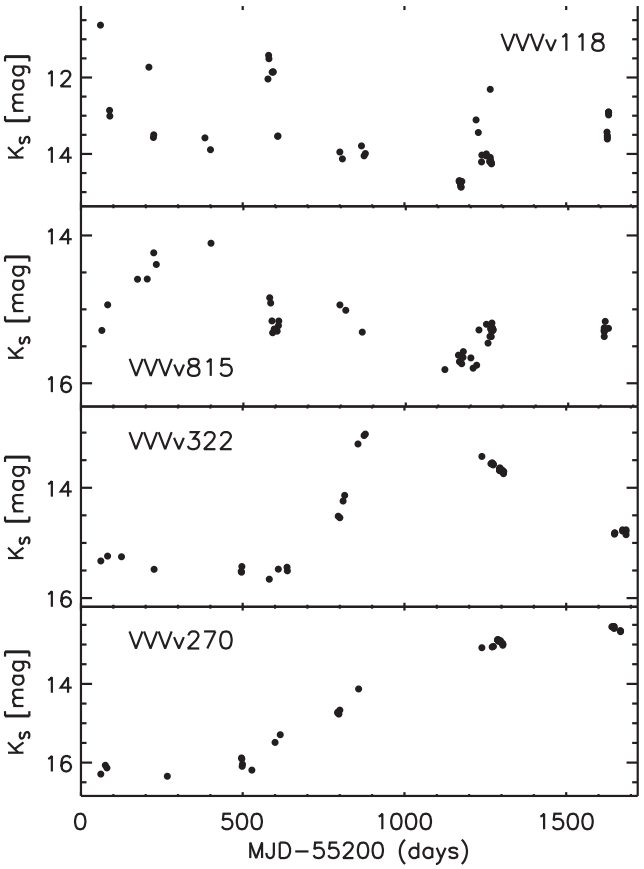

Figure 18. Examples of K_s light curves for different objects in the eruptive classification as explained in the text. From top to bottom, we show objects VVVv118, VVVv815, VVVv322 and VVVv270.

temporal behaviour observed in objects belonging to this class. As we have already mentioned, VVVv118 shows multiple short, high-amplitude rises. In general, objects show outbursts which last between 1 and 4 yr (see VVVv815 and VVVv322). We also detect a few cases where the outburst duration cannot be measured as it extends beyond 2014 data (e.g. VVVv270).

When comparing to the behaviour of known classes of eruptive variables, VVVv118 would resemble that of EXors and VVVv270 could potentially be an FUor object (based only on photometric data). However, most of the objects have outburst durations that are in between the expected duration for EXors and FUors.

Considering the outburst duration of the known subclasses of young eruptive variables, we are likely to miss detection of FUor outbursts if they went into outburst prior to 2010. In the case of EXors, which have outbursts that last from few weeks to several months, we would expect to detect more of these objects given the time baseline of VVV. However, our results show a lack of classical EXors, which could be a real feature or it could be related to the sparse VVV sampling. Thus, we need to test our sensitivity to short, EXor-like eruptions.

We simulate outbursts with time-scales from 2 months to 3 yr. First, we generate a very rough approximation of an eruptive light curve with outburst duration, T_o. The light curve consists of: (1) A quiescent phase of constant magnitude that lasts until the beginning of the outburst, which is set randomly at a point within the 2010–2012 period (between 0 and 1000 d). (2) A rise which is set arbitrarily to have a rate of 0.15 mag d^{-1}, lasting $T_{\rm rise}$ =10 d until reaching an outburst amplitude of 1.5 mag, which is a little below the median for the VVV eruptive variable candidates. (3) Plateau phase with a constant magnitude set to the peak of the outburst. This phase lasts for $T_o - T_{\rm rise} - T_{\rm decline} = T_o - 20$ d. (4) The decline, which lasts 10 d and has a rate of 0.15 mag d^{-1}. Finally, (5) a second quiescence phase. To every point in the light curve, we add a randomly generated scatter of ±0.2 mag. Once the light curve is generated, we measure the magnitude of the synthetic object at the observation dates of a particular VVV tile. If the synthetic object shows $\Delta K_s \geq 1$ mag, then it is marked as a detection. This procedure is repeated 1000 times for each outburst duration (which is set to be between 30 and 900 d). We also repeat this procedure for four different VVV tiles.

The simulation shows that the number of detections is very similar (∼80 per cent) for $T_o > 7$ months and declines slowly as T_o is reduced, falling by a factor of 2 for $T_o = 2$ months. However, this is not enough to cause the apparent lack of eruptive variables with EXor-like outbursts in our sample. We conclude that the longer (1–4 yr) durations that we observe are the typical values for IR eruptive variables, rather than a sampling effect.

The characteristics of our eruptive sample (see Paper II) agree with recent discoveries of eruptive variables that show a mixture of characteristics between the known subclasses of eruptive variables (see e.g. Aspin, Greene & Reipurth 2009). We note that classification of our sample into the known subclasses becomes even more problematic when taking spectroscopic characteristics into account, as e.g. VVVv270, the potential FUor from its light curve, shows an emission line spectrum or VVVv322 shows a classical FUor near-IR spectrum. In Paper II, we propose a new class of eruptive variable to describe these intermediate eruptive YSOs.

Sources classified as eruptive are very likely to be eruptive variables, where the changes are explained by an increase of the accretion rate on to the star due to instabilities in the disc of YSOs (see e.g. Audard et al. 2014). We find 39 objects in subgroup (a), 45 in (b) and 106 in (c). The whole class of faders/bursts represents 43 per cent of the likely YSO sample.

(iv) EBs: we find 24 objects with this light-curve morphology, representing 5 per cent of the sample. We are able to measure a possible period in 15 of them. The remaining nine objects are left with this classification given the resemblance of their light curves to the objects with measured periods. We expect that a number of them will be field EBs contaminating our YSO sample. However, inspection of the 2–23 μm spectral index for each object, α, (see Fig. 23), indicates that 12 objects are classified as either class II or flat-spectrum sources. If these are in fact YSOs, they would represent a significant discovery as YSO EBs are invaluable anchors for stellar evolutionary models, which generally lack empirical data on stellar radii. Fig. 19 shows the light curve and location of one candidate YSO EB, VVVv317, with P=6.85 d and $\alpha = -1.57$. The spectral index places it at the edge of the classification of class II YSOs.

In Fig. 20, we compare the amplitude distributions ($K_{\rm s,max} - K_{\rm s,min}$) of the different categories of variable YSO. We can see that the EBs and STVs typically have the smallest amplitudes (means of 1.18 and 1.33 mag, respectively), whereas the faders and eruptive variables have the highest amplitudes (means of 1.95 and 1.72 mag and medians of 1.75 and 1.61 mag, respectively). The dippers and long-term periodic variables have mean amplitudes of 1.64 and 1.57 mag, respectively, which are similar to the mean amplitude of 1.56 mag for the full sample. The substantial number of STVs with amplitudes only a little over 1 mag is consistent with

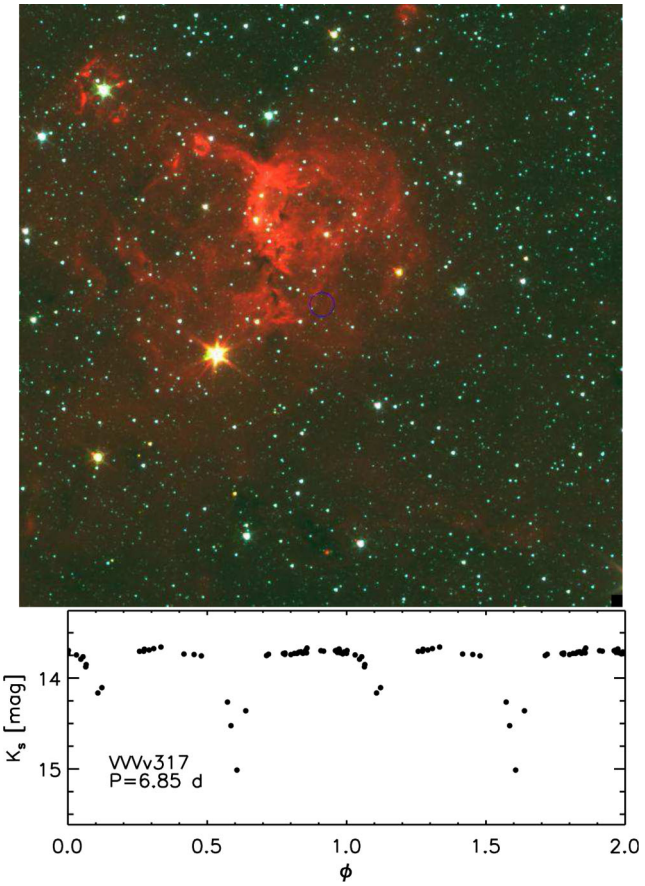

Figure 19. (top) False colour image (blue = 3.6 μm, green = 4.5 μm and red = 8 μm) from GLIMPSE, showing the location of candidate YSO EB VVVv317 (blue ring). (bottom) Phased light curve of the object.

