The Astrophysical Journal, 864:11 (19pp), 2018 September 1 https://doi.org/10.3847/1538-4357/aacc65

An Automated Tool to Detect Variable Sources in the Vista Variables in the Vía Láctea Survey: The VVV Variables (V^4) Catalog of Tiles d001 and d002

N. Medina[1,2], J. Borissova[2,1], A. Bayo[2,3], R. Kurtev[2,1], C. Navarro Molina[1,2], M. Kuhn[1,2], N. Kumar[4,5], P. W. Lucas[5], M. Catelan[1,6], D. Minniti[1,7,8], and L. C. Smith[9,10]
[1] Millennium Institute of Astrophysics (MAS), Santiago, Chile; nicolas.medina@postgrado.uv.cl
[2] Instituto de Física y Astronomía, Universidad de Valparaíso, Av. Gran Bretaña 1111, Playa Ancha, Casilla 5030, Chile
[3] Núcleo Milenio Formación Planetaria—NPF, Universidad de Valparaíso, Av. Gran Bretaña 1111, Valparaíso, Chile
[4] Instituto de Astrofísica e Ciências do Espaco, Universidade do Porto, CAUP, Rua das Estrelas, 4150-762 Porto, Portugal
[5] Centre for Astrophysics, University of Hertfordshire, College Lane, Hatffeld, AL10 9AB, UK
[6] Instituto de Física, Facultad de Física, Pontificia Universidad Católica de Chile, Casilla 306, Santiago 22, Chile
[7] Departamento de Ciencias Físicas, Facultad de Ciencias Exactas, Universidad Andrés Bello, Av. Fernandez Concha 700, Las Condes, Santiago, Chile
[8] Vatican Observatory, V00120 Vatican City State, Italy
[9] Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK
[10] Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK
Received 2018 January 12; revised 2018 May 13; accepted 2018 May 16; published 2018 August 24

Abstract

Time-varying phenomena are one of the most substantial sources of astrophysical information, and their study has led to many fundamental discoveries in modern astronomy. We have developed an automated tool to search for and analyze variable sources in the near-infrared K_s band using the data from the VISTA Variables in the Vía Láctea (VVV) ESO Public Large Survey. This process relies on the characterization of variable sources using different variability indices calculated from time series generated with point-spread function (PSF) photometry of sources under analysis. In particular, we used two main indices, the total amplitude ΔK_s and the eta index η, to identify variable sources. Once the variable objects are identified, periods are determined with generalized Lomb–Scargle periodograms and the information potential metric. Variability classes are assigned according to a compromise between comparisons with VVV templates and the period of the variability. The automated tool is applied on VVV tiles d001 and d002 and led to the discovery of 200 variable sources. We detected 70 irregular variable sources and 130 periodic ones. In addition, nine open-cluster candidates projected in the region are analyzed, and the infrared variable candidates found around these clusters are further scrutinized by cross-matching their locations against emission star candidates from VPHAS+ survey H_α color cuts.

Key words: catalogs – open clusters and associations: general – stars: pre-main sequence – techniques: photometric

Supporting material: machine-readable tables

1. Introduction

Time-varying phenomena are arguably one of the most powerful sources of astrophysical information. In recent decades, the development of astronomical instrumentation and automation has enabled many time-domain surveys, such as the wide-field optical imaging surveys the Catalina Real-time Transient Survey (Drake et al. 2009), Pan-STARRS (Kaiser et al. 2002); and *Gaia* (Perryman 2005). In the near future, even more ambitious programs, such as the Large Synoptic Survey Telescope (LSST; Krabbendam & Sweeney 2010), are planned to start monitoring the optical sky. While optical surveys are getting wider and deeper, the extension and the systematic exploration of the variable sky toward the infrared are also under development in order to better cope with the problem of the interstellar extinction. The VISTA Variables in the Vía Láctea survey (VVV; Minniti et al. 2010; Saito et al. 2012a) is one of these infrared surveys and is comparable to the optical ones both in areal and time-domain coverage (e.g., Arnaboldi et al. 2007, 2012). It has been designed to catalog $\sim 10^9$ sources, where a great part of those are expected to be variable stars. All of these sources will be used to map the structure of the optically obscured Galactic disk and bulge by using some main distance indicators, such as the red-clump giants and pulsating variable stars (RR Lyrae stars, classical Cepheids, anomalous Cepheids and Miras, and semi-regular variables), as well as to provide a census of young stellar objects (YSOs) across the southern Galactic plane. Some focused studies have been carried out in the southern disk region in search of variable stars using these data, for example, Contreras Peña et al. (2017a, 2017b), cataloging high-amplitude variable stars with an emphasis on YSOs; Borissova et al. (2016), searching for YSOs around young stellar clusters; Dékány et al. (2015), searching classical Cepheids in the bulge; and Elorrieta et al. (2016), Gran et al. (2016), and Minniti et al. (2017), focusing on the RR Lyrae stars. However, the systematic and uniform searches of the variability phenomena in the VVV disk area are still missing. On the other hand, the VVV multi-epoch observations produced a huge amount of information, a data set of challenging size. It is highly necessary to develop tools, processes, and techniques able to perform sophisticated analysis in an automated way in order to efficiently exploit this unique data set. In this paper, we present an automated tool designed to search for, classify, and analyze variable sources in the near-infrared K_s band. Our tool is fed VVV tile images to extract time series and identify different types of variable sources. The main goal is to understand the behavior of K_s-band variability in large regions of the sky, with the ultimate goal of processing the 1.8 deg^2 images of the VVV observations. The identified variables will be used to derive properties in active star-forming regions (SFRs) to determine distances using the periodic stars with available

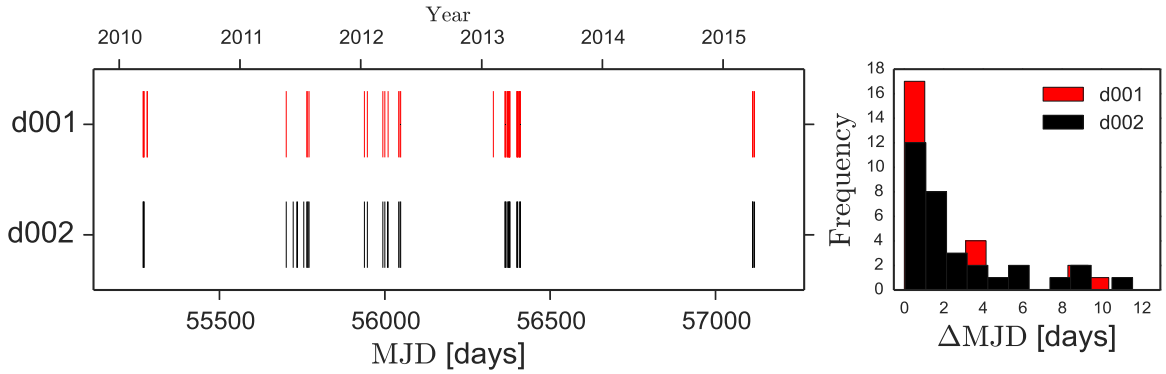

Figure 1. Left: log of observations of VVV tiles d001 and d002 between 2010 and 2015, based on the log representation of Rebull et al. (2014). Each photometric measurement is marked by a "|" symbol. The bars are thicker in places with high cadence. Right: histogram of differences between consecutive observations ΔMJD. It is possible to see that the typical time interval between observations is between 0.3 and 2 days.

period–luminosity relations, such as RR Lyrae and Cepheids, and to identify parameters of different variable stars. The information gathered from these sources will be collected in the VVV Variables (V^4) catalog.

The structure of the paper is as follows. In Section 2, we present the photometry and calibration process on the VVV tiles and characteristics of the extracted time series. Then, in Section 3, the methodology is explained, where we focus on identifying irregular and periodic variables, mainly using different variability indices and periodograms. Next, in Section 4, the preliminary classification of the selected sources is presented. This classification is based on the principal properties of the sources, such as the shape of time series and light curves, and the period in the case of variable sources. We determined the general properties of selected variable sources as characteristic features of the light curves and locations on the color–magnitude and color–color diagrams. Also, we described the environment of the variable sources in the young stellar cluster candidates projected on VVV tiles d001 and d002. Finally, in Section 5, we present the catalog of variable sources in these tiles. The individual characterization of the variable objects is beyond the scope of this paper, and they will be analyzed in an upcoming work, once the follow-up spectroscopic analysis is completed.

2. The VVV Data

The VVV survey is an ESO Infrared Large Public survey (Minniti et al. 2010; Saito et al. 2012b) that uses the 4 m VISTA telescope located at Cerro Paranal Observatory, Chile. The survey was designed for mapping 562 deg^2 in the Galactic bulge and the southern disk in five near-infrared broadband filters—Z ($\lambda_{\rm eff} = 0.87$ μm), Y ($\lambda_{\rm eff} = 1.02$ μm), H ($\lambda_{\rm eff} = 1.25$ μm), J ($\lambda_{\rm eff} = 1.64$ μm), and K_s ($\lambda_{\rm eff} = 2.14$ μm)—with a time coverage spanning 5 yr between 2010 and 2015 in the K_s band. The telescope has a near-infrared camera, VIRCAM (Dalton et al. 2006), consisting of an array of 16 detectors with 2048 × 2048 pixels. A set of single exposures (a paw print) are combined into a tile, covering $1.5 \times 1.^\circ1$ in the sky. To cover the VVV area, the disk field was divided into 152 tiles and the bulge into 196 tiles (see Saito et al. 2012a for more details).

To test our method, we choose the first two VVV disk tiles, namely d001 and d002, due to their low crowding and interstellar reddening when compared to the rest of the VVV disk area. The preliminary reduced images were retrieved from the VISTA Science Archive[11] (VSA) database (Cross et al. 2012), keeping the quality flags. In total, we analyzed up to 55 and 41 K_s images for tiles d001 and d002, respectively. Figure 1 represents the log of observations of the VVV tiles.

2.1. Photometry and Calibrations

The PSF photometry was obtained using the `Dophot` software (Schechter et al. 1993; Alonso-García et al. 2012) in all available tile images in the field of view (FoV). We based this procedure on the method explained in Navarro Molina et al. (2016). We assessed the reliability of the photometry using the `Dophot` parameter chi, which quantifies the PSF quality. The sources with chi > 3 were rejected due to the large associated uncertainty. The calibration process to the VISTA system was done using the aperture photometry catalogs produced by the Cambridge Astronomical Survey Unit[12] (CASU). We selected sources with stellar ("−1") or borderline stellar ("−2") morphological classification to perform the cross-match using `STILTS` (Taylor 2006), using the catalog of the first epoch as reference with a $0.''34$ tolerance (VIRCAM pixel size). The conversion factors and uncertainties were estimated via a 2σ clipping linear fit to the `Dophot` PSF photometry versus the CASU isolated selected sources. By following this procedure, we have found 624,983 sources in d001 and 683,643 sources in d002 in common in the K_s band. The photometry in the J and H bands was performed using `Dophot` in a similar manner, i.e., using the CASU catalogs to calibrate the PSF photometry.

2.2. Cadence of the Observations

As has been pointed out, tiles d001 and d002 accumulate up to 55 and 41 epochs observed between 2010 and 2015, respectively. The left panel of Figure 1 shows the gaps, baseline, and maximum size of the time step between epochs. The right panel of Figure 1 shows the distribution of the difference of consecutive observations ΔMJD, zoomed up to ΔMJD > 12 days. The minimum time interval between the observation is ∼0.35 days, with distribution maximum between 0.35 and 2 days. Thus, from the time cadence of the observations, we can expect to detect variability related to timescale accretion variations, star spots, episodic accretion

[11] http://horus.roe.ac.uk/vsa
[12] http://casu.ast.cam.ac.uk

events, rotational modulation, and variable extinction in the YSOs (Rebull et al. 2014; Contreras Peña et al. 2017a). On the other hand, VVV produces unevenly spaced light curves, which provides its challenges, but we still expect to identify different types of periodic variability in a wide range of timescales (see, for example, Elorrieta et al. 2016; Gran et al. 2016; Minniti et al. 2017).

2.3. The K_s-band Time Series

The K_s-band time series of the sources were constructed by cross-correlation of all the catalogs for all available epochs. We filtered our initial sets of time series with some ad hoc quality and robustness criteria: (1) a minimum of 25 photometric measurements; (2) a total amplitude $\Delta K_s > 0.2$ mag, where $\Delta K_s = (K_s^{\max} - K_s^{\min})$; and (3) an upper limit in flux of $K_s > 11$ mag (to avoid objects that may suffer from saturation in some VVV epochs). The first restriction represents the minimum number of epochs that allows us to search for reliable periods. The second is motivated by a conservative estimation of the errors of photometry and transformation to the standard system. These initial filters reduced the source numbers obtained from photometry by approximately 30% (for example, from 669,825 to 433,102 for d001). Moreover, the photometric measurements are prone to be affected by systematic errors that are hard to clarify and quantify, given atmospheric or instrumental problems. For example, Alonso-García et al. (2015) reported a problem related to highly variable PSFs in the tile images due to different geometric distortions in the combination process of the paw-print images. Thus, to remove outliers in the time series, we implement the modified Thompson τ technique, which is based on the definition shown in Thompson (1985). The modified Thompson τ statistic is defined as

$$\tau = \frac{t_{\alpha/2}(n-1)}{\sqrt{n}\sqrt{n-2+t_{\alpha/2}^2}}, \quad (1)$$

where n is the number of data points, and $t_{\alpha/2}$ is the critical value of Student's t-distribution given a confidence parameter α. For each photometric data point K_s in a time series, the standard deviation of the time series $\sigma_{\rm ts}$ and the absolute deviation $\delta_i = |Ks_i - \overline{K}s|$ is calculated, where $\overline{K}s$ is the mean Ks magnitude. Individual photometric measurements were removed from time series when $\delta_i > \tau\sigma_{\rm ts}$ using a confidence level of 95% (i.e., $\alpha = 0.05$). One of the consequences of this approach is that we will remove poorly sampled transients event from our time series. Figure 2 shows the performance of this method acting on a time series.

