

## **The Vista Variables in the Via Lactea (VVV) ESO Public Survey: Current Status and First Results**

Catelan, M; Minniti, D; Lucas, PW; Alonso-Garcia, J; Angeloni, R; Beamin, JC; Bonatto, C; Borissova, J; Contreras, C; Cross, N; Dekany, I; Emerson, JP; Eyheramendy, S; Geisler, D; Gonzalez-Solares, E; Helminiak, KG; Hempel, M; Irwin, MJ; Ivanov, VD; Jordan, A; Kerins, E; Kurtev, R; Mauro, F; Bidin, CM; Navarrete, C; Perez, P; Pichara, K; Read, M; Rejkuba, M; Saito, RK; Sale, SE; Toledo, I

arXiv record <http://arxiv.org/abs/1105.1119>

For additional information about this publication click this link.

<http://qmro.qmul.ac.uk/xmlui/handle/123456789/14370>

Information about this research object was correct at the time of download; we occasionally make corrections to records, please therefore check the published record when citing. For more information contact [scholarlycommunications@qmul.ac.uk](mailto:scholarlycommunications@qmul.ac.uk)

---

*Carnegie Observatories Astrophysics Series, Vol. 5:  
RR Lyrae Stars, Metal-Poor Stars, and the Galaxy  
ed. A. McWilliam (Pasadena: Carnegie Observatories)*

# The Vista Variables in the Vía Láctea (VVV) ESO Public Survey: Current Status and First Results

M. CATELAN<sup>1</sup>, D. MINNITI<sup>1</sup>, P. W. LUCAS<sup>2</sup>, J. ALONSO-GARCÍA<sup>1</sup>, R. ANGELONI<sup>1</sup>, J. C. BEAMÍN<sup>1</sup>, C. BONATTO<sup>3</sup>, J. BORISSOVA<sup>4</sup>, C. CONTRERAS<sup>1,2</sup>, N. CROSS<sup>5</sup>, I. DÉKÁNY<sup>1</sup>, J. P. EMERSON<sup>6</sup>, S. EYHERAMENDY<sup>1</sup>, D. GEISLER<sup>7</sup>, E. GONZÁLEZ-SOLARES<sup>8</sup>, K. G. HELMINIAK<sup>1</sup>, M. HEMPEL<sup>1</sup>, M. J. IRWIN<sup>8</sup>, V. D. IVANOV<sup>9</sup>, A. JORDÁN<sup>1</sup>, E. KERINS<sup>10</sup>, R. KURTEV<sup>4</sup>, F. MAURO<sup>7</sup>, C. MONI BIDIN<sup>7</sup>, C. NAVARRETE<sup>1</sup>, P. PÉREZ<sup>1,11</sup>, K. PICHARA<sup>1</sup>, M. READ<sup>5</sup>, M. REJKUBA<sup>12</sup>, R. K. SAITO<sup>1</sup>, S. E. SALE<sup>1,4</sup>, I. TOLEDO<sup>1</sup>

- (1) Pontificia Universidad Católica de Chile, Santiago, Chile
- (2) University of Hertfordshire, Hatfield, UK
- (3) Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
- (4) Universidad de Valparaíso, Valparaíso, Chile
- (5) The University of Edinburgh, Edinburgh, UK
- (6) Queen Mary, University of London, London, UK
- (7) Universidad de Concepción, Concepción, Chile
- (8) University of Cambridge, Cambridge, UK
- (9) European Southern Observatory, Santiago, Chile
- (10) The University of Manchester, Manchester, UK
- (11) Universidad de Chile, Santiago, Chile
- (12) European Southern Observatory, Garching, Germany

---

## Abstract

Vista Variables in the Vía Láctea (VVV) is an ESO Public Survey that is performing a variability survey of the Galactic bulge and part of the inner disk using ESO's Visible and Infrared Survey Telescope for Astronomy (VISTA). The survey covers  $520 \text{ deg}^2$  of sky area in the  $ZYJHK_S$  filters, for a total observing time of 1929 hours, including  $\sim 10^9$  point sources and an estimated  $\sim 10^6$  variable stars. Here we describe the current status of the VVV Survey, in addition to a variety of new results based on VVV data, including light curves for variable stars, newly discovered globular clusters, open clusters, and associations. A set of reddening-free indices based on the  $ZYJHK_S$  system is also introduced. Finally, we provide an overview of the VVV Templates Project, whose main goal is to derive well-defined light curve templates in the near-IR, for the automated classification of VVV light curves.