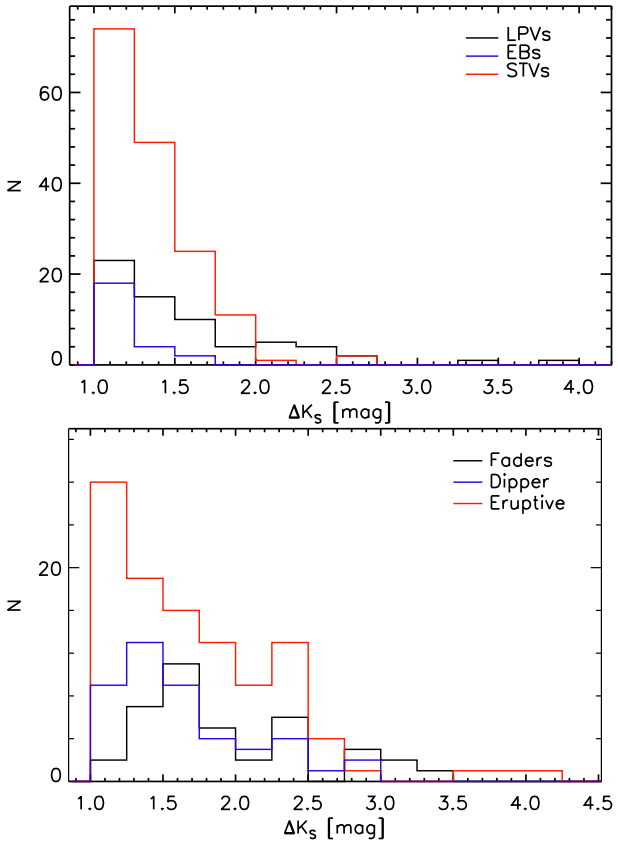

Figure 20. ΔK_s distribution for the different light-curve morphology YSO classes. (top) Distribution for LPVs (black line), STVs (red line) and EBs (blue line). (bottom) Distribution for aperiodic variables, faders (black line), dippers (blue line) and eruptive objects (red line).

our suggestion that in the shorter period variables in this category, variability may be explained by rotational modulation of dark or bright spots on the photosphere. The relatively high amplitudes of the eruptive variables are not unexpected, whilst the high amplitudes of the faders could be explained if some of these objects are eruptive variables returning to quiescent states.

4.2 Near-IR colour variability

Considering the various classes of light curve defined above for SFR-associated variables, it is reasonable to expect that extinction causes the variability in dippers and perhaps some of the STVs. Extinction variability can be observed in eruptive objects. However, we do not expect the main cause of variability in this class to be due to this mechanism. It is less clear what to expect for faders and long-term periodic variables. We can test this by looking at colour variability data.

The VVV survey was initially designed with only one epoch of contemporaneous JHK_s colour data but a second such epoch was added to the programme for observation in 2015, both to benefit the YSO variability science and to help to understand VVV variables of unknown nature. We note that many objects in the SFR sample are not detected in J and H band (see Section 4.3), making a colour comparison impossible. Also, in many objects, the two epochs do not span a large fraction of the full range of magnitudes in the light curve. For example, in many eruptive objects, we do not have direct comparison between quiescent and outburst states. Nevertheless, comparison of the change in colour versus magnitude still provides some valuable information on the mechanisms driving variability, particularly for those sources in which source magnitudes differ substantially at the two epochs.

Following a similar procedure to Lorenzetti et al. (2012) and Antoniucci et al. (2014), we compare the change in colour, $(H - K_s)$, versus magnitude, H, between 2010 and 2015 for the different classes of YSOs (see Fig. 21). In the figure, we see that in most cases, the changes in both colour and magnitude are small, thus objects cluster around the origin. This is especially true in EBs. The overall distribution of eruptive, dippers, LPVs and STVs, on the other hand, seems to be elongated along an axis passing through the 'bluer when brighter' and 'redder when fainter' quadrants. This agrees with the behaviour expected from changes due to accretion or extinction in YSOs and resembles the near-IR variability observed in EXors (Lorenzetti et al. 2009, 2012), the classical T Tauri sample of Lorenzetti et al. (2012) and the mid-IR variability of candidate EXors from Antoniucci et al. (2014). It is interesting to see that many objects classified as faders fall in the bluer when fading quadrant. This behaviour is still consistent with YSO variability due to changes in disc parameters described by Meyer et al. (1997) and observed in some Cygnus OB7 YSOs (see e.g. Wolk et al. 2013). The different behaviour might also be caused by the different geometry (inclination) of the system with respect to the observer. Scattering of light by the circum-steller disc or envelope may also be contributing to the sources that are bluer when fainter.

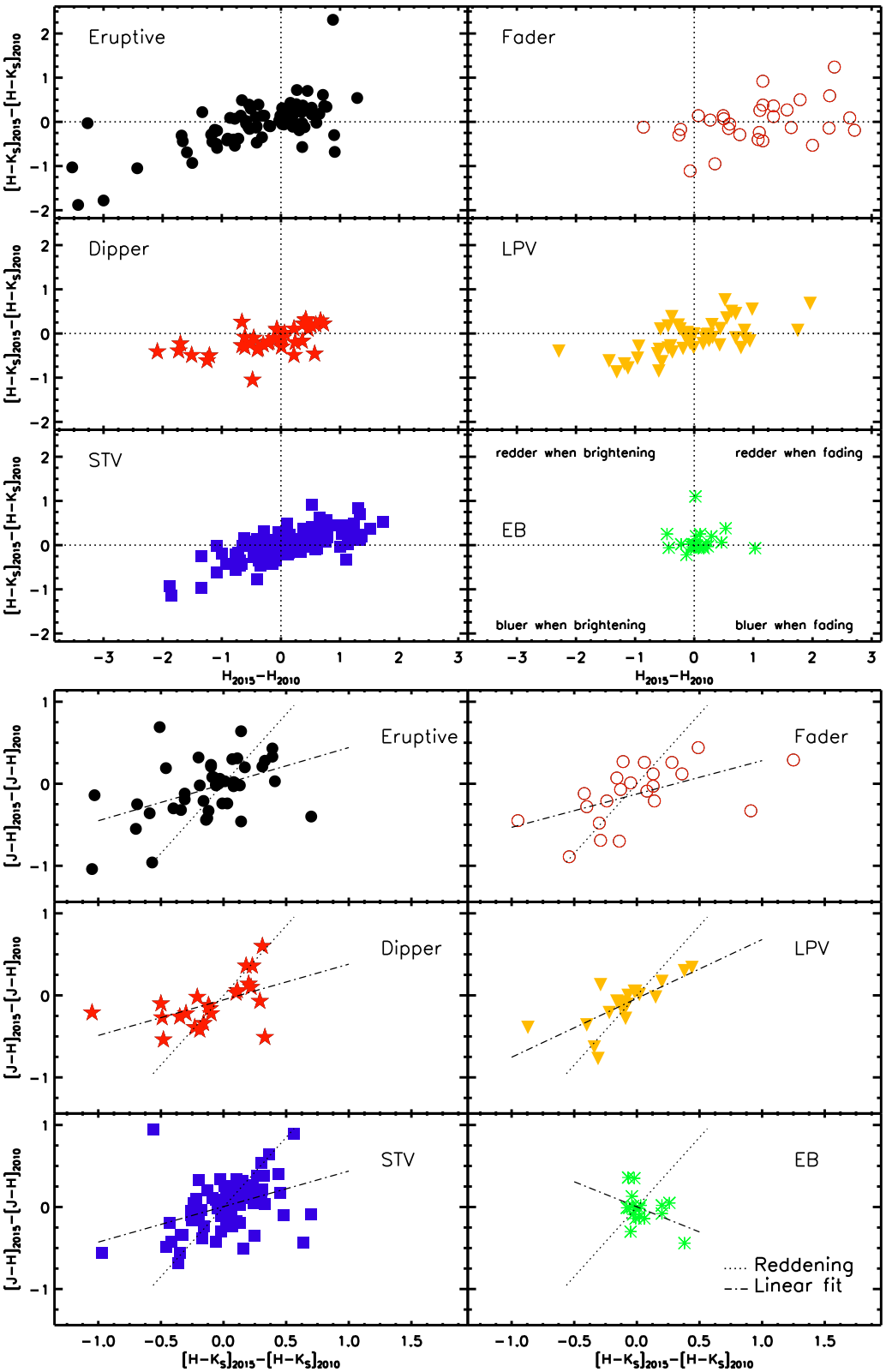

Figure 21. $\Delta(H-K_s)$ versus ΔH (top) and $\Delta(J-H)$ versus $\Delta(H-K_s)$ (bottom) for YSOs with an available second JHK_s epoch from VVV. In the plots, we mark the different classes from light-curve morphology. In the bottom plot, we mark the expected changes which occur parallel to the reddening vector (dotted line) as well as the best fit to the observed change (dot–dashed line). In the bottom right of the upper panel, we mark four distinct regions as explained in the text.

Fig. 21 also shows the observed change in both $(J - H)$ and $(H - K_s)$ colours for YSOs detected in the three filters in both epochs. There we also plot the expected change if the variability occurs parallel to the reddening line (independent of the direction of the change) as well as a linear fit to the different YSO classes. We would expect that variability similar to that observed in EXors (see e.g. fig. 1 in Lorenzetti et al. 2012) would show a behaviour that is not consistent with reddening. It is hard to say much from EBs as they do not show much variability. The overall change in STVs, faders and dippers appears to be different from the reddening path, although it appears to depend on the selection of objects from those samples. The path followed by LPVs is also different from the reddening line, but it does suggest a more similar behaviour to extinction compared to the other classes. Eruptive variables appear to follow a very different path from reddening. We would expect such behaviour if the variability is similar to that observed in EXors (see e.g. fig. 1 in Lorenzetti et al. 2012). Given that variable extinction and eruptive variability are the only known mechanisms to produce variability well in excess of 1 mag, the fact that colour variability disfavours the former in eruptive systems suggests that the latter is more likely.