3. Methodology

In time-series analysis, different sets of statistics, commonly called "variability indices," are frequently used to quantify changes in luminosity with time. Depending on the definition of these indices, a population of variables that have similar behavior can be identified, and one can try to separate stochastic variability from "further organized" flux variations. Examples of these indices are the Welch–Stetson $I_{\rm ws}$ (Welch & Stetson 1993) and Stetson $J_{\rm Stet}$ and $K_{\rm Stet}$ indices (Stetson 1996). These quantities have been used to identify sources that exhibit large photometric variations along the time. Defined in this manner, sources with larger $J_{\rm stet}$ values are the most probable variable sources. Different authors in the literature define

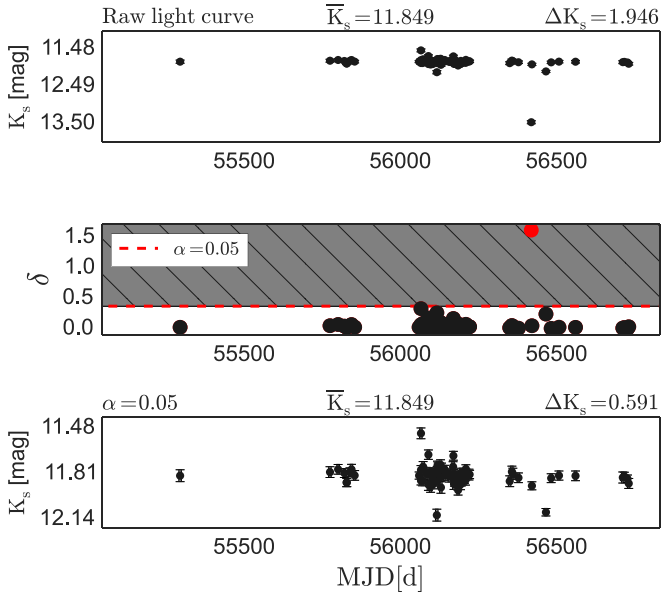

Figure 2. Performance of Thompson τ technique acting on a time series. Top plot: raw light curve extracted from the photometric process. Middle plot: absolute deviation δ of each measurement in function of MJD. The dashed red line indicates the rejection region (gray region) using $\alpha = 0.05$. The red point that falls in this region is removed from the sample. Lower plot: modified time series, which will be used in the analysis.

Table 1
Variability Indices Computed in This Analysis

Set of Variability Indices	
Index	Reference
η index	von Neumann (1941)
Stetson J	Stetson (1996)
Stetson K	Stetson (1996)
$\sigma_{\rm ts}/\mu$ ratio	Shin et al. (2009)
Classical χ^2	Rebull et al. (2014)
Total amplitude ΔK_s	Contreras Peña et al. (2017a)

particular limit values of $J_{\rm stet}$ for this task (such as $J_{\rm stet} \geqslant 0.55$, Carpenter et al. 2001; $J_{\rm stet} \geqslant 0.9$, Rebull et al. 2014). In the literature, it is possible to find many more different indices (or features) tailored to identify different types of variable sources. Thus, deciding which index is useful to detect a specific type of variability does not only depend on the definition of the index itself but also on the properties of the available data. Here we briefly summarize some variability indices applicable when only one photometric band is available and the type of variability that they can detect.

In this work, we considered a set of six variability indices (see the references in Table 1) in order to characterize the behavior of the variable sources along time. Mainly, we used the amplitude ΔK_s and the η index to identify irregular variables. All of these parameters are estimated directly from their K_s-band time series. All of the procedures and the automated process that we developed are summarized in the flux diagram of Figure 3.

3.1. Identifying Irregular Variable Sources

Eruptive pre-main-sequence (PMS) stars are traditionally classified as FU Orionis (FUors; Herbig 1966) and EX Lupi

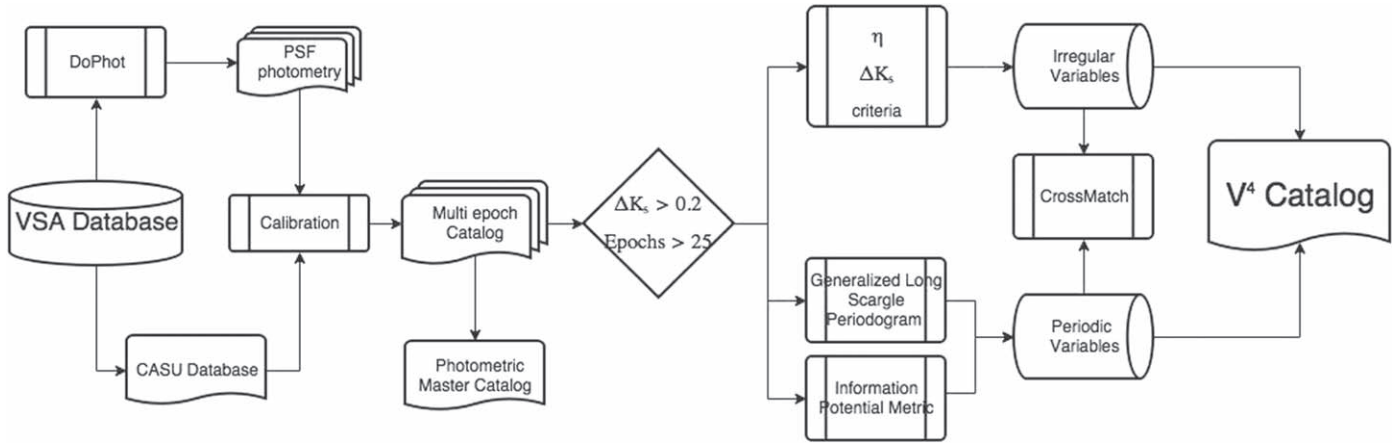

Figure 3. Schematic view of the automated process developed and used in this study to categorize the sources in the V^4 catalog.

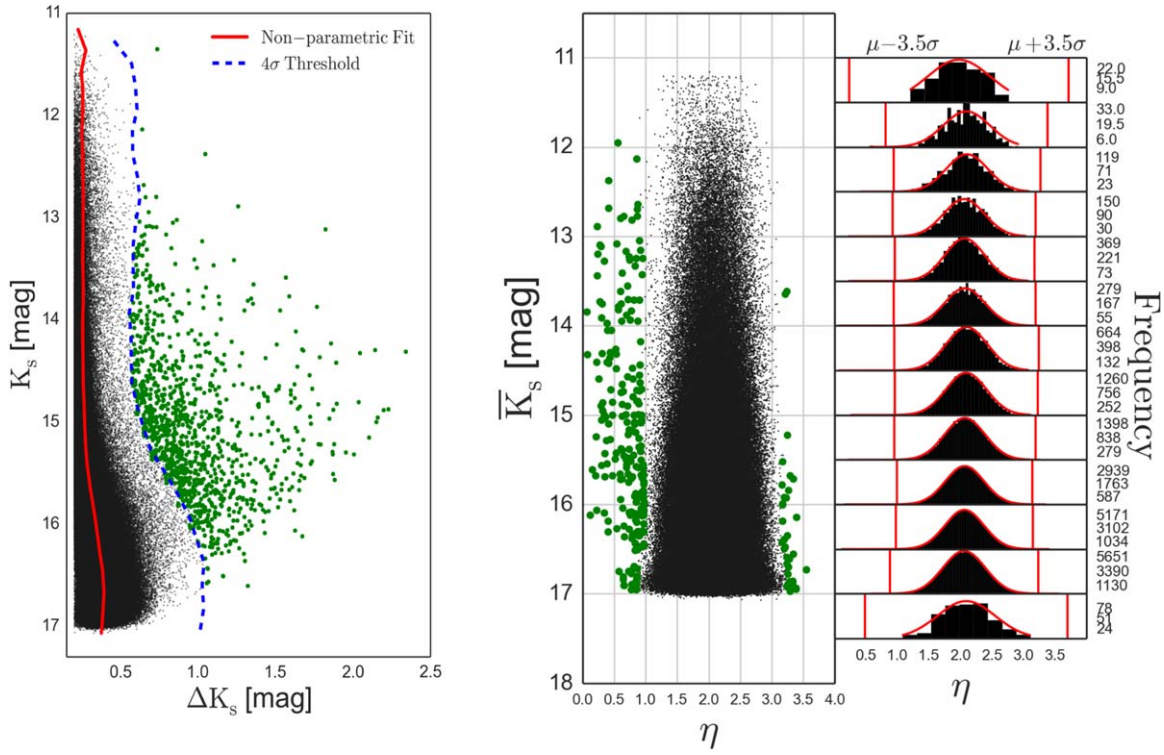

Figure 4. Representation of the two selection criteria aiming at identifying irregular variables. The dark points are all stars in the final photometric catalog, and the green filled circles highlight the selected sources by each method. Left panel: total amplitude ΔK_s selection. The red solid line represent the nonparametric fit, while the dashed blue line displays the 4σ threshold. Right panels: η index selection. The red lines in the bins of 0.5 mag represent the 3.5σ threshold of the Gaussian fit.

(EXors; Herbig 1989) types. Their variations occasionally have high amplitudes (up to 2–6 mag in the optical bands) in a short timescale. Different physical processes have been proposed to explain the variations of these objects, among them accretion or variable extinction induced by their circumstellar disks. These high-amplitude variables are potential tracers of a new generation of stars, so we expect to detect them within and close to the SFR of the Galaxy.

As an example, in Contreras Peña et al. (2017a), the total amplitude ΔK_s was used as a discriminant to identify variable sources with $\Delta K_s > 1$ mag. This method was very useful to help identify likely YSOs among irregular and periodic variable stars projected against SFRs. The "amplitude index" also performs satisfactorily in identifying sources with a large-amplitude ΔK_s, like some eclipsing binary (EB) systems and likely pulsating asymptotic giant branch (AGB) stars, as Miras and semi-regular sources, which have periods longer than $P \geqslant 100$ days and are frequently grouped under the name "long-period variables" (LPVs).

Following the same idea, we made a nonparametric fit on our photometric catalogs of d001 and d002. In order to quantify the behavior of ΔK_s as a function of mean magnitude $\overline{K_s}$, we measured the dispersion in bins and selected those above 4σ. This allowed us to define dynamical thresholds, taking into account that the estimated σ depends on the stellar population projected on the different tiles (for example, a tile containing a projected SFR will have a different threshold assigned than a tile with Population II stars). The left panel of Figure 4 shows the amplitude ΔK_s selection for the sources found in tile d001.

To complement the previously described criterion, we performed an additional selection using the η index (von Neumann 1941; Shin et al. 2009; Sokolovsky et al. 2017). This statistic is defined as the squared addition of successive differences between adjacent observations in a time series,

$$\eta = \frac{1}{(N-1)\sigma_{\rm ts}^2} \sum_{i=1}^{N-1} (m_{i+1} - m_i)^2, \quad (2)$$

where m_i are magnitude measurements, and N is the available number of epochs. The properties of the η index are well known for a stationary Gaussian distribution but not for astronomical time series, because they usually have an unequal sampling. For time series with uncorrelated photometric measurements (i.e., time series with uncorrelated normally distributed measurements), the η index would have a value of $\sim$2 and extreme values for series with long time variability trends. Given the aforementioned properties and the volume of the data, we expect that index η has a Gaussian distribution centered in $\eta \approx 2$. We separated $\overline{K_s}$ into 0.5 mag bins and considered 3.5σ confidence intervals on every bin to identify variable sources. The right panel of Figure 4 shows the η selection and the fit of the Gaussian functions in each histogram generated by separating the distribution in bins of 0.5 mag.

Each irregular variable should satisfy both ΔK_s, η criteria in order to be included in the paper as an irregular variable. All selected candidates have visual confirmation on the corresponding images; the candidates with close (less than $0\farcs4$) companions are removed.

3.2. Identifying Periodic Variable Stars

One of the main goals of the VVV Survey is to obtain a complete census of pulsating stars, such as RR Lyrae, Cepheids, semi-regular, and Mira variables across the Milky Way. These sources provide useful information from their quality as standard candles (using the developed period–luminosity relations in the near-IR) to determine the structure of our Galaxy. These stars are also useful to map the extinction affecting the projected areas. Another potential set of targets to be identified are the EB systems, which are known to provide the most robust/model free estimates of the fundamental stellar parameters.

In this study, we have implemented two methods to identify periodic sources within VVV data.

1. The generalized Lomb–Scargle periodogram (GLS; Zechmeister & Kürster 2009): a least-squares spectral analysis method based on the classical Lomb–Scargle periodogram (Lomb 1976; Scargle 1982). In particular, we used its implementation in the `astroML`[13] python library (Vanderplas & Ivezić 2007).
2. The informatic potential metric Q_m (IP metric; Huijse et al. 2011): a discriminant designed to identify the fundamental period of a time series using information theory. Within this framework, different sets of time series $\{x_n\}$ are assumed to be realizations of a continuous random variable X. The IP is defined as follows:

$$\mathrm{IP}_X(\{x_n\}) = \frac{1}{N^2} \frac{1}{\sqrt{2\pi}\,\sigma} \sum_{i=1}^{N} \sum_{j=1}^{N} \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2} \right). \quad (3)$$

We used a grid of trial periods P_t to fold the time series into phase space. The folded light curves are then segmented in H bins, and IP is computed for every bin (h). The IP metric Q_m is computed as the squared differences between the information potential of each bin and the global IP:

$$Q_m(P_t) = \frac{1}{H} \sum_{h=1}^{H} [\mathrm{IP}_X(\{x_n\}) - \mathrm{IP}_X(\{x_n\}_{n\in h})]^2. \quad (4)$$

To estimate the reliability of the periods found with both approaches, we calculated the statistical significance for the spectral power peaks in the GLS and the IP metric. Only objects with peak significance greater that 99.9% were considered for further analysis and characterization. We note that this formal peak significance assumes that the uncertainties are described by uncorrelated Gaussian noise.

3.3. Classification of Periodic Stars

To determine the variability type of the identified periodic stars, we consider the shape of the light curve and the period P determined by the methods previously described. A common tool to quantify the shape of a light curve is using templates, where the light curves of periodic sources are compared with templates to assign a well-defined variability class using one or a set of statistics to rely on this classification. These templates could be collected from public archives, literature, and other databases in order to create "training sets" that point to an automated classification. In this context, a significant (and still increasing) number of infrared light-curve templates have been assembled in the VVV Templates Project (Angeloni et al. 2014), where the main goal is to develop and test machine-learning algorithms for the automatic classification of VVV light curves. Nevertheless, we need to consider that in the NIR, RR Lyrae stars are a special case, given that the amplitudes of light curves decrease from optical to infrared wavelengths. This leads to it being more difficult to differentiate between RR Lyraes in the fundamental-mode (RRab) and first-overtone (RRc) subtypes using only templates. In this context, we used two criteria to classify these periodic sources.