## 1. Introduction

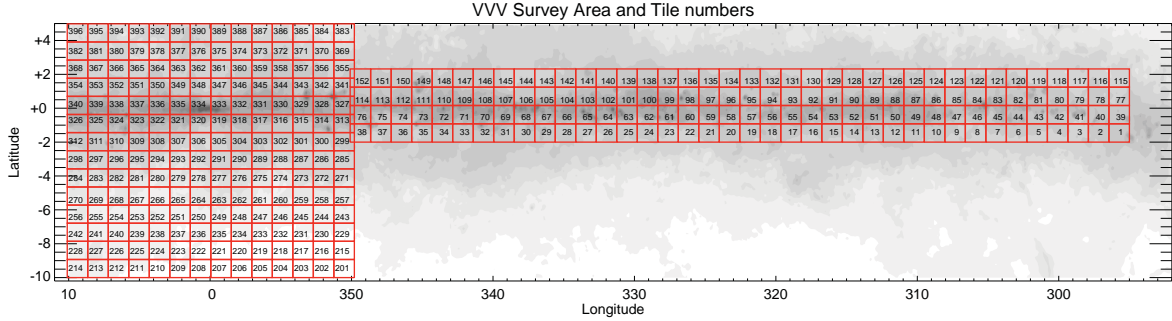
The Vista Variables in the Vía Láctea (VVV) ESO Public Survey<sup>1</sup> is an infrared variability survey of the Milky Way bulge and an adjacent section of the mid-plane where star formation activity is high (Minniti et al. 2010; Saito et al. 2010). The survey is carried out on ESO’s Visible and Infrared Survey Telescope for Astronomy (VISTA) 4m telescope, which is equipped with a wide-field near-infrared (IR) camera (VIRCAM; Dalton et al. 2006) that is a mosaic of 16  $2048 \times 2048$  detectors, each sensitive over the spectral region  $\lambda = 0.8 - 2.5 \mu\text{m}$ , and with an average pixel scale of  $0.34''\text{pixel}^{-1}$ . The total (effective) field of view (FoV) of the camera is  $1.1 \times 1.5 \text{deg}^2$ .

In addition to providing images covering a wide range in effective wavelength (*ZYJHK<sub>S</sub>* filters), the VVV Survey is able to probe much deeper into high extinction regions than its predecessors. In particular, near-IR color-magnitude diagrams (CMDs) of some VVV fields extend about 4 magnitudes deeper than their 2MASS counterparts (Saito et al. 2010), and also about 1 magnitude deeper than the UKIDSS Galactic Plane Survey (GPS) in the regions of overlap (Minniti et al. 2010). Most importantly, the VVV Survey will also provide a window into time-variable phenomena, by repeatedly covering the Galactic bulge and regions of the inner plane over a timespan of about 5 years, with a total of 1929 hours of observing time. Of order  $10^9$  point sources within a total sky area of  $520 \text{deg}^2$  will thus be monitored. According to the December 2010 version of the Harris (1996) catalog, the surveyed region includes 36 globular clusters<sup>2</sup> and  $\sim 314$  known open clusters, and is shown in Figure 1. The final products will be a deep IR atlas in 5 passbands and a catalogue containing an estimated  $\sim 10^6$  variable stars. The synergy with other surveys that provide optical, mid-IR, and far-IR data is also worth noting, including UKIDSS GPS, IPHAS, VPHAS+, the Spitzer, GLIMPSE, and MIPS GAL surveys, and the all-sky AKARI and WISE surveys.