We have checked the individual $(J - H)$ versus H, $(H - K_s)$ versus K_s and $(J - H)$ versus $(H - K_s)$ colour–magnitude (CMD) and colour–colour diagrams for the 15 eruptive variables that showed $\Delta K_s > 0.75$ mag between the two multiwavelength epochs (thus representing a significant fraction of the total amplitude in most systems). From 15 objects, 10 do not show changes consistent with extinction (e.g. they are bluer when fainter or show negligible colour change) and 5 were found to show variability approximately following the reddening vector. However, the colour behaviour in these five objects does not contradict the idea that accretion is the mechanism driving variability because: (1) we are not directly comparing quiescent versus outburst states, as the two near-IR epochs cover random points in the light curve; (2) As previously mentioned, extinction does play a role in outburst variability. For example, the near-IR colour variation of V1647 Ori follows the reddening path in fig. 13 of Aspin, Beck & Reipurth (2008). Extinction might also be involved in the observed variability or the recent eruptive object V899 Mon (Ninan et al. 2015).

The same analysis of individual CMD and colour–colour diagrams for the remaining classes shows that 3/7 LPV-YSOs, 5/9 STVs, 1/4 dippers and 3/11 faders have colour changes consistent with variable extinction (again considering only systems with $\Delta K_s > 0.75$ between the two multicolour epochs). No results can be derived from EBs as none of them show changes larger than 0.75 mag between the two epochs. It is interesting to see that the changes in the majority of faders are not consistent with extinction. This supports the idea that variability in many of the objects in this class could be related to accretion changes.

In Appendix B, we briefly summarize the colour and magnitude changes detected by the multi-epoch photometry from the *WISE* satellite. We note that this adds little to the preceding discussion of near-IR colour changes, though large mid-IR variability is observed in a minority of sources where the satellite happened to sample both a peak and a trough in the light curve.

4.3 Variability trends with SED class

In order to study the possible evolutionary stage of the variable stars in SFRs, we use the slope of the SEDs of the stars in the range $2 < \lambda < 24$ μm. Following Lada (1987), we define the parameter α as $\alpha = \mathrm{d}(\log(\lambda F_\lambda))/\mathrm{d}(\log(\lambda))$. The value of α is determined from a linear fit to SED points in the range $2 < \lambda < 24$ μm. Objects are then classified according to their value of α following Greene et al. (1994), also shown in Table 3. We note that this class definition might not necessarily relate to the actual evolutionary stage of the object. As stated in e.g. Robitaille et al. (2006), parameters such as inclination or stellar temperature can affect the shape of the SED at the wavelengths used to classify YSOs. We have removed numerous objects that showed Mira-like characteristics from their light curves but there may still be some contamination by non-YSOs amongst the remaining objects projected close to an SFR.

Table 3. Number of VVV variable stars belonging to the different evolutionary classes of YSOs, as determined from their SEDs.

Class	α	N	$N_{J\mathrm{drop}}$	$N_{JH\mathrm{drop}}$	$N_{\Delta K_s \geq 2}$
Class I	$\alpha > 0.3$	198	130	45	49
Flat	$-0.3 \leq \alpha \leq 0.3$	95	35	1	12
Class II	$-1.6 < \alpha < -0.3$	83	19	1	6
Class III	$\alpha \leq -1.6$	12	0	0	0
Undefined	n/a	53	10	2	3

We derive α using the photometry arising from VVV and *WISE*, given that these were taken in the same year (2010). When the objects are not detected in *WISE*, we use *Spitzer*/GLIMPSE photometry. The use of the latter is more likely to cause errors in the estimation of α due to the time difference between *Spitzer*/GLIMPSE and VVV measurements, but *Spitzer*/GLIMPSE benefits from higher spatial resolution. We find that if we use *Spitzer* instead of *WISE*, the difference in α shows a random offset of 0.1–0.2 for the majority of the sample detected in both surveys.

The number of objects belonging to different classes is shown in Table 3. We find that the majority (67 per cent) of objects in our sample are either class I or flat-spectrum sources (45 per cent and 22 per cent, respectively). Objects belonging to different classes show some differences in their global properties. Fig. 22 shows the near-IR colours of objects in SFRs. As expected, the vast majority of objects show colours consistent with them being YSOs. The $H - K_s$ colour tends to be redder for stars belonging to younger evolutionary stages. The fraction of objects detected at J and H bands also decreases with objects belonging to younger stages, as would be expected for typical deeply embedded class I objects. Table 3 shows that 66 per cent of class I objects are not detected in the J band, whilst 23 per cent of them are not detected in either the J nor H bands. These near-IR colour trends confirm that, as with typical YSOs, highly variable YSOs whose spectral index indicates an earlier evolutionary stage also have higher extinction by circumstellar matter, along with more IR emission from circum-stellar matter.

It might be thought that the relatively high reddening of most of the YSOs in Fig. 22 is due to foreground extinction given that in Paper II, we derive typical distances of a few kiloparsecs for these sources. We measure foreground extinction for a subsample of VVV objects (the 28 variable YSOs in Paper II) by estimating the extinction of red clump giants found at distances similar to those of our objects. The red clump giants are identified in the local K_s versus $(J - K_s)$ CMDs of the VVV objects (6 arcmin × 6 arcmin fields) and distances are estimated from their observed magnitudes K_s and mean $(J - K_s)$ colours using equation (1) in Minniti et al. (2011). The excess of the mean $(J - K_s)$ with respect to the intrinsic colour of red clump giants ($(J - K_s)_0 = 0.70$ mag; Minniti et al. 2011) gives us a measure of the foreground extinction to the front of the

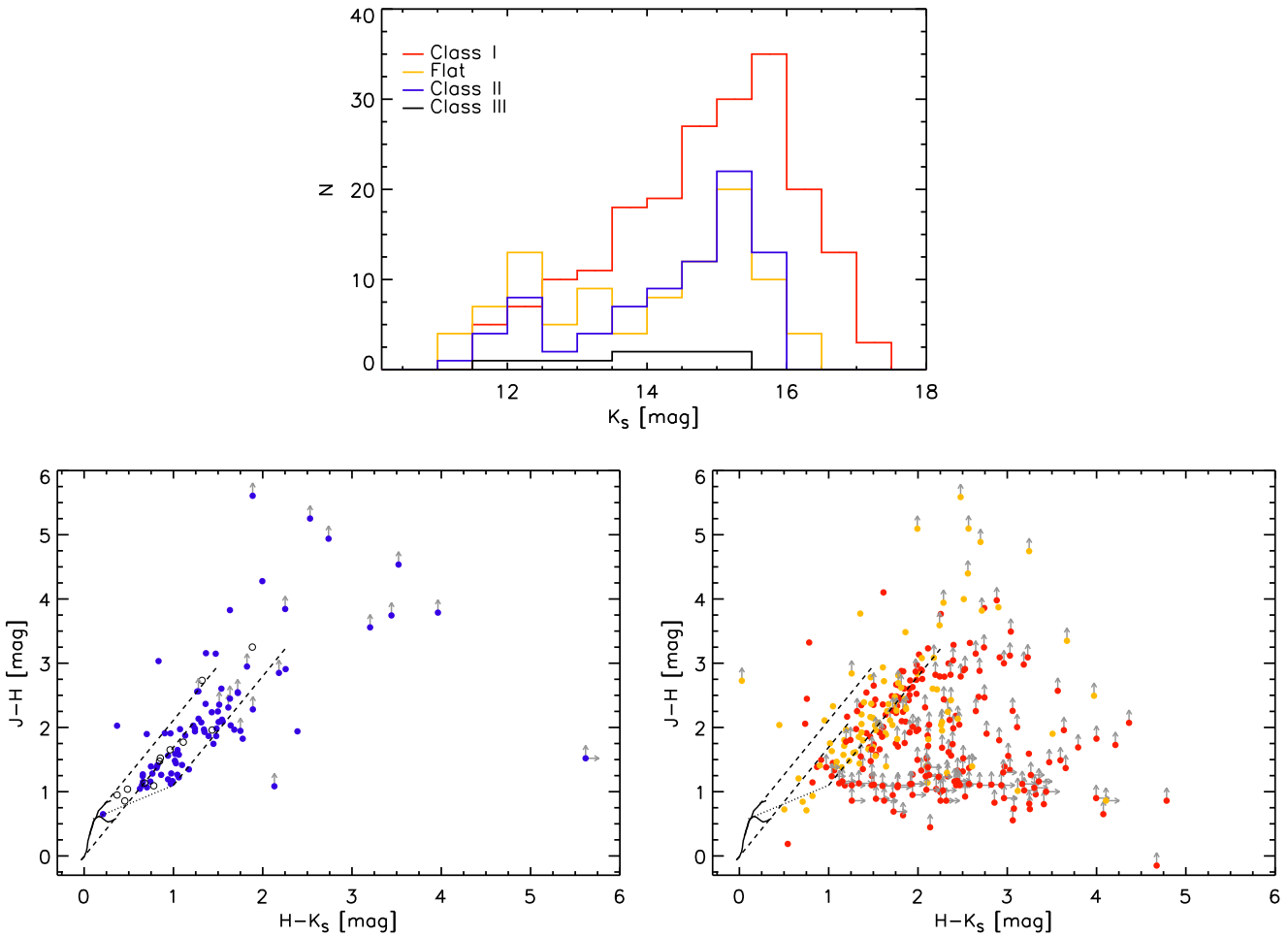

Figure 22. (top) K_s distribution (from 2010 data) of class I (red), flat-spectrum (orange), class II (blue) and class III (black) YSOs. (bottom left) Colour–colour diagram for class II (blue filled circles) and class III YSOs (black open circles) from VVV. In the figure, lower limits in colour are marked by arrows. The classical T Tauri locus of Meyer et al. (1997) is presented (long dashed line) along with intrinsic colours of dwarfs and giants (solid lines) from Bessell & Brett (1988). Reddening vectors of $A_V = 20$ mag are shown as dotted lines. (bottom right) Colour–colour diagram for class I (red) and flat-spectrum (orange) YSOs.

molecular cloud containing each YSO. In most cases, the YSO itself is redder than the red giant branch stars, due to extinction by matter within the cloud and by circum-stellar matter. This method yields values of $A_{K_s} \sim 0.8$–1.4 mag (i.e. $A_V \sim 7$–12 mag), which is higher than the typically very low diffuse interstellar extinction between the sun and nearby molecular clouds. We infer that the optical faintness and steeply rising near-IR SEDs of these objects are due not only to their early evolutionary stage but also to foreground extinction. Correcting our values of α accounting for $A_V = 11$ mag produces changes of 0.2–0.4 in this parameter. These changes would alter the classification of one-third of the sample where we can estimate α, mostly in objects where α has a value that is close to the limits set by Greene & Lada (1996). Despite this correction, flat-spectrum and class I sources still dominate the sample of YSOs with a measured value of α: with no correction, these represent 76 per cent of the sample, whereas with a correction to α of 0.3, the proportion remains high, at 59 per cent.