3.3.1. Using the VVV Template Project

We expect low accuracies in the period estimation due to the relatively small number of epochs, as we pointed out in Section 2.2. With this in mind, we have used templates of those variables that we expected to find in this galactic longitude, such as classic Cepheids and eclipsing variables. The δ Scuti stars were not considered in this analysis, given that just two light-curve templates are available in the K_s band. Nevertheless, their periods are less than 0.2 days and may have similar characteristics to EBs (Dong et al. 2017). Several δ Scuti sources were identified around open clusters in the K_s band (Palma et al. 2016), so we need more information for a good characterization of this type of variable.

[13] http://www.astroml.org/

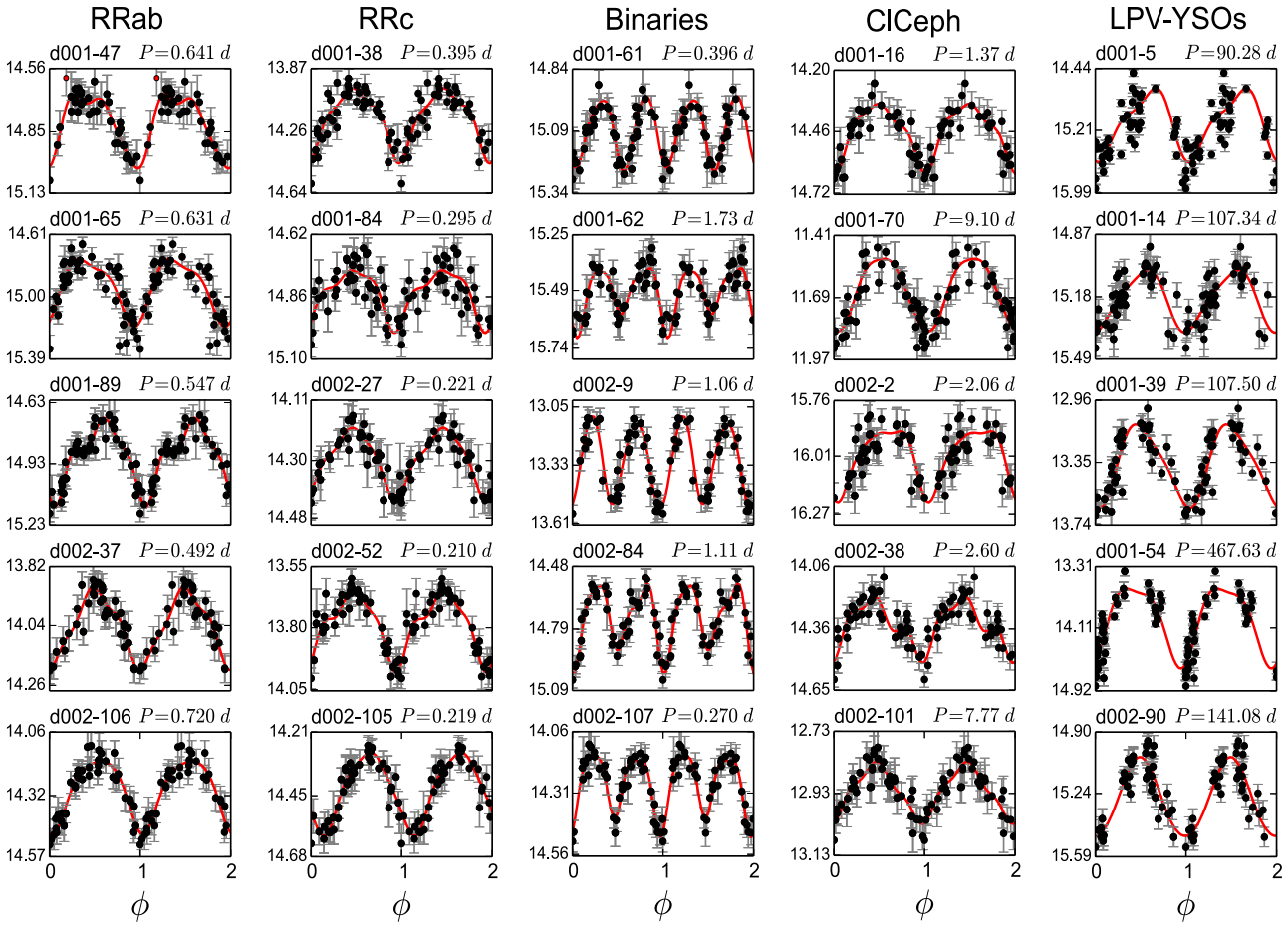

Figure 5. Examples of light curves of the periodic stars of the RRab, RRc, Cepheid, binary, and LPV types. The red points in the phase diagram are outliers of the Fourier fit. At the top of each plot, the identification and obtained period P are shown.

We used templates of classic Cepheids (103 templates in K_s with $0.97 < P < 133.90$) and EBs (76 templates in K_s with $0.305 < P < 16.092$) to compare with the light curves of our objects when the period P is contained in the indicated period range of the templates. To quantify these comparisons, we performed a Fourier fit of N harmonics into the phase space for each template and variable, given its period P and average magnitude $\langle m \rangle$ (see Figure 5). The amplitude A_k and phase ϕ_k for each k harmonic were determined. The Fourier series $f(t)$ at time t is given by

$$f(t) = \langle m \rangle + \sum_{k=1}^{N} A_k \sin\left(\frac{2\pi k t}{P} + \phi_k\right). \quad (5)$$

Each "synthesized" template was then normalized by subtracting the integral of the obtained Fourier series. The harmonic number N used to fit the model to periodic sources is $N = 4$. The periodic stars are classified using the template that had the best goodness of fit, using the reduced $\chi^2_{\rm red}$ statistics as the criterion. If a source has $\chi^2_{\rm red} > 1$, it will remain as not classified.

3.3.2. Classifying RR Lyrae Stars

The RR Lyrae stars can be subclassified by their locations in the Bailey diagram (Bailey 1902), given that RRc sources have shorter periods than the RRab type. Figure 5 of Gavrilchenko et al. (2014) shows that the RRab and RRc categories are located in different places on the Bailey diagram. They also discuss the arbitrary limit of $P = 0.4$ days to discriminate between these categories. Using this, we define the RRc region ($0.2 < P < 0.4$ days) and RRab region ($0.4 < P < 1$ days), and they are shown in Figure 8.

4. Application of the Automated Process on d001 and d002 VVV Tiles: The V^4 Catalog

4.1. Irregular Variable Sources

We had identified 72 variable sources that fulfill both the ΔK_s and η criteria. If a source presents periodicity, it is removed from the sample and added to the periodic sample. This was the case for two sources (d001-79 and d002-103) that present a large amplitude and period, typical signatures of dust-enshrouded AGB stars, invisible in the optical range, due to its thick circumstellar-envelope product of its high mass-loss rate.

Finally, we identified 70 irregular variable sources, 45 of them belonging to d001 and 25 to d002. Almost two-thirds of them (64.28%) are projected in the d001 tile and follow the cold gas/dust distribution as traced by the W3 ($\lambda_{\rm eff} = 12\ \mu$m) band *WISE* image (the background of Figure 6). The highest overdensity is observed at the borders of the SFR, where the nebulosity is overwhelming, thus suggesting active star formation. All objects in the sample are, to the best of our knowledge, reported here for the first time. Some of these sources are shown in Figure 7.

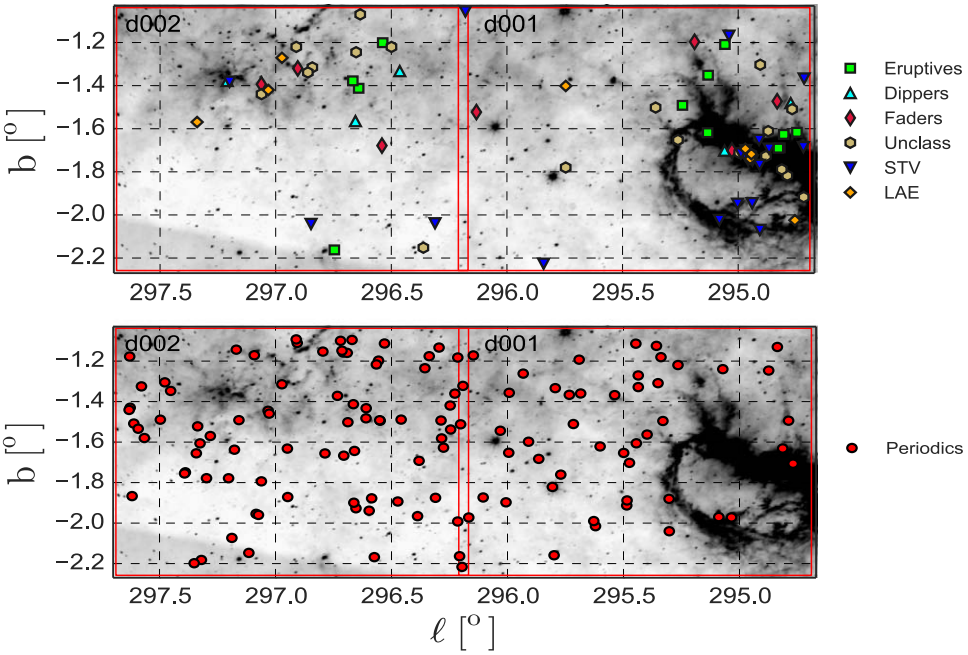

Figure 6. Region covered by the d001 and d002 tiles from the VVV survey, showing the spatial distribution of identified sources. Galactic north is up, galactic east is to the left. The symbols represent different types of variable stars found in the FoV and are explained in Section 4. In the background, the W3 ($\lambda_{\rm eff} = 12\ \mu$m) band *WISE* image is shown to illustrate the cold gas/dust distribution in the FoV.

The sample can be separated in two groups.

1. Short-term irregular sources. As discussed in Section 2.2, the cadence of VVV can reveal objects with large changes in their magnitudes in short periods of time, often associated with YSOs or PMS stars. Examples of such time series are shown in Figure 7. As noted previously, intrinsic changes in the sources can be explained by variable accretion (see, e.g., Meyer et al. 1997; Cody et al. 2014; Rebull et al. 2014) referred to as bursts if there is a brief, well-defined event and then a return to quiescence. Variable extinction is also possible. The individual characterization of the objects is beyond the scope of this paper, and they will be analyzed in an upcoming work, once the follow-up spectroscopic analysis is completed.
2. Long-term irregular variables. These kinds of sources have a slow change of their magnitudes over the time series, reaching large amplitudes ΔK_s in longer time intervals. In general, the sources do not exhibit large-amplitude changes in short time intervals but increase/decrease their luminosities monotonically and then, in certain cases, return to their mean magnitude. Possible mechanisms here are the eruptive episodes or long-term extinction events, followed by quiescent periods. These time series can also reveal stellar sources such as supernovae, microlensing events (Minniti et al. 2015), LPVs, aperiodic long-term variability objects, and even extragalactic variable sources such as quasars. Examples of long-term irregular time series can be seen in the right panels of Figure 7.

In order to analyze the morphological behavior of VVV irregular variable sources, Contreras Peña et al. (2017a), influenced by previous works such as Findeisen et al. (2013), proposed the following classifications for their sample of high-amplitude variables ($\Delta K_s > 1$ mag): faders, dippers, short-timescale variables (STVs), and eruptives. However, half of our irregular variable sample has an amplitude $\Delta K_s < 1$ mag (right panel of Figure 9). Similar variable sources with low amplitude have been reported in the literature (see, for example, Carpenter et al. 2001; Wolk et al. 2013) with light curves similar to those reported in this work. Therefore, given that the classification for irregular variables is defined mainly by the shape of the time series, we extended the classification proposed in Contreras Peña et al. (2017a) for low-amplitude sources ($\Delta K_s < 1$ mag), making an exception with the eruptive classification, which describes sources with eruptions on timescales of hours to days. Sources that present eruptive long-timescale variability ($t > 1$ yr) and $\Delta K_s < 1$ mag will be classified as "low-amplitude eruptive" (LAE). We used this classification to characterize our sources when applicable. In Table 2, the main characteristics of the different proposed classes are explained.

4.2. Periodic Variable Stars

Our automated tool detected 22 periodic variable sources with the GLS analysis in both the d001 and d002 tiles, with a range of periods of $4.03 < P < 1400$ days. Given the cadence of our data and the limitations of the GLS, we cannot obtain reliable periods shorter than 2 days. The IP metric, on the other hand, shows a great performance identifying periodic sources over the entire range of periods. Several stars have been detected by both methods, showing practically identical periods within the uncertainties. Thus, taking into account that the IP metric is computationally expensive, we used this method to search for shorter periods, creating a grid of periods between 0.05 and 5 days. The IP metric identified 108 sources with $0.2 < P < 3.1$ days.

Figure 5 shows examples of the periodic variables, while in Figure 8, we show the period–amplitude diagram. As can be seen, most of the stars have short periods, typical for RR Lyrae stars.

In some cases, it was not possible to distinguish between different variability types—for example, LPVs and YSOs (Figure 5)—so we put a mixed classification, LPV-YSO, and are working on additional (spectroscopic) data to clarify.