The scientific goals of the VVV Survey are described in great detail in Minniti et al. (2010). One of the main goals is to gain insight into the inner Milky Way’s origin, structure, and evolution. This will be achieved by, for instance, obtaining a first, precise 3D map of the Galactic bulge, whose detailed shape appears to be quite complex (e.g., McWilliam & Zoccali 2010; Nataf et al. 2010; De Propris et al. 2011). Of particular importance, in order to achieve this goal, will be the RR Lyrae stars. Not only are the RR Lyrae variables present in the direction of the bulge in great numbers (Soszyński et al. 2011), but they also follow precise period-luminosity relations in the near-IR bandpasses (e.g., Catelan et al. 2004). Moreover, being very old, they are “living eyewitnesses” of virtually the entire formation history of the Milky Way (see, e.g., Catelan 2009, for a recent review). Another important goal of the VVV Survey is the discovery of globular clusters, open clusters, and stellar associations which have heretofore remained “hidden” from view behind the high extinction regions towards

<sup>1</sup> <http://vvvsurvey.org/>

<sup>2</sup> An additional globular cluster, FSR1767 (Bonatto et al. 2007), is also in the surveyed region, but does not yet appear in the Harris (1996) catalog (see also §3.2).



**Figure 1.**— The VVV Survey area, with individual “tiles” (i.e., individual observation fields, obtained from a combination of “pawprints”; see, e.g., Minniti et al. 2010) numbered. This is a plot of Galactic latitude  $b$  version Galactic longitude  $\ell$ , overplotted on a differential extinction contour map.

**TABLE 1**  
Overview of VVV Observations:  
First Year

Filters	Planned <sup>a</sup>	Executed <sup>a</sup>	Percentage
Bulge			
$JHK_S$	196	176	90%
$ZY$	196	78	40%
var <sup>b</sup>	980	113	12%
Disk			
$JHK_S$	152	152	100%
$ZY$	152	128	84%
var <sup>b</sup>	760	547	72%

<sup>a</sup> In units of observing blocks (OBs).

<sup>b</sup> For the first year, the plan was to execute a total of 5 observations for each of the 196 (152) bulge (disk) fields, giving a total of 980 (760) OBs.

the inner Galaxy. In this paper, we provide an update on the current status of the project, as well as a first look into some of the science that has so far been enabled by the VVV data.

## 2. The VVV Survey: Current Status

### 2.1. Pipeline Processing at CASU

VVV observations are carried out in service mode. During the first year,  $JHK_S$  images typically had  $0.8''$  seeing, whereas  $ZY$  images typically had a seeing of  $1''$ . Once observations are carried out, the VVV data are processed by a pipeline at the Cambridge Astronomy Survey Unit (CASU; see, e.g., Irwin et al. 2004; Lewis et al. 2010). In addition to properly reduced and calibrated images, the CASU pipeline also delivers aperture photometry for the observed fields. At the same time, members of our team perform profile-fitting photometry for the denser fields (e.g., star clusters). Work is currently in progress also to enable difference-imaging analyses (e.g., Alard 2000; Bramich 2008) in the multi-epoch,  $K_S$ -band data, for the detection of stellar variability in selected fields.

CASU has recently implemented v1.1 of its pipeline, and completed processing all the data obtained so far with the VISTA telescope. Some of the improvements that went into the new version of the pipeline include the following:<sup>3</sup>

- Improved calculation of aperture corrections in very crowded fields, which goes into the computation of the photometric zero points.
- Calculation of separate zero points for each of the  $6 \times 16 = 96$  sections of a VISTA tile. This corrects for both the significant non-uniformity of the VIRCAM point-spread function (PSF) across the FoV and changes in the PSF between the 6 pawprints that make up a tile. (See §3.1 for a description of “pawprints” and “tiles.”)
- Improved correction for the effect of extinction on the photometric transformations from 2MASS to VISTA, which is part of the zero point calculation.
- First data release of FITS images, confidence maps, and FITS single-band source catalogues will be released to ESO by the end of April/early May.