We choose not to apply a correction in Table 3 and our subsequent analysis because the extinctions and distances are uncertain (e.g. the red giant branch is not well defined in the CMDs for about a third of sources) and we cannot be sure that this method is entirely correct. Our derived extinctions are a factor of ∼2 higher than indicated by the 3D extinction map of Marshall et al. (2006), which was also based on red clump giants but used 2MASS data on a coarser angular scale. Moreover, in Paper II, we compare the ratio of 70 to 24 μm flux in the spectroscopic subsample with that found in a nearby sample of YSOs. At these far-IR wavelengths, where extinction is very low, we find similar flux ratios for class I systems in both data sets, indicating that most VVV class I YSOs have been correctly classified. In the following discussion, we directly compare embedded (class I and flat-spectrum) YSOs with sources in nearby SFRs. We simply ask the reader to note first that α may be slightly inflated by interstellar extinction and secondly that this more distant sample will be biased towards more luminous YSOs of intermediate mass (see Paper II).

The distribution of ΔK_s (Fig. 23, upper left panel) shows that peak of the distribution is found at larger ΔK_s for younger objects. We also observe a higher fraction of objects with $\Delta K_s > 2$ mag for class I objects (25 per cent) than for flat-spectrum (13 per cent) and class II (7 per cent) objects. The comparison of $\alpha_{\rm class}$ versus ΔK_s in the bottom panel of the same figure further illustrates the increase in amplitude at younger evolutionary stages, as well as the higher incidence of $\Delta K_s > 1$ variability at the younger stages. The class I and flat-spectrum YSOs constitute 87 per cent of the $\Delta K_s > 2$ subsample, dominating it even more than the full SFR-associated sample.

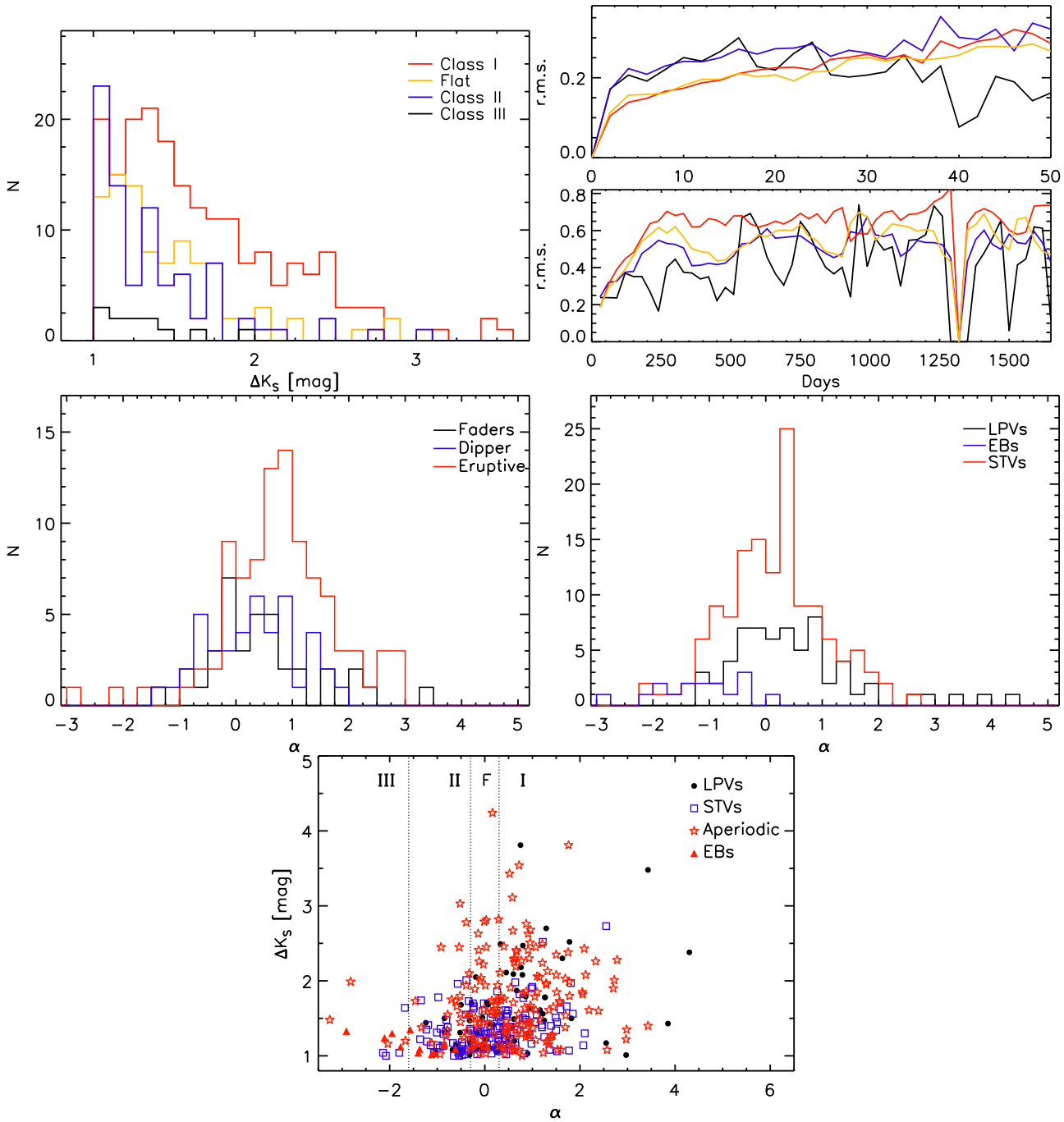

Figure 23. (top, left) ΔK_s distribution of the different YSO classes. (top, right) Mean rms variability for the different YSO classes in intervals up to 50 d (calculated with time bins of 1 d), and for intervals up to the full 1600 d baseline of the data set (using time bins of 30 d). The colour coding in the top (left and right) figures is the same as in Fig. 22. (middle, left) α distribution for fader (black line), dipper (blue line) and eruptive (red lines) objects. (middle, right) Same distribution but for long-term periodic objects (black lines), stars with short-term variability (red line) and EBs (blue line). (bottom) α versus ΔK_s for objects associated with SFRs. The limits for the different YSO classes are marked by dashed lines. Objects in this figure are divided according to the morphological light-curve classification (Section 4.1). The different classifications are marked in the plot.

We note that 70 objects with $\Delta K_s > 2$ mag also have redder near-IR colours than the full sample: the proportions of *J*-band non-detections and *JH* non-detections rises from 44 per cent and 11 per cent in the full SFR-associated sample to 56 per cent and 24 per cent in the $\Delta K_s > 2$ subsample. This further highlights the simple fact that efficient detection of the majority of YSOs with the most extreme variations requires observation at wavelengths $\lambda \geq 2\,\mu$m. Most importantly, this further emphasizes that younger objects have higher accretion variations. Eruptive variables are the largest component of the $\Delta K_s > 2$ sample, comprising 30/70 objects.

In Fig. 23, we also show the mean variability of YSOs belonging to different evolutionary classes as a function of time baseline. This is calculated by averaging the values of rms variability versus time interval computed using every possible pairing of two points within

the light curve of each star. In the figure, we show the variability over intervals up to 50 d (calculated with time bins of 1 d) and for intervals up to the full 1600 d baseline of the data set (using time bins of 30 d). The rms variability over short time-scales appears to be larger for more evolved objects than flat and class I sources. The variability increases with time for every YSO class and becomes flat at $t \sim 250$–350 d, although this is less clear for class III sources because of noise due to the lower number of objects in this class. Class I and flat sources have higher rms variability on these longer time-scales.

The higher rms variability in class II and III systems on time-scales <25 d can be explained by the fact that in these more evolved YSOs, the stellar photosphere contributes a greater proportion of the *K*-band luminosity of the system, whereas in less evolved YSOs, the luminosity is more dominated by the accretion disc. Consequently, we may expect a greater contribution to variability from cold spots and hot spots in the photosphere of class II and class III YSOs, which manifests itself on the time-scale of stellar rotation. We note that variability on rotational time-scales of a few days also contributes to the measured mean rms variability shown in Fig. 23 on all longer time-scales, which is why the variation increases rapidly on baselines from zero to 3 d (a typical rotation time-scale in YSOs; e.g. Alencar et al. 2010) and then increases more slowly thereafter.

It is also very interesting to see that the 250–350 d time-scale at which the maximum of mean rms variability is reached in all classes of YSO corresponds to variability on spatial scales of 1–2 au, assuming that the time-scale is determined by Keplerian rotation about low- to intermediate-mass YSOs. Connelley & Greene (2014) found from a spectroscopic variability study that mass accretion tracers in their sample of class I YSOs, such as Br γ and CO emission, are highly variable over time-scales of 1–3 yr and they proposed the above explanation of the time-scale.