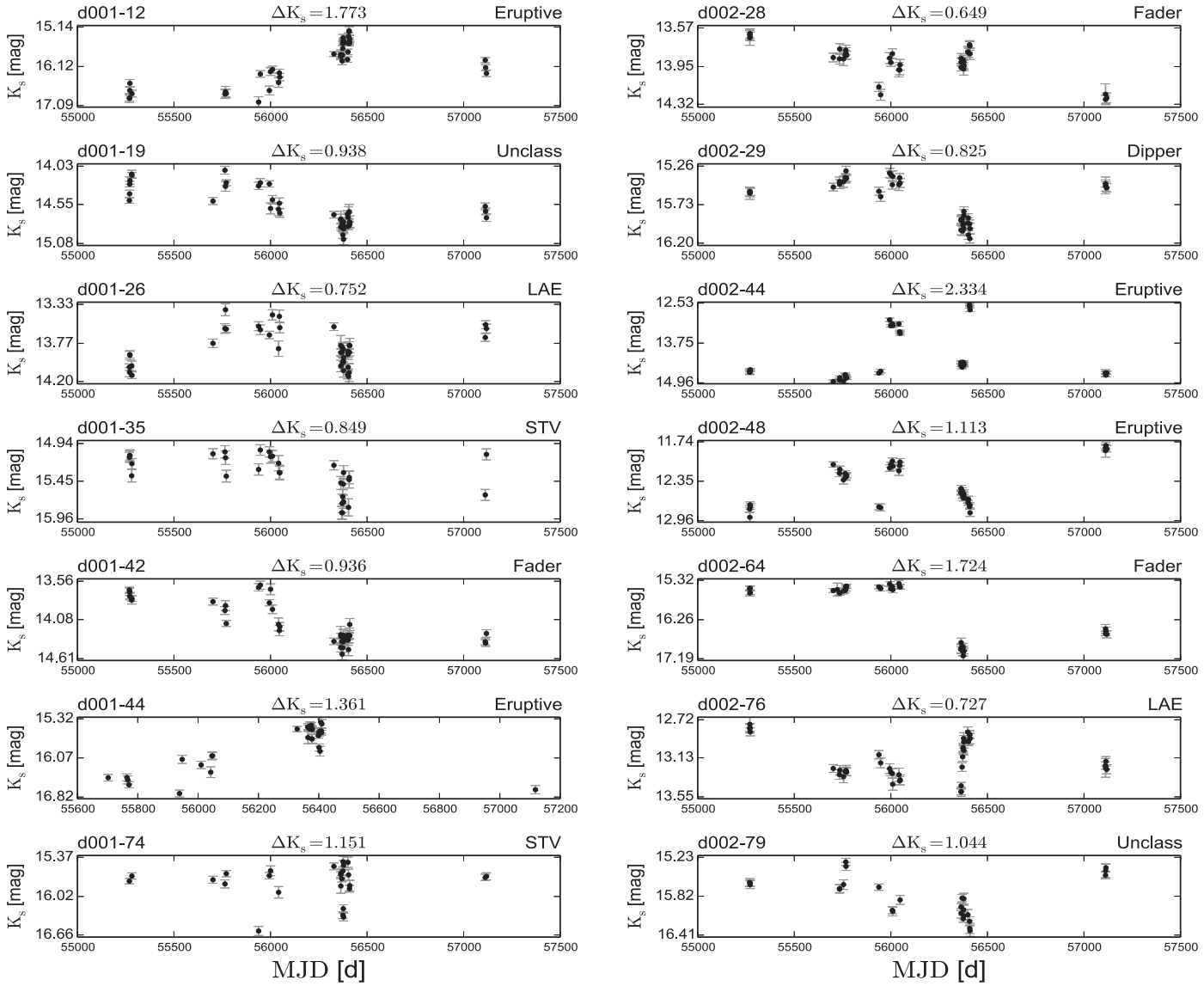

Figure 7. Examples of K_s-band time series of irregular variables. At the top of each plot, the identification, mean magnitude $\overline{K_s}$, amplitude ΔK_s, and variable type are shown.

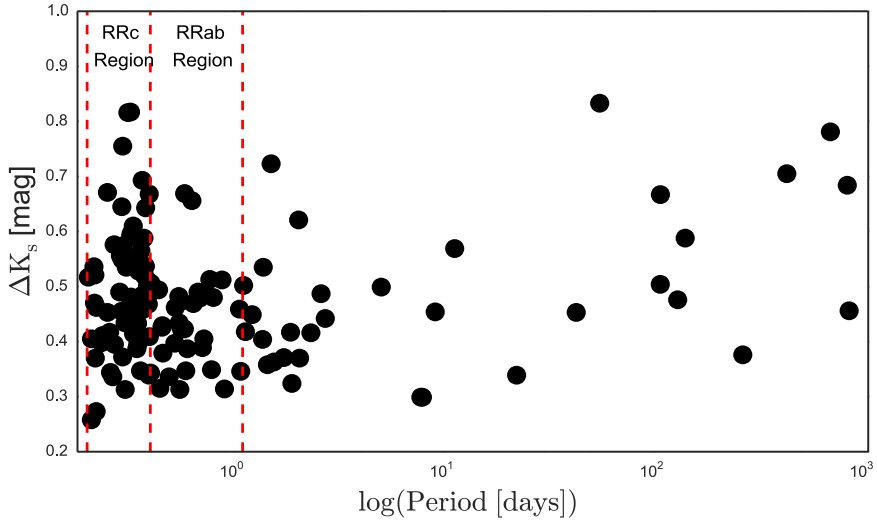

Figure 8. Period–amplitude diagram of periodic variables in our catalog.

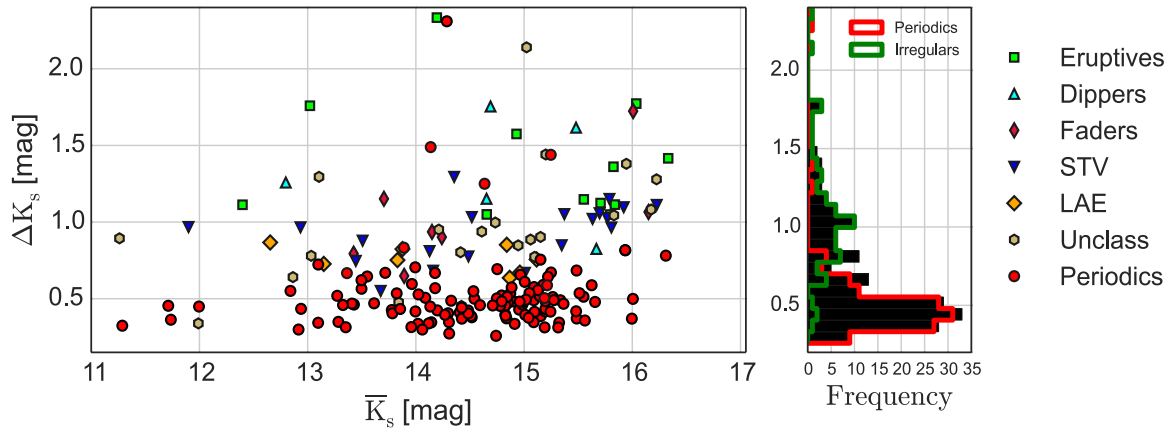

Figure 9. Left: the ΔK_s distribution of the selected sources vs. $\overline{K}_s$. Right: histogram of the ΔK_s distribution. In this figure, ΔK_s is truncated at 2.3 mag, given that the source d001-79 has a much larger amplitude than other sources ($\Delta K_s = 3.2$ mag). The symbols are the same as in Figure 6.

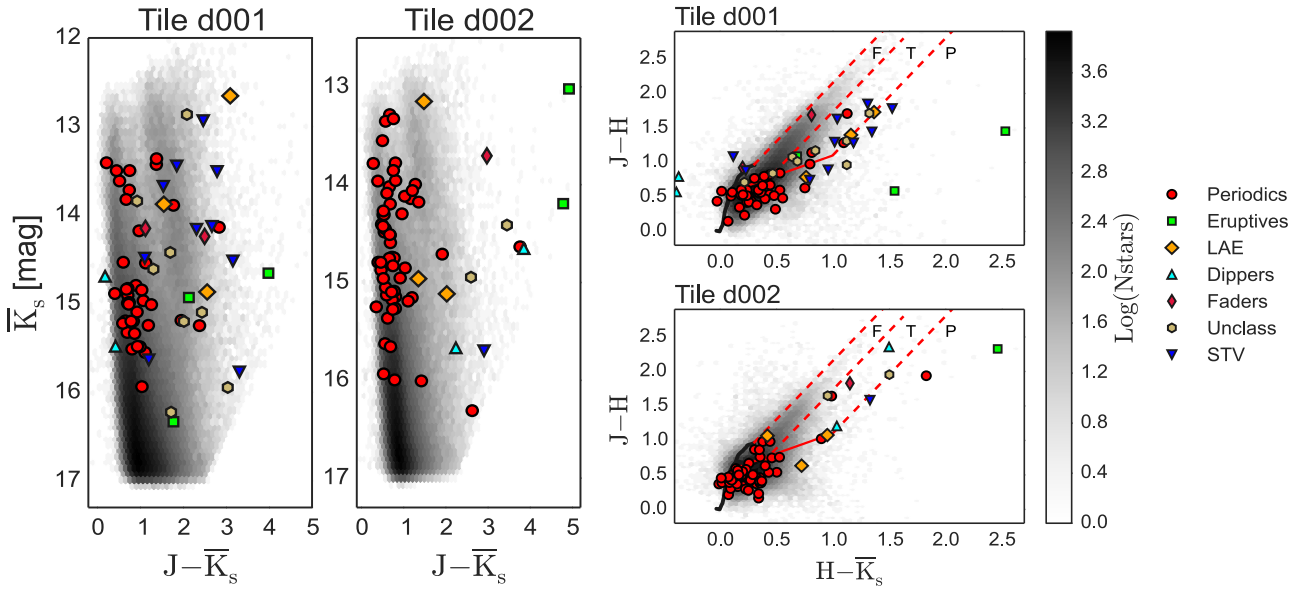

Figure 10. Color–magnitude and color–color diagrams of all variable stars in d001 and d002. The solid black lines are the intrinsic colors of the dwarf and giant stars from Bessell & Brett (1988), the solid red line is the locus of the unreddened CTTs (Meyer et al. 1997), and the dashed red lines are the reddening vectors of the early spectral type dwarfs, giant stars, and unreddened CTTs, assuming a visual extinction $A_V = 15$ mag. The symbols are the same as in Figure 6.

Table 2
Characterization of the Irregular Variables in the Categories Proposed by Contreras Peña et al. (2017a)

Class	Description
Dippers	Sources that show fading events and then return to their normal magnitude.
Eruptives	Sources that show outbursts with amplitude >1 mag and duration longer than a few days and typically at least a year.
LAEs	Sources that present outbursts with amplitude lower than 1 mag and duration typically longer than a year.
LPV-YSOs	Sources with a measured period but short-timescale scatter in the time series.
STVs	Sources with fast and constant scatter in their time series. They can also show brief rises in magnitude in timescales of weeks.
Faders	Sources that show a continuous decrease in brightness ($t > 1$ yr), or a big decrease in brightness in a source with relatively constant luminosity.

4.3. General Properties of Selected Sources

Figure 9 shows the ΔK_s distribution of the selected sources as a function of the mean $\overline{K}_s$ magnitude. The amplitude interval is in the range 0.5 mag $< \Delta K_s <$ 3.2 mag, with an average value of $\langle \Delta K_s \rangle = 1.1$ mag. The histogram is influenced mainly by the periodic sources.

Figure 10 shows the position of the variable stars in the color–magnitude and color–color diagrams. To plot the nonvariable stars, we used the first epoch of J- and H-band

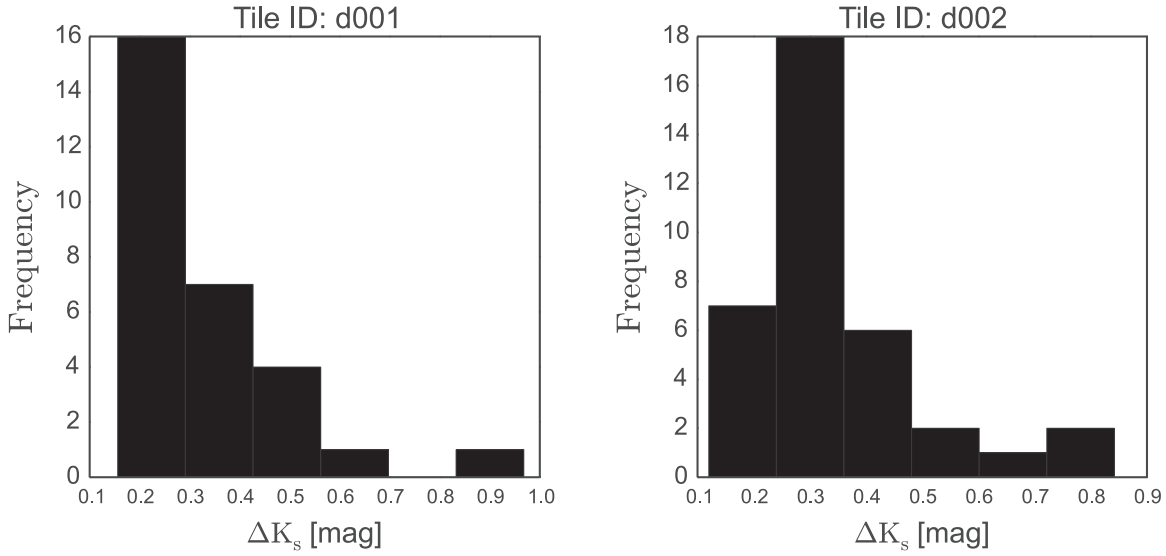

Figure 11. Histograms of the amplitudes of know variable stars from the VSX for d001 and d002.