### 2.2. Current Status of Observations

In Table 1, we show the completion rate for the planned VVV first year observations, as of October 2010. Clearly, while almost complete coverage of the planned fields in  $JHK_S$  has been obtained, allowing the production of deep near-IR CMDs, the same is not true for the  $ZY$  filters, for which several fields could not be observed during the first year. Note also that the bulk of the planned variability campaign could not be carried out during the first year, especially in the case of the bulge fields.

<sup>3</sup> See also <http://casu.ast.cam.ac.uk/surveys-projects/vista/data-processing/version-log> for a comparison between different versions of the CASU pipeline.

Unfortunately, the coating of the VISTA telescope mirrors has been degrading more rapidly than anticipated, leading to a loss in sensitivity, particularly at short wavelengths. As a consequence, operation of the VISTA telescope has been halted recently, so that the mirror coatings can be replaced. This means that VVV observations for ESO Period 86 will not be completed as expected, likely leading to an extension of our original 5-year goal for completion of the project.

This notwithstanding, the available data already represent a treasure trove that can be explored in a variety of ways. In what follows, we report on some of the first scientific results that have been obtained on the basis of the VVV data acquired so far.

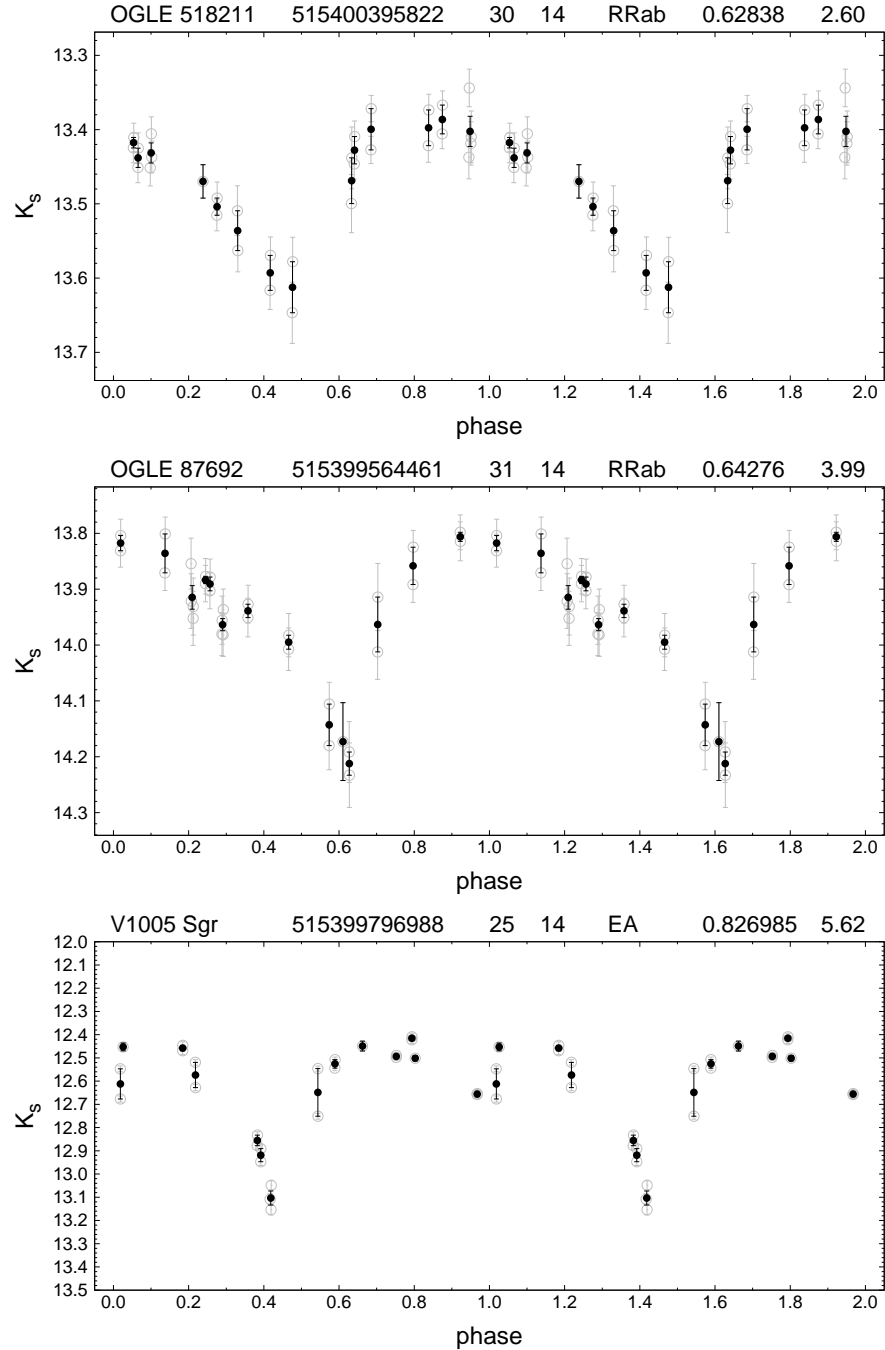
### 3. The VVV Survey: First Results

#### 3.1. Variable Stars in the Science Verification Fields

As can be seen from Table 1, the completion rate of planned observations for our variability campaign is still quite low. However, as part of ESO’s science verification of the VISTA telescope, additional data for our fields were obtained, in the period 19-30 Oct., 2009 (Arnaboldi et al. 2010). This consisted of multi-color imaging, in the  $ZYJHK_S$  filters, of a field located at Galactic coordinates  $b = -3.1^\circ$ ,  $\ell = 2.2^\circ$ , supplemented with an additional set of 13  $K_S$ -band images of the same field. We have searched the images available for this field for the presence