Studies of the optical, near-IR and mid-IR temporal behaviour of YSOs have shown that the great majority are variable at these wavelengths, with their light curves showing a diversity of amplitudes, time-scales and morphologies (see e.g. Findeisen et al. 2013; Cody et al. 2014; Rebull et al. 2014; Rice et al. 2015; Wolk et al. 2015). These studies and the earlier large-scale study of Megeath et al. (2012) showed that the amplitude of the variability increases for younger embedded objects, though these works contained few YSOs, if any, with $\Delta K_s > 1$. Rice et al. (2012) found indications that amplitudes $\Delta K > 1$ mag are more common amongst class I systems (13 ± 7 per cent, based on two high-amplitude objects in a sample of 30 in the Braid Nebula within Cygnus OB7), whereas such high amplitudes are found to be less common in more evolved YSOs (see e.g. Carpenter, Hillenbrand & Skrutskie 2001).

4.4 Eruptive variability

We have found from the morphological classification that 106 of our SFR-associated high-amplitude variables have light curves that show sudden and large increases in brightness. Our near-IR colour variability data, though limited, appear to verify that variability does not arise from changes in the extinction along the line of sight in most cases. As we have discussed previously, large magnitude changes in our sample are more likely explained by either changes in the accretion rate or in the extinction along the line of sight. Since we appear to discard the latter effect in our eruptive variables, we infer that large changes in the accretion rate are the most likely explanation for the observed variability in these objects.

In Fig. 23 (middle panels), we show histograms of the spectral index of the YSOs of each light-curve type. We also see that the YSOs classified as eruptive have larger values of α (i.e. redder SEDs) than the other categories of YSO variables. The redder SEDs of the eruptive objects supports the idea that fluctuations in the accretion rates are larger and much more common at early stages of PMS evolution than in the class II stage. This is especially true given that class II YSOs typically outnumber class I YSOs in nearby SFRs by a factor of ∼3.7–4.8 (see e.g. table 1 in Dunham et al. 2014) due to the greater duration of the class II stage. We have 70 eruptive variables classified as class I YSOs and 5 classified as class II YSOs (the remainder being flat-spectrum or class III systems). If we assume that the YSOs classified as eruptive are mainly genuine eruptive variables (which is supported by our spectroscopic follow-up in Paper II), this tells us that the incidence of eruptive variability is ∼50–70 times higher in class I YSOs than class II YSOs. If we consider the possible correction to α due to foreground extinction (see Section 4.3), then we have 52 class I YSOs and 15 class II YSOs in the eruptive category, so eruptive variability is still 13–17 times more common in class I YSOs. We conclude that the difference is at least an order of magnitude.

It is interesting to see that some of the theoretical models that explain the outbursts observed in young stars predict that luminosity bursts are more common in the class I stage compared to later stages of stellar evolution. These models mainly involve gravitational instability (GI) in the outer disc, which is most likely to occur during the embedded phase. GI can cause the outer disc to fragment, forming bound fragments that later migrate into the inner disc. The infall of these fragments can lead to mass accretion bursts. GI can also produce a persistent spiral structure which efficiently transfers mass to small radii. The continuous pile up of mass at lower radii can trigger magneto-rotational instabilities (MRI), which lead to sudden disc outbursts (see e.g. Zhu, Hartmann & Gammie 2009; Audard et al. 2014; Vorobyov & Basu 2015, and references therein). Thus, invoking these mechanisms might explain the higher occurrence of eruptive variables at younger stages. However, we note that most mechanisms that explain eruptive variability, such as bursts due to binary interaction (Bonnell & Bastien 1992), MRI activated by layered accretion (Zhu et al. 2009) or thermal instabilities (Hartmann & Kenyon 1996) predict outbursts during class I through class II stages of YSO evolution.

The classification as eruptive variables only comes from K_s light curves and spectroscopic follow-up is needed to confirm their YSO nature. However, we note that potentially adding 106 more objects to the YSO eruptive variable class would increase the known members by a factor of 5. Moreover, our survey covers just a portion of the Galactic plane. Therefore, eruptive variability in YSOs might be more common than previously thought, especially for the most embedded and young objects. Spectroscopic follow-up of a subsample of the objects shows that a large fraction of them are indeed eruptive YSOs (see Paper II).

5 INCIDENCE OF ERUPTIVE VARIABILITY

To determine whether episodic accretion plays an important role in the assembly of stars, it is essential to have an estimate of the incidence of the phenomenon. Hartmann & Kenyon (1996) estimate that stars must spend ∼5 per cent of their lifetime in high states of accretion to gain their final mass during the infall phase. Enoch et al. (2009) and Evans et al. (2009) find that 5–7 per cent of class I stars in their sample are at high accretion states. In this section, we attempt to derive an initial estimate of this number using our sample of eruptive variables from VVV and the intrinsically red *Spitzer* sources of Robitaille et al. (2008).

We investigated the intrinsically red *Spitzer* sources of Robitaille et al. (2008), specifically those objects classified by the authors as likely YSOs. The use of GLIMPSE and MIPSGAL (Carey et al. 2009) photometry allows us to study the SEDs of this sample, where we find 2059 class I YSOs in the Galactic disc area studied in our work. From these objects, 51 are found in our list of high-amplitude variables in SFRs, 26 of them having the eruptive variable classification. These numbers would imply that approximately 2.5 per cent of class I YSOs show large-amplitude variability and that eruptive variability is observed in 1.3 per cent of YSOs at this evolutionary stage.

Once again, we need to take into account the completeness of our sample due to our selection criteria. In Section 2.1, we showed that our strict selection criteria caused our sample of high-amplitude variables to be only ∼50 per cent complete, down to $K_s = 15.5$ mag. In Section 2.1, we noted that at magnitudes fainter than $K_s = 15.5$ mag, the completeness of our selection falls more steeply. Thus, the incompleteness factor in the Robitaille et al. sample is likely to be higher.

We studied the completeness of the Robitaille et al. sample in four widely separated VVV tiles (d052, d065, d072 and d109) using 2010–2014 K_s data, where we found 138 counterparts in the VVV catalogues. If we only consider 2010–2012 data, 22 objects show high-amplitude variability,[2] are located in SFRs and have light curves that do not resemble those of AGB stars. From these, eight are part of the list of high-amplitude variables presented in this work, which implies an incompleteness factor of 22/8 or 2.75, caused by the quality cuts that were required in Section 2.1 to reduce the number of false positives to a manageable level. This figure is higher than the factor of ∼2 (50 per cent completeness) quoted above because many of the class I YSO variables from the Robitaille et al. (2008) list have mean $K_s > 15.5$ mag. If we consider the whole period of 2010–2014, then the number of high-amplitude variables increases to 29 raising the incidence by a further factor of 1.32.

We should also consider whether there is a bias towards erupting systems in the Robitaille et al. (2008) sample, which could in general be detected at greater distances or lower down the mass function than quiescent systems. Such a bias would only affect our calculation if erupting YSOs detected by GLIMPSE in ∼2003 also erupted a second time in 2010–2014. We tested for this by comparing the GLIMPSE *I*2 (4.5 μm) and *WISE W*2 (4.6 μm) magnitudes of the 26 systems from the Robitaille et al. (2008) sample that are included in our list of eruptive YSOs. *W*2 and *I*2 magnitudes for YSOs are typically similar (e.g. Antoniucci et al. 2014). Considering the 20/26 of these systems with *W*2 detections in the *WISE* AllSky catalogue (data from 2010, typically before the eruption), the median $I2 - W2$ is 0.02 mag. If we transform the *Spitzer I*2 magnitudes to the *W*2 passband using the equation given in Antoniucci et al. (2014), then the median difference between the *Spitzer* and *WISE* data is −0.14 mag. This small number indicates that any luminosity bias towards eruptive systems can be neglected in this initial estimate of their incidence.

After correcting for incompleteness, the incidence of high-amplitude variability among class I YSOs rises to 6.8 per cent, including only stars that varied in the 2010–2012 time period. This rises to 9 per cent if we extend the time period to 2010–2014. More importantly, the incidence of eruptive variables amongst class I YSOs reaches 3.4 per cent(or 4.6 per cent in the 2010–2014 data). The 4.6 per cent figure has a statistical uncertainty of 40 per cent, not including any biases arising from our use of the sample of Robitaille et al. (2008), so we should express the incidence of eruptive variability as about 3–6 per cent over a 4 yr time-scale. These figures happen to agree with the incidence of outbursts inferred by Enoch et al. (2009) and Evans et al. (2009) from the observation of class I YSOs with high bolometric luminosities. However, their observations are perhaps more likely to trace long-duration, FUor-like outbursts than those detected by VVV.

If we assume that all class I YSOs go through episodes of enhanced accretion at the same average rate, then our estimate that ∼4 per cent of them burst over a 4 yr period would imply that roughly every source suffers at least one burst over 100 yr. Froebrich & Makin (2016) study the distribution of the separation between large H_2 knots in jets. The knots likely trace the accretion burst history of a particular source. Froebrich & Makin find that the bright knots have typical separations that correspond to about 1000 yr, which they conclude is too large to be EXor-driven and too small to be FUor-driven. This number is also larger than the 100 yr burst frequency of the VVV sample. However, further study needs to be done to obtain a more robust estimate and the assumption of similar behaviour for all class I YSOs is of course questionable.

6 SUMMARY AND CONCLUSIONS

We have searched for high-amplitude IR variables in a 119 deg^2 area of the Galactic mid-plane covered by the VVV survey, using a method that tends to exclude transients and eruptive variables that saturated during outburst or very faint in quiescence, owing to our requirement for a high-quality detection at every epoch. We discovered 816 bona fide variables in the 2010–2012 data with $\Delta K_s > 1$ in that time interval. Nearly all of these were previously unknown as variable stars, though a significant minority had been identified as embedded YSO candidates in the Robitaille et al. (2008) catalogue of stars with very red [4.5]–[8.0] *Spitzer*/GLIMPSE colours.