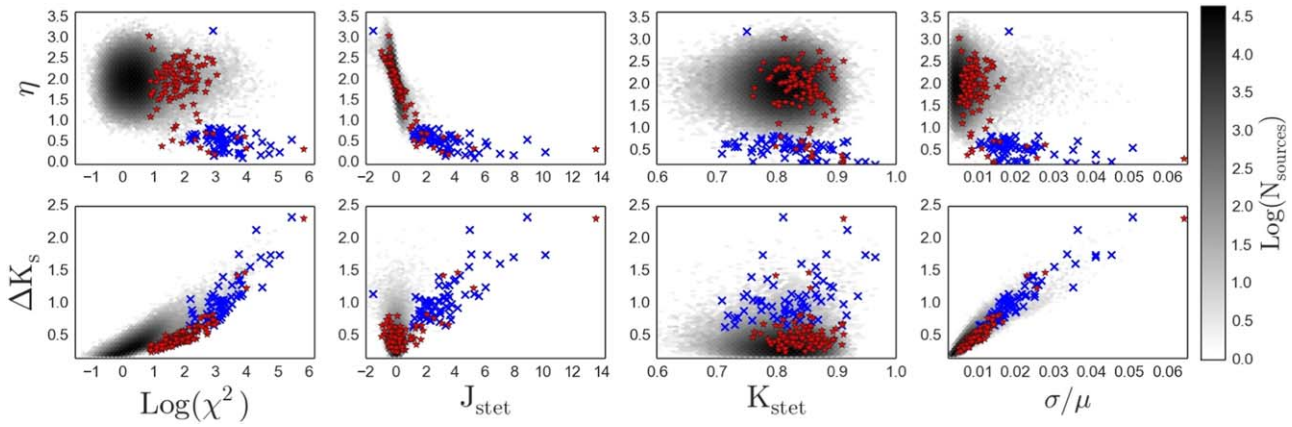

Figure 12. Correlation plots using the main features in this analysis. The blue crosses are the irregular variables, and the red stars are the periodic sources.

tile images taken in 2010. To analyze this, we followed the method described in Ojha et al. (2004). Three regions are defined. The "F" region is located between the reddening vectors of the giant and dwarf stars. The "T" region is between the reddening vector of the giant stars and the classical T Tauri star (CTT) locus, where Class II YSOs and Herbig Ae/Be stars (Hillenbrand et al. 1992) can be identified. In the so-called "P" region, located below the reddening vector of the CTTs, the likely protostellar objects are situated. Thus, the corresponding variability types are assigned. Row 32 of Table 5 contains information on the region in the color–color diagram for each source.

4.4. The Completeness and Accuracy of the Catalog

The catalog is limited to sources brighter than $K_s = 11$ mag, given the saturation of the limit of VVV. We are only sensitive to the stars fainter than this magnitude limit. In the International Variable Star Index Catalog[14] (VSX) are listed 57 variables in d001 and 46 in d002 fainter than $K_s = 11$ mag. From these, 30 sources in d001 and 36 in d002 are identified in the photometry catalog, but all of them have a typical ΔK_s around 0.2–0.3 mag, which is close to our conservative amplitude limit described in Section 2.3. This range of amplitudes is relatively low in comparison to the average amplitude, $\langle \Delta K_s \rangle = 1.1$ mag, of the selected variable sources (see Figure 11). Taking into account that these variables are detected in the optical wavelengths, the amplitudes are too low to be detected in the K_s band with our searching method. Four sources from the VSX are recovered in our catalog, namely d001-20, d001-81, d002-8, and d002-60; the rest of the stars with amplitudes greater than 0.3 have less than 25 measurements. Thus, in the magnitude interval 11 mag $< K_s <$ 15.5 mag, we recovered only 6% of the known optical variable stars, and more than 90% of our discoveries are new. In the context of previous K_s-band studies, Minniti et al. (2017) and S. Eyheramendy et al. (2018, private communication) reported 1 and 13 RRab stars in d001 and d002, respectively. The Minniti et al. (2017) RR Lyrae star, d002-0143595, has been rediscovered in our catalog as source d002-20, with practically the same period $P = 0.456794$ (with a discrepancy lower than 0.01%). Twelve of the 13 RR Lyrae stars from the Eyheramendy et al. list are rediscovered in our catalog. The only source missed by our procedure has $\Delta K_s = 0.195$ mag and is rejected from our initial conditions. Thus, the completeness of the catalog 11 mag $< K_s <$ 15.5 mag is very close to

[14] https://www.aavso.org/vsx/index.php?view=search.top

Table 3
Basic Information on the Star Cluster Candidates in the Region

Tile ID	Name	R.A.$_{2000}$ (h:m:s)	Decl.$_{2000}$ (d:m:s)	ℓ (deg)	b (deg)	Radius (arcmin)	$E(J-K)$ (mag)
d001	VVV CL005	11:38:59	−63:28:42	294.9481	−1.7353	0.50 ± 0.15	1.4 ± 0.3
	La Serena 001	11:39:13	−63:29:04	294.9726	−1.7292	0.52 ± 0.10	2.8 ± 0.6
	La Serena 002	11:39:22	−63:28:11	294.9896	−1.7188	0.52 ± 0.17	3.1 ± 0.4
	La Serena 003	11:40:28	−63:27:58	295.1026	−1.6779	0.48 ± 0.09	3.2 ± 0.5
	La Serena 009	11:45:04	−63:17:44	295.5571	−1.3788	0.72 ± 0.10	2.2 ± 0.3
d002	VVV CL007	11:53:50	−64:20:28	296.7463	−2.1629	0.33 ± 0.08	2.2 ± 0.2
	VVV CL008	11:55:29	−63:56:24	296.7611	−1.7328	0.42 ± 0.10	1.4 ± 0.2
	VVV CL009	11:56:03	−63:19:00	296.8331	−1.1105	0.60 ± 0.20	1.0 ± 0.1[a]
	La Serena 015	11:55:23	−63:25:30	296.7131	−1.2336	0.33 ± 0.07	1.4 ± 0.2

Note.
[a] Chené et al. (2013).

90%. We are expecting drops of the completeness for the fainter than $K_s > 15.5$ mag objects, but it is hard to estimate due to a lack of literature data.

Figure 12 presents the main features η (top row) and ΔK_s (bottom row) as a function of different variability indices used in this article (see Table 1). The distribution of all detected sources in tile d001 is shown in the background. The red stars represent the periodic sources, and the blue crosses are the irregular ones. As can be seen from the figure, the irregular variable sources are well separated from the distribution and follow different trends in the plots. The source d001-75 is an exceptional case, having an amplitude of $\Delta K_s = 1.151$ mag, $\eta = 3.193$, and a low number of observations (26 epochs). Excluding d001-75, the irregular sources fulfill $J_{\rm stet} > 0.95$, which agrees with the limit used in Rebull et al. (2014). In the case of the periodic sources with short periods, in general, they are located in the place where the main distribution is contained given the apparently uncorrelated shape of its time series, produced by the lack of observations.

4.5. Variable Stars around Open Clusters in the d001 and d002 Tiles

Nine open-cluster candidates are projected in the FoV of d001 and d002, namely VVV CL005, 007, 008, and 009 (Borissova et al. 2011) and La Serena 001, 002, 003, 009, and 015 (Barbá et al. 2015). The coordinates of the clusters are listed in Table 3, and the VVV color images of the clusters in d001 are shown in Figure 13 for illustration. Little is known about their properties, except for VVV CL009, which has been investigated in Chené et al. (2013) and Hervé et al. (2016). According to these papers, VVV CL009 is a young (4–6 Myr), moderately massive stellar cluster (total mass greater than 1000 $M_\odot$) containing at least an OIf/WN7 and two O8-9V stars. The cluster distance of 5 kpc is estimated by Chené et al. (2013) using spectroscopic parallaxes. In the context of this paper, we search for variable YSOs around these clusters that could be cluster members. To do this, we use the shape of the light curves, their position on the color–magnitude diagrams, the projection within the visual diameter of the clusters, and proper-motion diagrams taken from Smith et al. (2018).

To estimate the projected cluster radius, we combined the existing VVV K_s-band images (55 in the case of d001 and 41 for d002) using the standard `IRAF` procedures, and then the PSF photometry with `Daophot` in `IRAF` was performed. The obtained magnitudes were transformed to the 2MASS system using common stars. The details of these procedures can be seen in Borissova et al. (2011, 2014). These photometric catalogs were used to construct the stellar surface-density maps by performing direct star counting in the K_s band with a 5″ bin radius, assuming spherical symmetry. The maps are normalized by the area of the rings to determine the stellar density. The resulting spatial distribution maps of the stellar surface density are shown in Figure 14.

As can be seen from Figure 14, the overdensity of the stars is clearly visible, and the density peaks are at least 3 times higher that the surface density of the surrounding fields, thus confirming the cluster/group nature of the candidates. The cluster boundary was determined by fitting the theoretical profile presented in Elson et al. (1987). The obtained visual radii of the clusters are listed in Table 3, where the errors correspond to uncertainties from the model fit.

The procedure employed for determining fundamental cluster parameters such as age, reddening, and distance is described in Borissova et al. (2011, 2014) and Chené et al. (2012, 2013). Briefly, to construct the color–magnitude diagram, we perform PSF photometry of the 5′ × 5′ J, H, and K_s fields surrounding the selected candidate using the `Dophot` pipeline. Data for saturated stars (usually $K_s \leqslant 11.5$ mag, depending on the crowding) were replaced by data from the 2 MASS Point Source Catalog (Skrutskie et al. 2006). Since 2MASS has a much lower angular resolution than VVV, when replacing stars, we carefully examined each cluster to avoid the contamination effects of crowding using the Point Source Catalog quality flags available in the 2MASS catalog.

To separate the field stars from probable cluster members, we used the field-star decontamination algorithm of Bonatto & Bica (2010). The algorithm divides the K_s, $(H-K_s)$, and $(J-K_s)$ quantities into a grid of cells. For each cell, the algorithm estimated the expected number density of member stars by subtracting the respective field-star number density. Thus, each grid setup produced a total number of member stars $N_{\rm mem}$. Repeating the above procedure for different setups, we obtained the average number of member stars. Each star was ranked according to the number of times it survived after all runs (survival frequency), and only the $N_{\rm mem}$ highest-ranked stars were taken as cluster members. For the present cases, we obtained survival frequencies higher than 85%. To additionally clean up the diagrams, we used the relative proper-motion

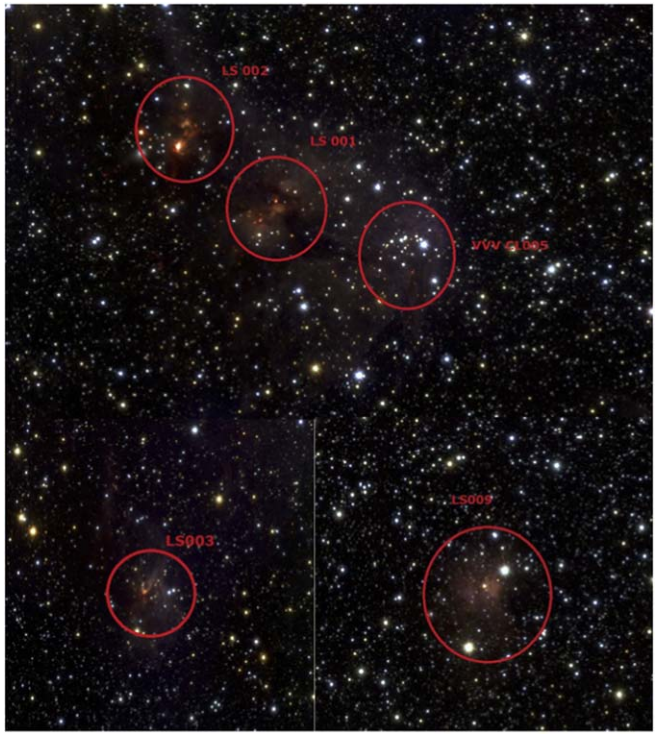

Figure 13. False-color VVV images of star clusters projected in d001. North is up, and east is to the left. The red circles represent the visual diameter of the clusters as defined in the text (see below).

catalog recently constructed by Smith et al. (2018). In general, the cluster members should form a clearly visible overdensity with respect to field stars in the proper-motion diagram. In our case (see Figure 15), it is impossible to separate the cluster members from the field stars, because the cluster members closely follow the motion of the Galactic disk. Nevertheless, they mark compact groups slightly shifted from the disk population. To calculate the radius of the group, we started from the photometrically decontaminated candidates, calculated the mean proper motion and its error (quadratically adding to this error the mean of the individual proper-motion errors), and drew the circle with 3σ radius (blue circle in Figure 15). Thus, the stars with motions projected farther than 3σ from the circle are rejected.

Taking into account that our candidates are classified as young clusters in the discovery papers of Borissova et al. (2011) and Barbá et al. (2015), the photometry/astrometry alone cannot give accurate distance and age determinations. Usually, spectroscopic parallaxes from follow-up observations of selected members are needed. Here only VVV CL005 has one follow-up object observed, thus it was impossible to obtain accurate basic parameters of the clusters. Instead, we use the PARSEC isochrones compilation for solar metallicity and ages of 5 and 400 Myr (Bressan et al. 2012; Marigo et al. 2017) to illustrate the position of the most probable cluster members and estimate their mean reddening (last column in Table 3).

4.5.1. Notes on Individual Clusters: VVV CL005

The cluster VVV CL005 is a young star cluster candidate, defined in Borissova et al. (2011) as a small group of 24 stars; it is projected close to the IC 2944 H II region and to the IC 2948 cluster ($3\farcm8$), on the part of the cloud [SMN83] Lam Cen 1. The brightest star within the cluster radius, namely HD 308829, is classified as a Be star of B8 spectral type and is suggested to be a member of the IC 2944 cluster (Cl* IC 2944 THA 51). The star is found to be a periodic variable star with $P = 0.8709$ days (Pojmanski 1998), but we cannot follow its variability with VVV, because the star is saturated in our images ($K = 9.68$ mag). The published distance to the IC 2944 H II region varies from 1.8 kpc (McSwain & Gies 2005) to 2 kpc (Sana et al. 2011) in the literature. The color–magnitude diagram of VVV CL005 (Figure 16) shows a poorly populated main sequence and some reddened stars, suggesting indeed a very young stellar group. With regard to the distance, it is impossible to estimate it from comparison with the theoretical isochrones, because the isochrones for ages younger than 5 Myr in this mass interval in the near-infrared bands are vertical and practically identical. Then, the spectroscopic distance is calculated using the spectral classification of Obj1 (R.A. = 11:38:57.73 and decl. = −63:28:22.4). The object is observed with the Astronomy Research Cornell Infra Red Imaging Spectrograph (ARCoIRIS). This is a cross-dispersed, single-object, long-slit, infrared imaging spectrograph mounted on the Blanco 4 m telescope, CTIO. The wavelength range is from 0.80 to 2.47 μm, with a spectral resolving power of about 3500 Å. The comparison with different spectral templates taken from VOSA (Bayo et al. 2008) gives the most probable spectral type as F4-F6V, used to estimate the spectroscopic parallax. We calculated a reddening and distance modulus of $E(J - K_s) = 1.4 \pm 0.3$ and $(M - m)_0 = 10.45 \pm 0.43$ mag (1.23 ± 0.24 kpc), respectively. The distance is comparable with the distance estimates of 1 kpc as measured from interstellar silicon monoxide (SiO) sources (Harju et al. 1998). Five variables (d001-25, 27, 28, 29, and 30; see Table 5) are probable cluster members, taking into account their projected position radius from the cluster center and the position in the proper-motion and color–magnitude diagrams (Figures 15, 16). All of them show irregular variability with relatively large amplitudes (0.63 mag $< \Delta K_s <$ 0.88 mag) and thus can be classified as YSOs.

4.5.2. La Serena 001

The star cluster candidate La Serena 001 (LS001) is projected very close to VVV CL005, on the same H II region, and contains a few very reddened stars (see Figure 16). These objects are deeply embedded in dust and gas. We adopted the same distance as for CL005 and, using the 5 Myr isochrone, determined $E(J - K) = 2.8 \pm 0.6$. Note the large uncertainty of the reddening determination, which can be a result of a strong differential reddening inside the region. One variable star (d001-32) is found in the vicinity of this group in formation. It shows an irregular time series and $\Delta K_s = 0.96$ mag. Taking into account the time series and its position on the color–magnitude diagram (in the "P" region), most probably it is a YSO candidate.

4.5.3. La Serena 002

The star cluster candidate La Serena 002 (LS002) is projected close to LS001 in the north east direction. Several embedded and very red sources are visible, indicating stars in formation. The color–magnitude diagram is poorly populated; thus, as in the case of LS001, we determined a mean reddening of $E(J - K) = 3.1 \pm 0.4$, adopting the same distance modulus as above. One periodic variable star (d001-38) is

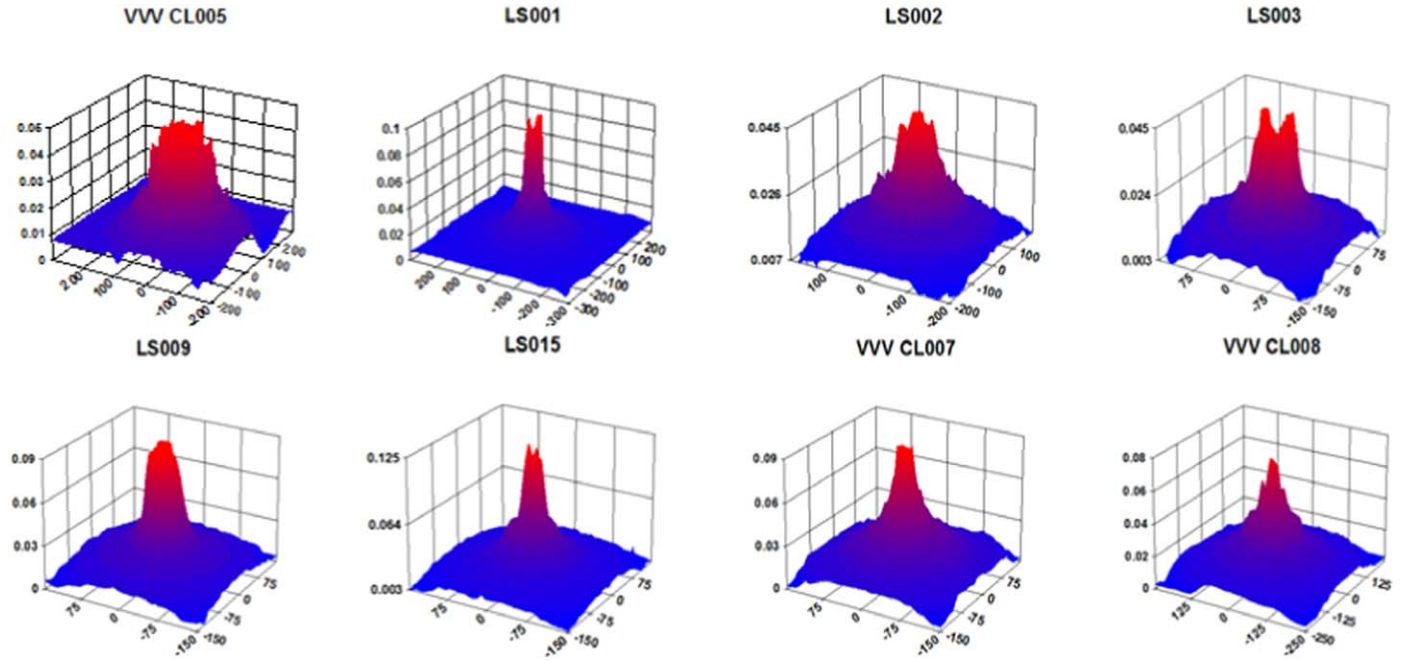

Figure 14. Stellar surface-density maps σ (stars arcmin^{-2}) of the clusters.

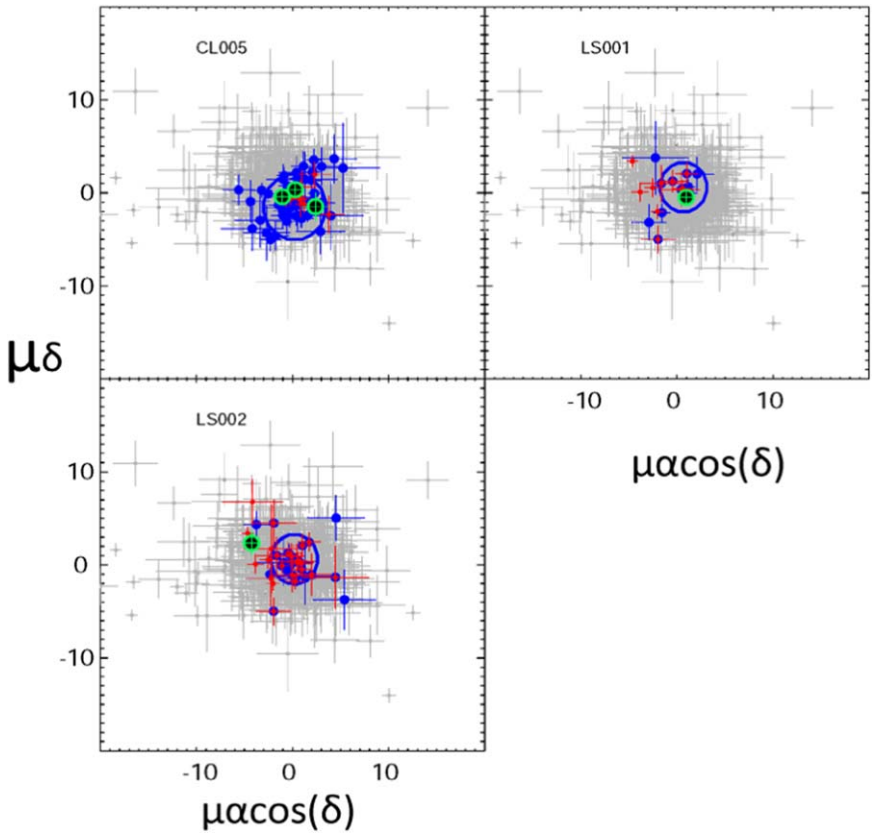

Figure 15. Relative proper motion of CL005, LS001, and LS002. The gray points are all stars in a $5' \times 5'$ area. Blue points stand for the most probable cluster members, obtained after a statistical photometric decontamination procedure; red points are the emission candidates (see text); and green points are the variable stars. The blue circles mark the suggested area of cluster members.

found, with $\Delta K_s = 0.453$ mag and period $P = 42.706$ days, which we classified as a Cepheid. Additionally, this source is far from the main locus of most probable cluster members in the color–magnitude and proper-motion diagrams; thus, we conclude that is a projected field star. Two more irregular variables (d001-35 and d001-36; see Table 5) with amplitudes in the range 0.84 mag $< \Delta K_s <$ 0.91 mag are most probably cluster members and YSO candidates.

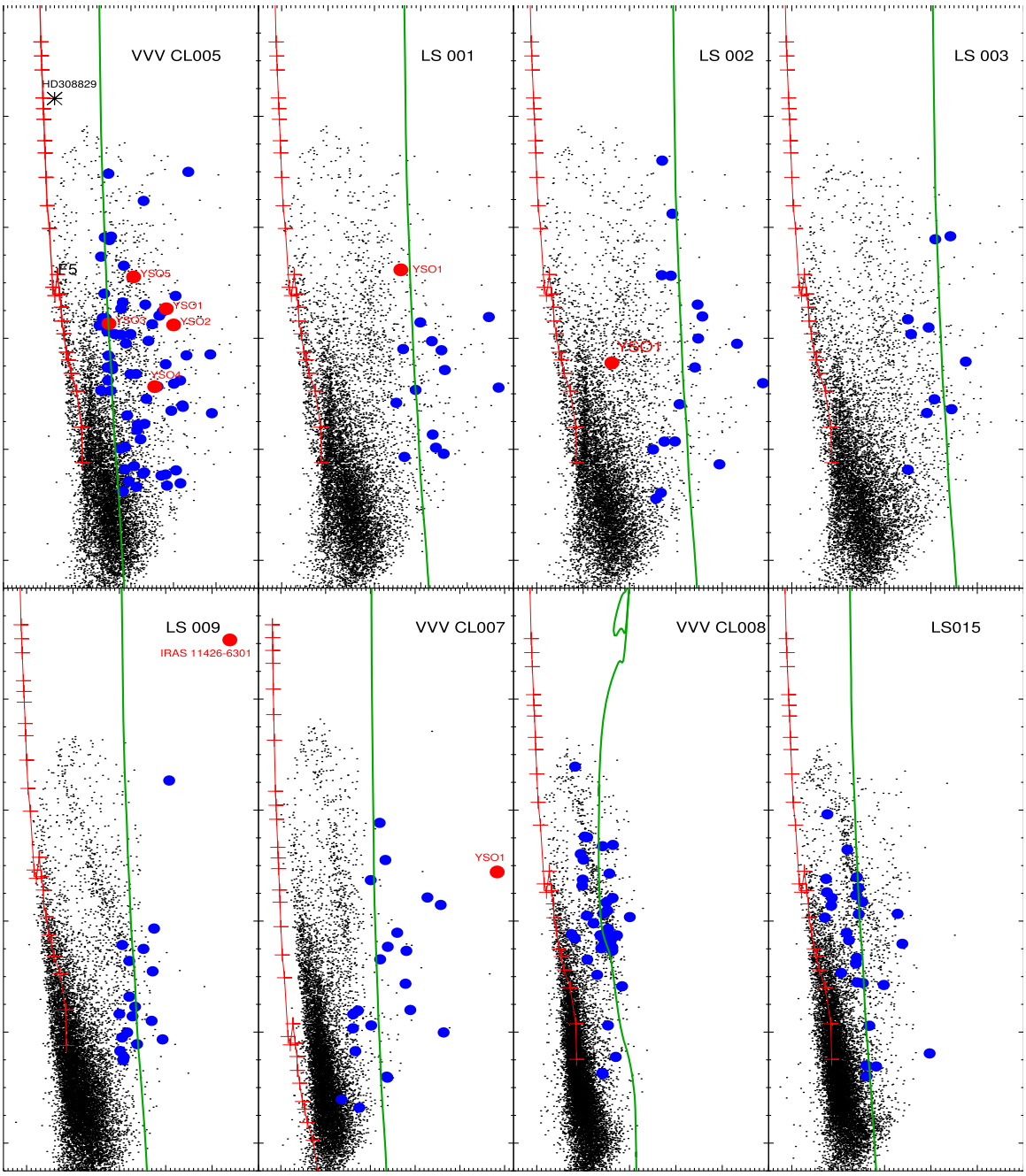

Figure 16. K_S vs. $(J - K_S)$ color–magnitude diagrams for the clusters in the FoV of d001 and d002. Black points are stars in a $5' \times 5'$ field around the cluster centers, and blue circles are the most probable cluster members after statistical photometric decontamination. Red circles are variable stars (see text). The Geneva isochrones of 5 Myr (400 Myr for VVV CL008) and $Z = 0.020$ are plotted with green lines. Red lines represent the sequence of the zero-reddening stars of luminosity class V (Schmidt-Kaler 1982) for illustration.

4.5.4. La Serena 003 and La Serena 009

The star cluster candidates La Serena 003 and 009 (LS003 and LS009) are groups of stars in formation that are deeply embedded in dust and gas. No stars with variations in the magnitude above our sensitivity limit of $\Delta K_s > 0.2$ mag are found in the vicinities of clusters. One YSO candidate, IRAS 11426–6301, in the field of LS009 is reported by Kwok et al. (1997). Later on, Mottram et al. (2011) and Lumsden et al. (2013) resolved the source to a YSO (G295.5570-01.3787A) and H II region (G295.5570-01.3787B, separated by $5''$) on the basis of far-infrared *MSX* measurements. They determined the radial velocity of 37.2 km s^{-1} and kinematic distance to both sources of 10.4 kpc. The bolometric luminosity of the YSO is calculated as 5980 $L_\odot$, while the H II region has 63,570 $L_\odot$, and the log mass of the whole clump is estimated to be 3.106 $M_\odot$ according to the same authors. Navarete et al. (2015) reported an extended H II emission associated with the region. Our K_s-band light curve (Figure 17) shows low-amplitude variability of 0.27 mag during the 2010–2015 time interval, with $\overline{K_s} = 11.365$ mag. Note, however, that the 2MASS magnitude (taken from the 2MASS Extended Sources Catalog; Skrutskie et al. 2006)) is $K = 8.94$ mag, and thus the star most probably shows long-term variability.