We have presented a fairly simple analysis of the sample using the 2010–2014 VVV light curves, supplemented by a recently obtained second epoch of multifilter JHK_s data and photometry from *WISE* and *Spitzer*. Our main conclusions are as follows.

(i) In agreement with the previous results from searches in the UKIDSS GPS (Contreras Peña et al. 2014), we observe a strong concentration of high-amplitude IR variables towards areas of star formation. The two-point correlation function and nearest neighbour distribution of VVV objects show evidence for clustering on angular scales typical of distant Galactic clusters and SFRs. The variable stars found in SFRs are characterized by having near-IR colours and SEDs of YSOs.

(ii) The most common types of variable outside SFRs are LPVs (typically dust-obscured Mira variables because other types are saturated in VVV) and EBs. By visual inspection of the light curves of the variables in SFRs, we were able to identify and remove most of these contaminating systems and provide a reasonably clean sample of high-amplitude YSOs. The YSOs make up about half of the full sample of 816 variables.

(iii) We analysed the light curves of the variables in SFRs, after removal of likely Mira-types, and we classify them as 106 eruptive variables, 39 faders, 45 dippers, 162 STVs, 65 long-term periodic variables ($P > 100$ d) and 24 EBs and EB candidates. Individual YSOs may display more than one type of variability and the low-amplitude variation on short time-scales seen in normal YSOs is common in every category.

[2] The true variability of these objects was confirmed via visual inspection of 1 arcmin × 1 arcmin images.

(iv) Spectroscopic follow-up of a substantial subset of the variables with eruptive light curves is presented in the companion paper (Paper II), confirming that the great majority of systems with eruptive light curves are indeed eruptive variables with signatures of strong accretion similar to those seen in EXors or FUors, or a mixture of the two. The two epochs of VVV JHK_s multicolour data indicate that extinction is not the main cause of the variability in systems with eruptive light curves, though there is a trend for the sources to become bluer when brighter, similar to EXors. The faders show a wider range of colour behaviour, with more examples of 'bluer when fainter' than the other categories.

(v) Unsurprisingly, very few of the EBs showed significant magnitude or colour changes between the two multicolour epochs. Amongst the STVs and dippers, we observe large colour and magnitude changes that appear to differ from the reddening vector, although this conclusion seems to depend on the selection of objects from these classes. We observed changes consistent with the reddening vector in a few systems. It is possible that the flux changes in some of the STVs are caused by extinction, e.g. as has been observed in some binary YSOs with a circum-binary disc (Windemuth & Herbst 2014; Rice et al. 2015).

(vi) The STVs typically have lower amplitudes than any other category except the EBs and EB candidates. Moreover, the STVs have a bluer distribution of spectral indices than the full YSO sample, including a high proportion of class II YSOs with relatively low extinction. It is likely that in many of the STVs with periods $P < 15$ d and K_s amplitudes, not far above 1 mag, the light curve is rotationally modulated by unusually prominent bright or dark spots on the photosphere.

(vii) Variables in the eruptive and fader categories tend to have higher amplitudes than the full YSO sample over the 4 yr period, with mean $\Delta K_s = 1.72$ and 1.95 mag, respectively, compared to 1.56 mag for the full YSO sample.

(viii) It is reasonable to suppose that variable accretion is the cause of photometric variability in a proportion of the faders and some of the long-term periodic variables and STVs, although spectroscopic confirmation is limited in these categories as yet. In the periodic variables, the accretion rate would presumably be modulated by a companion body (Hodapp et al. 2012). Adding these together with the ∼100 YSOs with eruptive light curves suggests that the sample contains between 100 and 200 YSOs in which the variability is caused by large changes in accretion. If we take the lower end of this range, this increases the number of probable eruptive variable YSOs available for study by a factor of 5. Some of these systems have K_s amplitudes not far above the 1 mag level, below which variability due to spots and smaller changes in accretion rate and extinction becomes much more common. However, we see no clear argument for a higher threshold given that the amplitudes have a continuous distribution that can be influenced in individual YSOs by the mass and luminosity of the central protostar as much as the change in accretion rate (Calvet et al. 1991).

(ix) As a whole, the high-amplitude variables are a very red sample, dominated by embedded YSOs (i.e. systems with Class I or flat-spectrum SEDs that are typically not observable at optical wavelengths). This contrasts with the optical selection of the classical FUors and EXors, the majority of which have a class II or flat-spectrum classification. While FUors are often discussed as systems with a remnant envelope, only three or four deeply embedded eruptive variables were known prior to this study and our recent UGPS 2-epoch study.

(x) Variables with eruptive light curves tend to have the reddest SEDs. The spectral indices in the eruptive category indicate that this type of variability is at least one order of magnitude more common among class I YSOs than class II YSOs. This demonstrates that eruptive variability is either much more common or recurs much more frequently amongst YSOs at earlier stages of PMS evolution, when average accretion rates are higher. We hope that this result will inform ongoing efforts to develop a theoretical framework for the phenomenon.

(xi) For the full sample of YSOs, the rms variability is higher at earlier evolutionary (SED) classes for time intervals longer than 25 d and reaches a maximum at 250–350 d for all SED classes. The full duration of the outbursts in eruptive systems is typically 1–4 yr. Some variables with eruptive light curves show more than one outburst. If some of the faders are eruptive variables in decline, then a similar or slightly longer duration would apply.

(xii) At time intervals shorter than 25 d, the evolutionary dependence is reversed, with class II YSOs showing higher amplitudes than flat-spectrum and class I YSOs. This suggests that the shorter time-scale variations are dominated by rotational modulation by spots on the photosphere (which are more readily observed in class II systems than embedded YSOs), whereas accretion variations usually take place on longer time-scales.

(xiii) The 1–4 yr duration of the eruptions is between that of FUors (>10 yr) and EXors (weeks to months) A small but growing number of eruptive YSOs with these intermediate durations have been found in recent years, some of which have a mixture of the spectroscopic characteristics of FUors and EXors. This has led to the recent concept that FUor and EXors may simply be part of a range of different eruptive behaviours on different time-scales, all driven by large variations in accretion rate. Until now it was unclear whether these recent discoveries were rare exceptions, but it now seems clear that they are not. In fact, we find that YSOs with intermediate outburst durations outnumber short EXor-like outbursts and are now the majority of known eruptive systems, A much longer duration survey would be required to determine the incidence of FUor-like outbursts amongst embedded systems. In Paper II, we propose a new class of eruptive variable to describe YSOs with eruptive outbursts of intermediate duration, which are usually optically obscured class I or flat-spectrum YSOs and display a variety of the EXor-like and/or FUor-like spectroscopic signatures of strong accretion.

(xiv) We investigated the intrinsically red *Spitzer* sources of Robitaille et al. (2008), specifically those objects classified by the authors as likely YSOs. We find 2059 such objects in the area covered by VVV. From these, 51 are found in the list of high-amplitude variables in SFRs from this work, with 26 of them being classified as eruptive variables. After correcting for incompleteness imposed by the strict selection criteria in the VVV sample, we estimate that high-amplitude variability is observed from 6.8 per cent (when considering 2010–2012 data) up to 9 per cent (when considering 2010–2014 data) of class I YSOs. More importantly, the incidence of eruptive variability amongst class I YSOs rises to 3.4–4.6 per cent. This estimate agrees with those inferred from observations of class I YSOs by Evans et al. (2009) and Enoch et al. (2009). However, the agreement might be a coincidence considering that their observations are likely tracing long-duration, FUor-like outbursts, rather than those detected in our study.

(xv) YSOs are the commonest type of high-amplitude IR variable detected by the VVV survey. We estimate a completeness-corrected source density of 7 deg^{-2} in the mid-plane of quadrant 4, in the approximate mean magnitude range $11 < K_s < 16$. These YSOs are detected at typical distances of a few kiloparsecs, there being no very nearby SFRs in the area surveyed; they are therefore likely to

be intermediate-mass YSOs. If we were able to detect such objects out to the far edge of the Galactic disc, the source density would rise to perhaps ∼40 deg^{-2}. This confirms our previous suggestion in Contreras Peña et al. (2014) that high-amplitude YSO variables have a higher source density and average space density than Mira variables. EBs are very common at low amplitudes and they may have a comparable space density to YSOs at K_s amplitudes of 1–1.6 mag (an approximate upper limit for EBs in *Kepler*). YSOs are more numerous at higher amplitudes and may well be more numerous for all amplitudes over 1 mag if the eruptive phenomenon extends to the lower part of the stellar initial mass Function.

ACKNOWLEDGEMENTS

This work was supported by the UK's Science and Technology Facilities Council, grant numbers ST/J001333/1, ST/M001008/1 and ST/L001403/1.

We gratefully acknowledge the use of data from the ESO Public Survey program 179.B-2002 taken with the VISTA 4.1 m telescope and data products from the Cambridge Astronomical Survey Unit. Support for DM and CC is provided by the Ministry of Economy, Development, and Tourism's Millennium Science Initiative through grant IC120009, awarded to the Millennium Institute of Astrophysics, MAS. DM is also supported by the Center for Astrophysics and Associated Technologies PFB-06 and Fondecyt Project No. 1130196. MSNK is supported by the Marie-Curie Intra-European fellowship project GESTATE (661249). This research has made use of the SIMBAD data base, operated at CDS, Strasbourg, France; also the SAO/NASA Astrophysics data (ADS). ACG was supported by the Science Foundation of Ireland, grant 13/ERC/I2907. We also acknowledge the support of CONICYT REDES project No. 140042 'Young variables and proper motion in the Galactic plane. Valparaiso–Hertfordshire collaboration'

C. Contreras Peña was supported by a University of Hertfordshire PhD studentship in the earlier stages of this research.

We thank Janet Drew for her helpful comments on the structure of the paper.

REFERENCES

Alencar S. H. P. et al., 2010, A&A, 519, A88
Antoniucci S., Giannini T., Li Causi G., Lorenzetti D., 2014, ApJ, 782, 51
Armstrong D. J., Gómez Maqueo Chew Y., Faedi F., Pollacco D., 2014, MNRAS, 437, 3473
Aspin C., Beck T. L., Reipurth B., 2008, AJ, 135, 423
Aspin C., Greene T. P., Reipurth B., 2009, AJ, 137, 2968
Audard M. et al., 2014, in Beuther H., Klessen R. S., Dullemond C. P., Henning T., eds, Protostars and Planets VI. Univ. Arizona Press, Tucson, AZ, p. 387
Avedisova V. S., 2002, Astron. Rep., 46, 193
Bans A., Königl A., 2012, ApJ, 758, 100
Baraffe I., Chabrier G., Gallardo J., 2009, ApJ, 702, L27
Baraffe I., Vorobyov E., Chabrier G., 2012, ApJ, 756, 118
Bate M. R., Clarke C. J., McCaughrean M. J., 1998, MNRAS, 297, 1163
Benjamin R. A. et al., 2003, PASP, 115, 953
Bessell M. S., Brett J. M., 1988, PASP, 100, 1134
Bonnell I., Bastien P., 1992, ApJ, 401, L31
Bouvier J., Grankin K., Ellerbroek L. E., Bouy H., Barrado D., 2013, A&A, 557, A77
Budding E., Erdem A., Çiçek C., Bulut I., Soydugan F., Soydugan E., Bakiş V., Demircan O., 2004, A&A, 417, 263
Calvet N., Patino A., Magris G. C., D'Alessio P., 1991, ApJ, 380, 617
Caratti o Garatti A. et al., 2011, A&A, 526, L1
Caratti o Garatti A. et al., 2012, A&A, 538, A64
Carey S. J. et al., 2009, PASP, 121, 76
Carpenter J. M., Hillenbrand L. A., Skrutskie M. F., 2001, AJ, 121, 3160
Catelan M. et al., 2013, preprint (arXiv:1310.1996)
Chen C.-H. R., Chu Y.-H., Gruendl R. A., Gordon K. D., Heitsch F., 2009, ApJ, 695, 511
Cioni M.-R. L. et al., 2011, A&A, 527, A116
Cody A. M. et al., 2014, AJ, 147, 82
Connelley M. S., Greene T. P., 2014, AJ, 147, 125
Contreras Peña C. et al., 2014, MNRAS, 439, 1829
Contreras Peña C. et al., 2016, MNRAS, in press (Paper II)
Culverhouse T. et al., 2011, ApJS, 195, 8
Derue F. et al., 2002, A&A, 389, 149
Drew J. E. et al., 2014, MNRAS, 440, 2036
Dunham M. M. et al., 2014, in Beuther H., Klessen R. S., Dullemond C. P., Henning T., eds, Protostars and Planets VI. Univ. Arizona Press, Tucson, AZ, p. 195
Eiroa C. et al., 2002, A&A, 384, 1038
Enoch M. L., Evans N. J., II, Sargent A. I., Glenn J., 2009, ApJ, 692, 973
Epchtein N. et al., 1994, Ap&SS, 217, 3
Evans N. J., II, et al., 2009, ApJS, 181, 321
Findeisen K., Hillenbrand L., Ofek E., Levitan D., Sesar B., Laher R., Surace J., 2013, ApJ, 768, 93
Froebrich D., Makin S. V., 2016, MNRAS, 462, 1444
Gonzalez O. A., Rejkuba M., Zoccali M., Valenti E., Minniti D., 2011, A&A, 534, A3
Greene T. P., Lada C. J., 1996, AJ, 112, 2184
Greene T. P., Wilking B. A., Andre P., Young E. T., Lada C. J., 1994, ApJ, 434, 614
Hartmann L., Kenyon S. J., 1996, ARA&A, 34, 207
Herbst W., Shevchenko V. S., 1999, AJ, 118, 1043
Hodapp K. W., Chini R., 2015, ApJ, 813, 107
Hodapp K. W., Chini R., Watermann R., Lemke R., 2012, ApJ, 744, 56
Hodapp K.-W., Hora J. L., Rayner J. T., Pickles A. J., Ladd E. F., 1996, ApJ, 468, 861
Huckvale L., Kerins E., Sale S. E., 2014, MNRAS, 442, 259
Hughes J. P., Slane P., Posselt B., Charles P., Rajoelimanana A., Sefako R., Halpern J., Steeghs D., 2010, Astronomer's Telegram, 2771, 1
Irwin M., 2009, UKIRT Newsl., 25, 15
Ishihara D., Kaneda H., Onaka T., Ita Y., Matsuura M., Matsunaga N., 2011, A&A, 534, A79
Javadi A., Saberi M., van Loon J. T., Khosroshahi H., Golabatooni N., Mirtorabi M. T., 2015, MNRAS, 447, 3973
Jiménez-Esteban F. M., García-Lario P., Engels D., Perea Calderón J. V., 2006a, A&A, 446, 773
Jiménez-Esteban F. M., García-Lario P., Engels D., Manchado A., 2006b, A&A, 458, 533
Kenyon S. J., Hartmann L. W., Strom K. M., Strom S. E., 1990, AJ, 99, 869
Koenig X. P., Leisawitz D. T., 2014, ApJ, 791, 131
Kóspál Á., Ábrahám P., Prusti T., Acosta-Pulido J., Hony S., Moór A., Siebenmorgen R., 2007, A&A, 470, 211
Lada C. J., 1987, in Peimbert M., Jugaku J., eds, Proc. IAU Symp. 115, Star Forming Regions. Kluwer, Dordrecht, p. 1
Lawrence A. et al., 2007, MNRAS, 379, 1599
López-Morales M., Clemens J. C., 2004, PASP, 116, 22
Lorenzetti D., Larionov V. M., Giannini T., Arkharov A. A., Antoniucci S., Nisini B., Di Paola A., 2009, ApJ, 693, 1056
Lorenzetti D. et al., 2012, ApJ, 749, 188
Lucas P. W. et al., 2008, MNRAS, 391, 136
Marshall D. J., Robin A. C., Reylé C., Schultheis M., Picaud S., 2006, A&A, 453, 635
Megeath S. T. et al., 2012, AJ, 144, 192
Meyer M. R., Calvet N., Hillenbrand L. A., 1997, AJ, 114, 288
Minniti D. et al., 2010, New Astron., 15, 433
Minniti D., Saito R. K., Alonso-García J., Lucas P. W., Hempel M., 2011, ApJ, 733, L43
Murakami H. et al., 2007, PASJ, 59, 369
Ninan J. P. et al., 2015, ApJ, 815, 4
Ochsenbein F., Bauer P., Marcout J., 2000, A&AS, 143, 23

Paczyński B., Szczygieł D. M., Pilecki B., Pojmański G., 2006, MNRAS, 368, 1311
Persi P., Tapia M., Gòmez M., Whitney B. A., Marenzi A. R., Roth M., 2007, AJ, 133, 1690
Pietrukowicz P. et al., 2013, Acta Astron., 63, 115
Plavchan P., Gee A. H., Stapelfeldt K., Becker A., 2008, ApJ, 684, L37
Price S. D., Egan M. P., Carey S. J., Mizuno D. R., Kuchar T. A., 2001, AJ, 121, 2819
Rebull L. M. et al., 2014, AJ, 148, 92
Reipurth B., Aspin C., 2010, in Harutyunian H. A., Mickaelian A. M., Terzian Y., eds, Proc. Conf. Evolution of Cosmic Objects through their Physical Activity. Gitutyun Publ. House, Yerevan, p. 19
Rice T. S., Wolk S. J., Aspin C., 2012, ApJ, 755, 65
Rice T. S., Reipurth B., Wolk S. J., Vaz L. P., Cross N. J. G., 2015, AJ, 150, 132
Robitaille T. P., Whitney B. A., Indebetouw R., Wood K., Denzmore P., 2006, ApJS, 167, 256
Robitaille T. P. et al., 2008, AJ, 136, 2413
Romanova M. M., Ustyugova G. V., Koldoba A. V., Lovelace R. V. E., 2013, MNRAS, 430, 699
Russeil D., 2003, A&A, 397, 133
Safron E. J. et al., 2015, ApJ, 800, L5
Saito R. K. et al., 2012, A&A, 537, A107
Saito R. K. et al., 2013, A&A, 554, A123
Samus N. N., Kazarovets E. V., Kireeva N. N., Pastukhova E. N., Durlevich O. V., 2010, Odessa Astron. Publ., 23, 102
Schechter P. L., Mateo M., Saha A., 1993, PASP, 105, 1342
Schlafly E. F., Finkbeiner D. P., 2011, ApJ, 737, 103
Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525
Scholz A., 2012, MNRAS, 420, 1495
Scholz A., Froebrich D., Wood K., 2013, MNRAS, 430, 2910
Simpson R. J. et al., 2012, MNRAS, 424, 2442
Skrutskie M. F. et al., 2006, AJ, 131, 1163
Stellingwerf R. F., 1978, ApJ, 224, 953
Tudose V., Fender R. P., Tzioumis A. K., Spencer R. E., van der Klis M., 2008, MNRAS, 390, 447
van Loon J. T., Cohen M., Oliveira J. M., Matsuura M., McDonald I., Sloan G. C., Wood P. R., Zijlstra A. A., 2008, A&A, 487, 1055
Vijh U. P. et al., 2009, AJ, 137, 3139
Vorobyov E. I., Basu S., 2015, ApJ, 805, 115
Ward J. L., Oliveira J. M., van Loon J. T., Sewiło M., 2016, MNRAS, 455, 2345
Wenger M. et al., 2000, A&AS, 143, 9
Whitelock P. A., Feast M. W., van Leeuwen F., 2008, MNRAS, 386, 313
Windemuth D., Herbst W., 2014, AJ, 147, 9
Wolk S. J., Rice T. S., Aspin C., 2013, ApJ, 773, 145
Wolk S. J. et al., 2015, AJ, 150, 145
Wright E. L. et al., 2010, AJ, 140, 1868
Zhu Z., Hartmann L., Gammie C., 2009, ApJ, 694, 1045

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

Table 1. K_s photometry of the 816 high-amplitude variable stars from VVV.
Table 2. Parameters of the high-amplitude variables from VVV. For the description of the columns, see Section 3.1. Here we show the first 40 sources in the list.