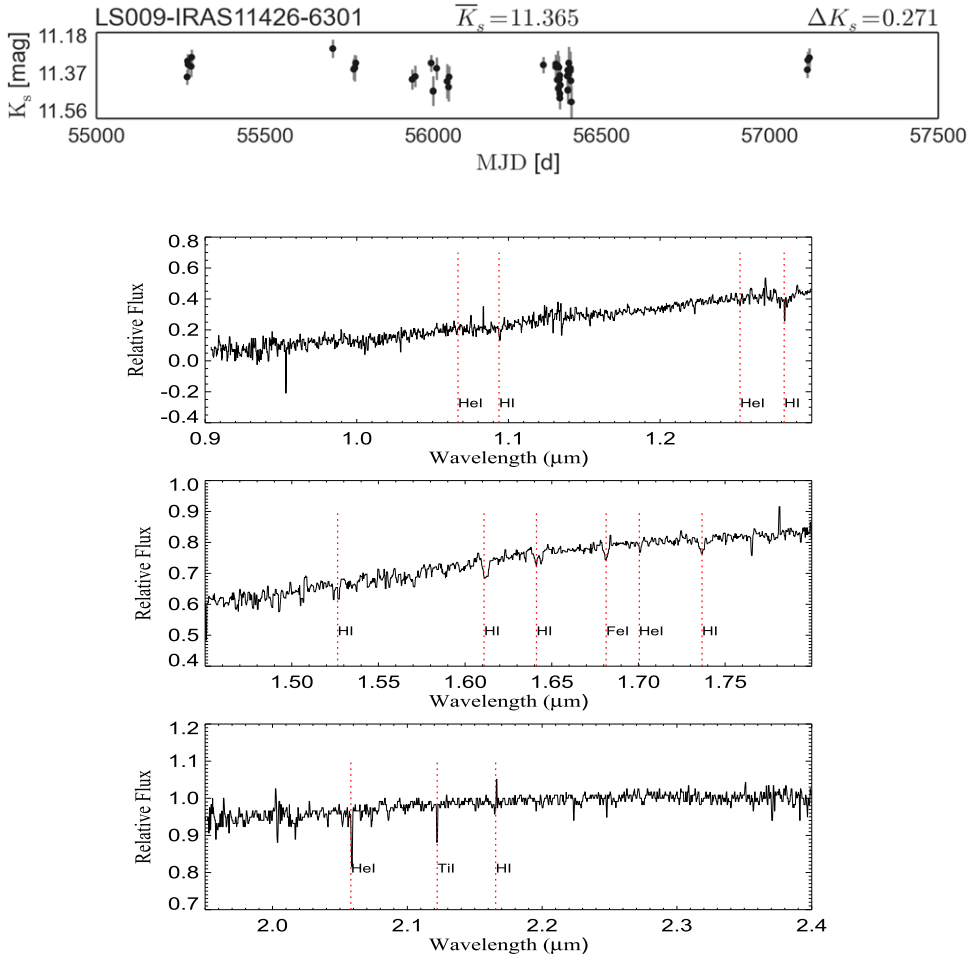

Figure 17. VVV time series and ARCoIRIS spectrum of the IRAS 11426–6301 YSO candidate.

IRAS 11426–6301 was observed in 2017 May with the ARCoIRIS spectrograph with 480 s integration time at 1.28 average airmass. We reduced the spectrum using the `Spextool` IDL package (version 4.1; Cushing et al. 2004), which is a data reduction algorithm specifically designed for the data format and characteristics of ARCoIRIS by Dr. Katelyn Allers. Telluric correction and flux calibration of the post-extraction spectra are achieved through the `xtellcorr` IDL package (Vacca et al. 2003). Figure 17 shows the spectrum normalized at 2.293 μm. As can be seen from the figure, the continuum level is flat to slightly rising. The overall spectral energy distribution is peaking at around 2.5 μm. The H I and He I lines are clearly visible in absorption, and some Ti lines and an Na I doublet (2.21 μm) in absorption also can be identified. The Ca I triplet (2.26 μm) and ^{12}CO bands (2.29 μm) are missing. While H I and He I generally represent early-type stars, the Ti and Na features are typical of low-mass YSOs. The absence of the CO band means the absence of a circumstellar disk. A possible explanation for this contradiction could be related to the photodissociation of the CO molecule. This phenomenon (where one even detects CI but no CO, or a much lower abundance than expected) has been observed around young A-type stars in deep searches for molecular gas in debris disks (Higuchi et al. 2017). Thus, the spectrum shows mixed features arising from both high- and low-mass young stars. Taking into account the kinematic distance of 10 kpc, this object could be an unresolved compact cluster or group of young stars.

4.5.5. VVV CL007

One variable (d002-35) is detected with our algorithm very close to the center of the VVV CL007 cluster. Taking into account its high amplitude, $\Delta K_s = 1.76$ mag; its position on the color–magnitude diagram (in the "P" region); and its classification as an eruptive variable (Contreras Peña et al. 2017a), it is most probably a YSO.

4.5.6. VVV CL008, VVV CL009, and La Serena 015

No variable stars are detected around VVV CL008, CL009, or La Serena 015. In Chené et al. (2013), we classified the Obj 4 (R.A. = 11:56:03.03 and decl. = −63:19:00.72) of VVV CL009 as a Be star.

In summary, only a few variable YSOs around the clusters are detected in the near-infrared.

4.6. Stars with Hα Photometric Emission

The clusters in tile d001 are very young, still in formation, and are surrounded by dust and gas; thus, we can use the photometric catalogs from the VPHAS+ survey (Drew et al. 2014) in order to search for additional YSOs and Hα emission candidates. The VPHAS+ catalog contains magnitudes in five filters:

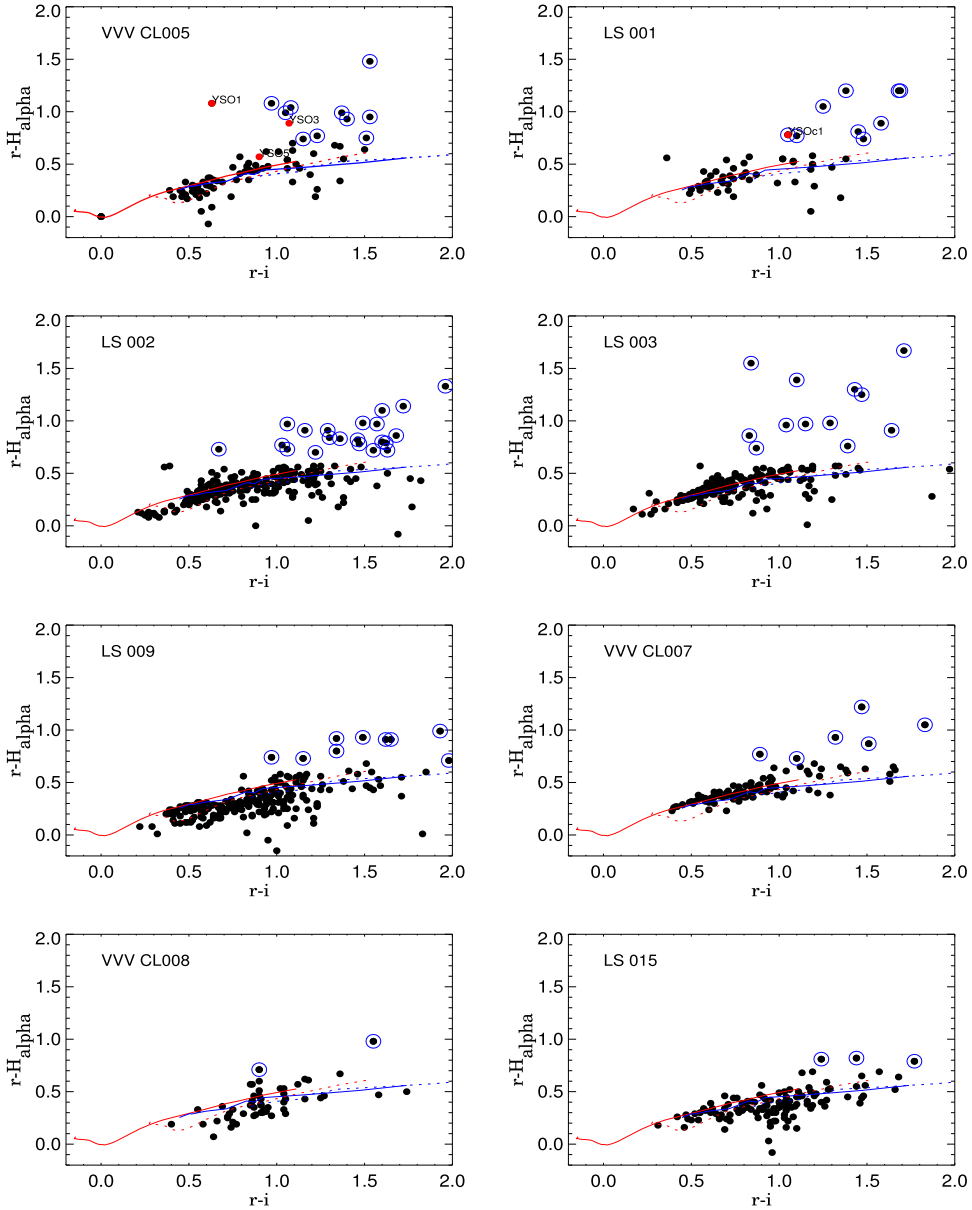

Figure 18. ($r - \mathrm{H}_\alpha$, $r - i$,) color–color diagrams. Blue circles show the selected emission stars. The red and blue solid lines are the synthetic unreddened main sequence and giant sequences, while the dashed ones correspond to those corrected for the reddening sequences (Drew et al. 2014). The red circles in the top plots are the infrared variable YSOs found in this work.

Table 4
VPHAS+ r, ($r - i$), and ($r - \mathrm{H}_\alpha$) Magnitudes and Colors of the YSO Candidates

ID	R.A.$_{2000}$ (deg)	Decl.$_{2000}$ (deg)	r (mag)	$r - i$ (mag)	$r - \mathrm{H}_\alpha$ (mag)
CL005 C1	174.71433	−63.48757	17.98 ± 0.02	1.08 ± 0.011	1.04 ± 0.01
CL005 C2	174.71582	−63.46827	19.39 ± 0.04	0.97 ± 0.025	1.08 ± 0.03
CL005 C3	174.71664	−63.48768	20.29 ± 0.10	1.53 ± 0.054	0.95 ± 0.09
CL005 C4	174.71783	−63.48682	18.37 ± 0.02	1.05 ± 0.011	0.99 ± 0.02
CL005 C5	174.71942	−63.4668	19.67 ± 0.05	1.37 ± 0.032	0.99 ± 0.04

(This table is available in its entirety in machine-readable form.)

u ($\lambda_{\rm eff} = 354$ nm), g ($\lambda_{\rm eff} = 475$ nm), r ($\lambda_{\rm eff} = 622$ nm), i ($\lambda_{\rm eff} = 763$ nm), and H_α ($\lambda_{\rm eff} = 659$ nm). Following Kalari et al. (2015), we construct the ($r - i$, $r - \mathrm{H}\alpha$) color–color diagrams using the DR2 of the VPHAS+ catalog. The main-sequence stars do not show any Hα emission with respect to the r-band photospheric continuum. We use the VST/OmegaCAM synthetic colors for main-sequence and giant stars in the ($r - \mathrm{H}\alpha$, $r - i$) plane (Drew et al. 2014) and select the stars with more than 5σ deviation from these synthetic sequences, corrected for the corresponding reddening (dashed lines in

Table 5
Information of content of V^4 Catalog

Column	Units	Description
ID	…	Identification given by the catalog
R.A.$_{2000}$	Degrees	Right Ascension of VVV source
Decl.$_{2000}$	Degrees	Declination of VVV source
GLon	Degrees	Galactic longitude of VVV source
GLat	Degrees	Galactic latitude of VVV source
$\overline{K_s}$	mag	Photometric mean value of K_s-band
$\overline{K_s}^{\rm err}$	mag	Estimation of $\overline{K_s}$ error using bootstrap technique
Epochs	…	Observation number of the source in their K_s time serie
ΔKs	mag	Photometric total amplitude of K_s-band
J	mag	Photometric J band
$J_{\rm err}$	mag	Photometric J-band error
H	mag	Photometric H band
$H_{\rm err}$	mag	Photometric H-band error
u	mag	Photometric u band
$u_{\rm err}$	mag	Photometric u-band error
g	mag	Photometric g band
$g_{\rm err}$	mag	Photometric g-band error
r	mag	Photometric r band
$r_{\rm err}$	mag	Photometric r-band error
$r2$	mag	Photometric $r2$ band
$r2_{\rm err}$	mag	Photometric $r2$-band error
i	mag	Photometric i band
$i_{\rm err}$	mag	Photometric i-band error
H_α	mag	Photometric H_α band
H_α err	mag	Photometric H_α-band error
η	…	Value of the variability index η
Class	…	Classification via VVV templates light curve or shape of the time serie
Tile ID	…	VVV tile where the source is located
Period	days	Identified period for a VVV source
A_{K_s}	mag	Extinction measured using Nishiyama et al. (2009)
Distance	Kpc	Distance measured from PL relations of RRab sources
CCD	…	Position in ($H - K_s$, $J - H$) color–color diagram
Reference	…	Reference to catalog of a documented source

(This table is available in its entirety in machine-readable form.)

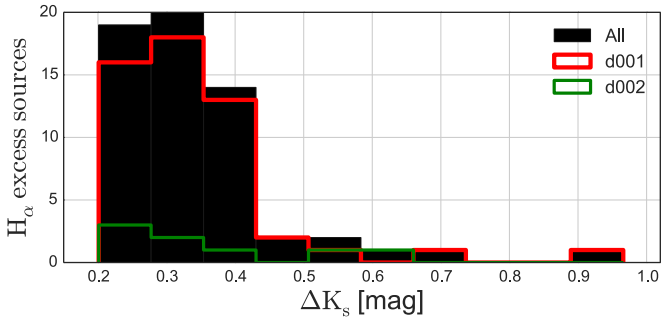

Figure 19. Distribution of amplitudes in the K_s band of the emission-line candidates.

Figure 18). The 73 selected sources are shown in Figure 18 and summarized in Table 4.

The cross-identification of these sources with the K_s variability catalogs shows that the peak of the amplitudes is around 0.3 mag (Figure 19). As we stated in Section 2.3, we set our detection limit on the spread of the magnitude measurements with time to be greater than 0.2 mag. For sources below this threshold, we treat these objects as constant at our sensitivity level. Thus, only a few of the infrared variable YSOs (d001-25 and d001-32) are emission-line object candidates.

5. Summary

We have developed an automated process for identification, classification, and analysis of variable sources using the VVV K_s-band time series, which are extracted directly from the $1.5 \times 1.^{\circ}2$ image tiles. This process was created to automatically analyze the VVV tiles, given the huge amount of available data. The sources that present variability in the NIR are cataloged in order to understand the physical process behind its variability, spatial distribution, evolutionary state, and relation with its environment. The gathered information from these sources will be collected in the VVV Variables (V^4) catalog, which will be publicly available in the VISTA Science Archive (VSA; http://vsa.roe.ac.uk/index.html) database and constantly updated by adding newly processed VVV tiles.