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 A: EBS

In order to compare the YSO and EB space densities, we look at the measured source density of EBs detected in Galactic disc fields by OGLE-III (Pietrukowicz et al. 2013) and estimate their distances with the help of a recent analysis of the physical properties of *Kepler* EBs (Armstrong et al. 2014).

Shallow surveys such as the Automated All-Sky Survey (Paczyński et al. 2006) are sensitive only to the more luminous high-amplitude EBs, which are mainly early-type Algol systems, similar to the high-amplitude EBs in the General Catalogue of Variable Stars (Samus et al. 2010). However, simulations by López-Morales & Clemens (2004) find that EBs seen by deeper surveys such as OGLE-III will be dominated by later type systems. Armstrong et al. (2014) analysed the large sample of EBs discovered by *Kepler*, providing temperature and radius estimates for the full sample. They note that their results are most accurate for cases where the temperatures of the primary and the secondary are very different, which is fortunate because this is characteristic of high-amplitude EBs, in which a hotter main-sequence star is eclipsed by a cooler giant star. Inspection of their results shows that high-amplitude EBs fall into two groups: EBs with early type primaries and EBs with F- or early G-type primaries. Simple calculations assuming Planckian SEDs indicate that only the latter group can produce eclipses deeper than 1 mag in the K_s passband (and this group also dominates in the OGLE I passband) because early-type stars emit little of their total flux in the IR. The same calculations indicate that no systems with K_s amplitudes above 1.6 mag exist in the *Kepler* sample, whereas YSOs with higher amplitudes are common in our sample; see Section 4. Applying this approach in the optical indicates that EBs should exist with amplitudes up to 3 mag in the I passband. This limit is confirmed by the OGLE-III sample of Pietrukowicz et al. (2013), thereby giving confidence that our limit of 1.6 mag at K_s is robust. Indeed, our VVV sample of 72 EBs contains no systems with higher amplitude than this. We can assume that the EBs from Pietrukowicz et al. (2013) are mainly composed of F- or early G-type stars, based on the *Kepler* results, together with the fact that later type primaries are not expected given the need for these systems to have produced a post-main-sequence star within the lifetime of the Galactic disc.

Pietrukowicz et al. (2013) report the discovery of 11589 EBs in area of 7.21 deg^2 across different fields on the Galactic mid-plane. The catalogue of EBs is reported to be 75 per cent complete to I=18. Correcting for the 75 per cent completeness indicates a source density of 2143 EBs deg^{-2}, from which only ∼1.6 per cent display $\Delta I > 1$ mag. Our earlier calculation (assuming Planckian SEDs) from the results of Armstrong et al. (2014) suggests that this drops to 0.6 per cent at K_s. This implies ∼12 high-amplitude EBs per deg^2. We can estimate the extinction to the EBs from their observed $V-I$ colours and then roughly estimate their typical distances either using the absolute I magnitudes of F- to early G-type stars or by converting extinction to distance using the red clump giant branch (as in Section 4.3). Both approaches indicate that the EBs from Pietrukowicz et al. (2013) are typically at heliocentric distances of a few kiloparsecs, which is similar to the estimated distances to our YSOs based on radial velocities, literature distances to the SFRs and SED fitting (see Paper II). Given that the surface density of EBs (∼12 deg^{-2}) is similar to the surface density of the VVV YSO population sampled in this study (∼7 deg^{-2}), the similar distances suggest that the high-amplitude EB and YSO populations may have similar average space densities, within an order of magnitude. The YSO population is certainly more numerous at K_s amplitudes over 1.6 mag. If the high-amplitude YSO population extends well below a solar mass, which is very possible, then they would probably be more numerous even at our ΔK_s=1 mag threshold. However, the VVV and OGLE-III data tell us only that both populations are substantial. The number of YSOs rises steeply up to the sensitivity

Table B1. Colour and magnitude changes measured by *WISE*.

Category	Median $\Delta W1$	Median $\Delta W2$	Median $\Delta(W1-W2)$	No. in sample
Faders	1.33	0.93	0.37	12
Eruptive	0.72	0.56	0.26	21
LPV-YSOs	0.62	0.67	0.13	23
STVs	0.50	0.38	0.10	11
Dippers	0.59	0.62	0.27	9

limit and a similar trend is obtained in EBs if we assume that a significant fraction of the unclassified variables shown in Fig. 10 are EBs.

APPENDIX B: MID-IR COLOUR AND MAGNITUDE VARIABILITY

Photometry from the *WISE* and NEOWISE missions in the $W1$ (3.4 μm) and $W2$ (4.6 μm) passbands are available at the four epochs for the whole sky, with simultaneous photometry in the two filters. The *WISE* satellite scanned the whole sky twice in 2010, at epochs separated by 6 months, and the extended mission, NEOWISE, has repeated the process in 2014. For any given sky location, each epoch is composed of multiple scans taken over a period of several days. For all the high-amplitude variables, we downloaded photometry for all these scans from the AllWISE Multiepoch Photometry table (for the 2010 data) and NEOWISE-R Single Exposure L1b Source Table (for the 2014 data), both of which are archived in IRSA.

Inspection of the $W1$ and $W2$ photometry indicated that the uncertainties typically become larger for saturated stars and for faint stars in crowded Galactic fields. For this analysis, we therefore considered only sources with $7 < W1 < 11$, and $6 < W2 < 11$ (defining these cuts with the AllWISE Source Catalogue). We combined the data from the multiple scans into the four widely separated epochs by binning the scans into groups (epochs) with a maximum baseline of 20 d and computing the medians of $W1$, $W2$ and $(W1 - W2)$ at each epoch.

In most cases, the *WISE* data do not sample the full peak to trough variation in the VVV light curves, so changes in $W1$ and $W2$ fluxes are often small. We define $\Delta W1$, $\Delta W2$ and $\Delta(W1 - W2)$ for each source as the largest difference observed at any two epochs (not necessarily the same pair of epochs for each quantity). Then considering each type of VVV light-curve category, we present the median of these quantities in Table B1. We see that both magnitude and colour variability are typically larger for faders than the other categories, as is the case in K_s. Eruptive sources, dippers and lpv-YSOs have similar magnitude changes, while STVs have smaller changes (as in K_s). EBs typically show negligible changes since the eclipses are very rarely sampled.

The WISE data were less useful than the JHK_s data for investigating the physical cause of the photometric variability in K_s because the extinction versus wavelength relation appears to depend strongly on environment (Koenig & Leisawitz 2014). Also, extinction is lower in the *WISE* passbands than at K_s so extinction may have less effect on the $W1 - W2$ colour than even a modest change in accretion rate.

In most STVs and LPV-YSOs, we can confidently state the extinction is not the main cause of the measured $W1-W2$ colour changes because the trend over the four epochs is 'bluer when fainter' or a negligible colour change or there is no clear trend. In the eruptive and fader categories, about half the variables (10/21 and 7/12) have a 'redder when fainter' trend that might be due to variable accretion, variable extinction or a combination of the two (in some sources, one of the four epochs does not follow the trend of the other three epochs). The remainder of the eruptive variables and faders show a different trend or no clear result. In some sources with a 'redder when fainter' trend in the *WISE* data, the two epochs of JHK_s data appear to rule out extinction as the cause of variability. In dippers, the majority (5/9) sources have a 'redder when fainter' trend that might be consistent with extinction.

These results are not very informative because of the limited sampling of the *WISE* data. In most cases, $K_s - W1$, $K_s - W2$ changes cannot be investigated because there are very few sources for which two of the *WISE* epochs are contemporaneous with VVV K_s epochs.

Fortunately, there are some sources with clear long-term trends in the 2010–2014 VVV light curves for which we can usefully compare *WISE* data from 2010 and 2014. Four examples of such sources are VVVv270, VVVv631, VVV118 and VVVv562, all of which are members of the spectroscopic subsamples that are discussed individually in Paper II. To bring together those sources here, we note that all four have unusually large amplitudes in K_s, $W1$ and $W2$ ($2.6 < \Delta K_s < 4.2$ mag). The first two show a rising trend (classed as eruptive light curves), whereas the other two show a declining trend: VVVv562 is a fader but VVVv118 is an eruptive variable with several brief eruptions superimposed on a long-term decline. In every case, the trend is similar in the three wavelengths but the amplitude is larger in K_s than in $W1$ and $W2$, typically by a factor between 1 and 2. Three sources have $\Delta W1 \approx \Delta W2$ (VVVv118, VVVv562, VVVv631), whereas in VVVv270, $\Delta W1 \approx 2\Delta W2$.

The spectroscopic evidence in Paper II indicates that these four variables are all bona fide eruptive variables driven by episodic accretion, so the different amplitudes in the different filters presumably reflect differing changes in the luminosity of the different regions of the accretion disc responsible for most of the emission at each wavelength, as well as the effect of temperature changes on the flux emitted by each region at different wavelengths.

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