This process is based on `Dophot` PSF photometry to create a multi-epoch K_s-band catalog of detected sources in the FoV. We also obtained the J- and H-band photometry using the data taken in 2010. All of the PSF photometry was calibrated using the aperture catalogs made by CASU. To test this method, we select the d001 and d002 tile regions, covering an area of $\sim$3.6 deg^2. A total of 1,308,626 point sources in the range 10.8 mag $\leqslant \overline{K_s} \leqslant$ 17.2 mag were detected. The time series of sources with more than 25 epochs and amplitude $\Delta K_s > 0.2$ mag were selected and analyzed through different methods in order to

detect real variables. To avoid outliers, suspicious photometric measurements were removed from the light curves using the modified Thompson τ technique.

Our automated tool identified 200 sources with prominent NIR variability. Using two main variability indices (ΔK_s, η), we identified 70 variable sources without periodic or semi-periodic behavior that are bona fide irregular variables. On the other hand, we used periodograms (GLS, IP) to identify periodic sources, finding 130 of them. All identified sources have an average K_s magnitude distributed in the range 11.2 mag $< \overline{K_s} <$ 16.4 mag and a total amplitude contained in the range 0.2 mag $< \Delta K_s <$ 3.3 mag. About 90% of these sources are previously unknown as variable stars.

For each source with available J and H photometric measurements, its position in the ($H - K_s$, $J - H$) color–color diagram was reported, dividing this diagram into three parts (the "F," "T," and "P" regions) following the procedures of Ojha et al. (2004) in order to extract information about the evolutionary state of those sources.

In the case of the irregular variable sources, they were classified using the framework of Contreras Peña et al. (2017a), who used the time-series morphology to classify the sources into five categories (dippers, eruptives, LPV-YSOs, STVs, and faders). The LPV-YSO sources, together with the periodic variable sample, will be analyzed in a subsequent paper. Sources that could not be classified unambiguously into the aforementioned categories are marked as "unclassified." In total, we classified 20 STVs, 12 eruptives, 5 dippers, 7 faders, and 8 LAE variables, and 18 sources remained unclassified. These variable stars have amplitudes $\Delta K_s > 0.63$ mag, $\eta < 0.82$, and $J_{\rm stet} > 0.95$, without counting the source d001-75, which has $\eta = 3.193$ and $J_{\rm stet} = -1.55$.

We examined some parameter distributions in the specific parameter space (Graczyk & Eyer 2010) in order to separate the variability types. For example, a χ^2 parameter (in logarithmic scale) can separate the periodic sources from the main distribution when $\log(\chi^2) > 0.8$. In the future, we will use this to optimize the process of generation and analysis of time series and light curves in the VVV in the supervised algorithms of machine learning.

We classified 25 RRab, 42 RRc, 13 classic Cepheids, 33 binaries, 7 LPVs, and 7 LPV-YSOs, and three periodic variables remain unclassified. The periodic sources have a distribution of periods in the range 0.2days $< P <$ 1430 days and an average K_s magnitude distributed in the range 11.3 mag $< \overline{K_s} <$ 16.32 mag.

We also analyzed nine open-cluster candidates using surface-density maps, color-based decontamination, and proper-motion decontamination algorithms to determine the radii of the clusters. We have estimated the mean reddening $E(J - Ks)$ by comparison with the PARSEC isochrones for solar metallicity. We were only able to determine a distance modulus for the VVV CL005 cluster using the spectroscopic parallax. This cluster also has the largest number of irregular variables as probable cluster members (d001-25, 27, 28, 29, and 30). The cluster LS002 has two irregular variables, while clusters LS001 and VVV CL007 have one irregular variable, each other, as probable cluster members. All irregular sources projected close to open clusters in the region are YSO candidates.

Given the low number of irregular variable sources we found around open clusters, we used the VPHAS+ survey to identify excess in the H_α band. We have selected 73 stars with more than 5σ difference from the VST/OmegaCAM synthetic colors for main-sequence and giant stars in the ($r - i$, $r - H_\alpha$) plane (Drew et al. 2014). We noted that 64% of this sample has low-amplitude $\Delta K_s < 0.4$ mag, so this could be a good complementary method to find YSOs.

We gratefully acknowledge data from the ESO Public Survey program ID 179.B-2002 taken with the VISTA telescope and products from the Cambridge Astronomical Survey Unit (CASU), as well as data products from observations made with ESO telescopes at La Silla, Paranal Observatory, under program ID 177.D-3023 as part of the VST Photometric H_α Survey of the Southern Galactic Plane and Bulge (VPHAS+). Support for NM, JB, RK, CMN, MK, MC, and DM is provided by the Ministry for the Economy, Development and Tourism, Programa Iniciativa Científica Milenio grant IC120009, awarded to the Millennium Institute of Astrophysics (MAS). We also thank the referee for comments that greatly improved the present manuscript. AB is thankful for support from the Millennium Science Initiative (Chilean Ministry of Economy). DM is supported by the BASAL Center for Astrophysics and Associated Technologies (CATA) through grant PFB-06 and by FONDECYT Regular grant No. 1170121.

Software: `Matplotlib` (Hunter 2007), `AstroPy` (The Astropy Collaboration et al. 2018), `Sklearn` (Pedregosa et al. 2011), `AstroML` (Vanderplas & Ivezić 2007), `Dophot` (Schechter et al. 1993; Alonso-García et al. 2012), `VOSA` (Bayo et al. 2008), `STILTS` (Taylor 2006), `Spextool` Cushing et al. (2004), `xtellcorr` (Vacca et al. 2003).

ORCID iDs

N. Medina https://orcid.org/0000-0002-4684-1927
J. Borissova https://orcid.org/0000-0002-5936-7718
A. Bayo https://orcid.org/0000-0001-7868-7031
R. Kurtev https://orcid.org/0000-0002-9740-9974
C. Navarro Molina https://orcid.org/0000-0002-3786-9135
M. Kuhn https://orcid.org/0000-0002-0631-7514
N. Kumar https://orcid.org/0000-0002-7066-4828
P. W. Lucas https://orcid.org/0000-0002-8872-4462
L. C. Smith https://orcid.org/0000-0002-3259-2771

References

Alonso-García, J., Dékány, I., Catelan, M., et al. 2015, AJ, 149, 99
Alonso-García, J., Mateo, M., Sen, B., et al. 2012, AJ, 143, 70
Angeloni, R., Contreras Ramos, R., Catelan, M., et al. 2014, A&A, 567, A100
Arnaboldi, M., Neeser, M. J., Parker, L. C., et al. 2007, Msngr, 127, 28
Arnaboldi, M., Rejkuba, M., Retzlaff, J., et al. 2012, Msngr, 149, 7
Bailey, S. I. 1902, AnHar, 38, 252
Barbá, R. H., Roman-Lopes, A., Nilo Castellón, J. L., et al. 2015, A&A, 581, A120
Bayo, A., Rodrigo, C., Barrado Y Navascués, D., et al. 2008, A&A, 492, 277
Bessell, M. S., & Brett, J. M. 1988, PASP, 100, 1134
Bonatto, C., & Bica, E. 2010, A&A, 516, A81
Borissova, J., Bonatto, C., Kurtev, R., et al. 2011, A&A, 532, A131
Borissova, J., Chené, A.-N., Ramírez Alegría, S., et al. 2014, A&A, 569, A24
Borissova, J., Ramírez Alegría, S., Alonso, J., et al. 2016, AJ, 152, 74
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Carpenter, J. M., Hillenbrand, L. A., & Skrutskie, M. F. 2001, AJ, 121, 3160
Chené, A.-N., Borissova, J., Bonatto, C., et al. 2013, A&A, 549, A98
Chené, A.-N., Borissova, J., Clarke, J. R. A., et al. 2012, A&A, 545, A54
Cody, A. M., Stauffer, J., Baglin, A., et al. 2014, AJ, 147, 82
Contreras Peña, C., Lucas, P. W., Kurtev, R., et al. 2017b, MNRAS, 465, 3039
Contreras Peña, C., Lucas, P. W., Minniti, D., et al. 2017a, MNRAS, 465, 3011

Cross, N. J. G., Collins, R. S., Mann, R. G., et al. 2012, A&A, 548, A119
Cushing, M. C., Vacca, W. D., & Rayner, J. T. 2004, PASP, 116, 362
Dalton, G. B., Caldwell, M., Ward, A. K., et al. 2006, Proc. SPIE, 6269, 62690X
Dékány, I., Minniti, D., Majaess, D., et al. 2015, ApJL, 812, L29
Dong, H., Schödel, R., Williams, B. F., et al. 2017, MNRAS, 471, 3617
Drake, A. J., Djorgovski, S. G., Mahabal, A., et al. 2009, ApJ, 696, 870
Drew, J. E., Gonzalez-Solares, E., Greimel, R., et al. 2014, MNRAS, 440, 2036
Elorrieta, F., Eyheramendy, S., Jordán, A., et al. 2016, A&A, 595, A82
Elson, R. A. W., Fall, S. M., & Freeman, K. C. 1987, ApJ, 323, 54
Findeisen, K., Hillenbrand, L., Ofek, E., et al. 2013, ApJ, 768, 93
Gavrilchenko, T., Klein, C. R., Bloom, J. S., & Richards, J. W. 2014, MNRAS, 441, 715
Graczyk, D., & Eyer, L. 2010, AcA, 60, 109
Gran, F., Minniti, D., Saito, R. K., et al. 2016, A&A, 591, A145
Harju, J., Lehtinen, K., Booth, R. S., & Zinchenko, I. 1998, A&AS, 132, 211
Herbig, G. H. 1966, VA, 8, 109
Herbig, G. H. 1989, in ESO Workshop Proc. 33, ed. B. Reipurth (München: ESO), 233
Hervé, A., Martins, F., Chené, A.-N., Bouret, J.-C., & Borissova, J. 2016, NewA, 45, 84
Higuchi, A. E., Sato, A., Tsukagoshi, T., et al. 2017, ApJL, 839, L14
Hillenbrand, L. A., Strom, S. E., Vrba, F. J., & Keene, J. 1992, ApJ, 397, 613
Huijse, P., Estevez, P. A., Zegers, P., Principe, J. C., & Protopapas, P. 2011, ISPL, 18, 371
Hunter, J. D. 2007, CSE, 9, 90
Kaiser, N., Aussel, H., Burke, B. E., et al. 2002, Proc. SPIE, 4836, 154
Kalari, V. M., Vink, J. S., Drew, J. E., et al. 2015, MNRAS, 453, 1026
Krabbendam, V. L., & Sweeney, D. 2010, Proc. SPIE, 7733, 77330D
Kwok, S., Volk, K., & Bidelman, W. P. 1997, ApJS, 112, 557
Lomb, N. R. 1976, Ap&SS, 39, 447
Lumsden, S. L., Hoare, M. G., Urquhart, J. S., et al. 2013, ApJS, 208, 11
Marigo, P., Girardi, L., Bressan, A., et al. 2017, ApJ, 835, 77
McSwain, M. V., & Gies, D. R. 2005, ApJS, 161, 118
Meyer, M. R., Calvet, N., & Hillenbrand, L. A. 1997, AJ, 114, 288
Minniti, D., Contreras Ramos, R., Alonso-García, J., et al. 2015, ApJL, 810, L20
Minniti, D., Dékány, I., Majaess, D., et al. 2017, AJ, 153, 179
Minniti, D., Lucas, P. W., Emerson, J. P., et al. 2010, NewA, 15, 433
Mottram, J. C., Hoare, M. G., Urquhart, J. S., et al. 2011, A&A, 525 A149
Navarete, F., Damineli, A., Barbosa, C. L., & Blum, R. D. 2015, MNRAS, 450, 4364
Navarro Molina, C., Borissova, J., Catelan, M., et al. 2016, MNRAS, 462, 1180
Ojha, D. K., Tamura, M., Nakajima, Y., et al. 2004, ApJ, 616, 1042
Palma, T., Minniti, D., Dékány, I., et al. 2016, NewA, 49, 50
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, J. Mach. Learning Res., 12, 2825
Perryman, M. A. C. 2005, in ASP Conf. Ser. 338, Astrometry in the Age of the Next Generation of Large Telescopes, ed. P. K. Seidelmann & A. K. B. Monet (San Francisco, CA: ASP), 3
Pojmanski, G. 1998, AcA, 48, 35
Rebull, L. M., Cody, A. M., Covey, K. R., et al. 2014, AJ, 148, 92
Saito, R. K., Hempel, M., Minniti, D., et al. 2012b, A&A, 537, A107
Saito, R. K., Minniti, D., Dias, B., et al. 2012a, A&A, 544, A147
Sana, H., James, G., & Gosset, E. 2011, MNRAS, 416, 817
Scargle, J. D. 1982, ApJ, 263, 835
Schechter, P. L., Mateo, M., & Saha, A. 1993, PASP, 105, 1342
Schmidt-Kaler, T. 1982, BICDS, 23, 2
Shin, M.-S., Sekora, M., & Byun, Y.-I. 2009, MNRAS, 400, 1897
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
Smith, L. C., Lucas, P. W., Kurtev, R., et al. 2018, MNRAS, 474, 1826
Sokolovsky, K. V., Gavras, P., Karampelas, A., et al. 2017, MNRAS, 464, 274
Stetson, P. B. 1996, PASP, 108, 851
Taylor, M. B. 2006, in ASP Conf. Ser. 351, Astronomical Data Analysis Software and Systems XV, ed. C. Gabriel et al. (San Francisco, CA: ASP), 666
The Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, arXiv:1801.02634
Thompson, R. 1985, J. Royal Stat. Soc. Ser. B, 47, 53
Vacca, W. D., Cushing, M. C., & Rayner, J. T. 2003, PASP, 115, 389
Vanderplas, J., & Ivezić, Ž. 2007, ApJ, 812, 18
von Neumann, J. 1941, Ann. Math. Statist., 12, 367
Welch, D. L., & Stetson, P. B. 1993, AJ, 105, 1813
Wolk, S. J., Rice, T. S., & Aspin, C. 2013, ApJ, 773, 145
Zechmeister, M., & Kürster, M. 2009, A&A, 496, 577