of known variable stars, from the MACHO and OGLE catalogs (Alcock et al. 1998; Soszyński et al. 2011) and the General Catalog of Variable Stars (GCVS; Kholopov et al. 1998), and are able to identify several previously catalogued variables, including RR Lyrae stars and eclipsing variables. In Figure 2, we show three examples of the light curves obtained. This includes the ab-type RR Lyrae stars OGLE 518211 (*upper panel*) and OGLE 87692 (*middle panel*), additionally to the Algol-type (detached) eclipsing binary V1005 Sgr (*bottom panel*).

In Figure 2, gray symbols denote the individual measurements (and their formal errors) by standard VISTA Science Archive [VSA] aperture photometry, performed on single VISTA images called “pawprints.” The latter have a non-contiguous sky coverage due to the gaps between the 16 chips of VIRCAM, resembling the pawprint of an animal (for an example, see Fig. 15 below, right panel). Around each pointing of VISTA, 6 pawprints with proper offsets are acquired in succession, in order to provide a contiguous sampling of a sky area; stacks of these images are called “tiles.” Thus, a point source is observed at least two and up to six times in a row (depending on its position), within a short time interval (5 min 8 sec for the bulge in the first year). Magnitude values corresponding to such exposure sequences were averaged, and are shown as black dots in Figure 2, with error bars representing the standard deviation of the mean.

The time between the acquisition of two consecutive pawprints that correspond to the same tile is much shorter than the timescale of the light variation for the most



**Figure 2.**— Light curve for three variable stars in the VISTA science verification fields. *Top:* OGLE 518211, an ab-type RR Lyrae star. *Middle:* OGLE 87692, another ab-type RR Lyrae. *Bottom:* V1005 Sgr, a detached eclipsing binary. For all plots, from left to right, the following information is given: star ID in the OGLE catalog or in the GCVS; internal VVV source ID; number of data points in gray; number of data points in black; variability type; period in days; and a variability parameter  $d$ . See text for further details.



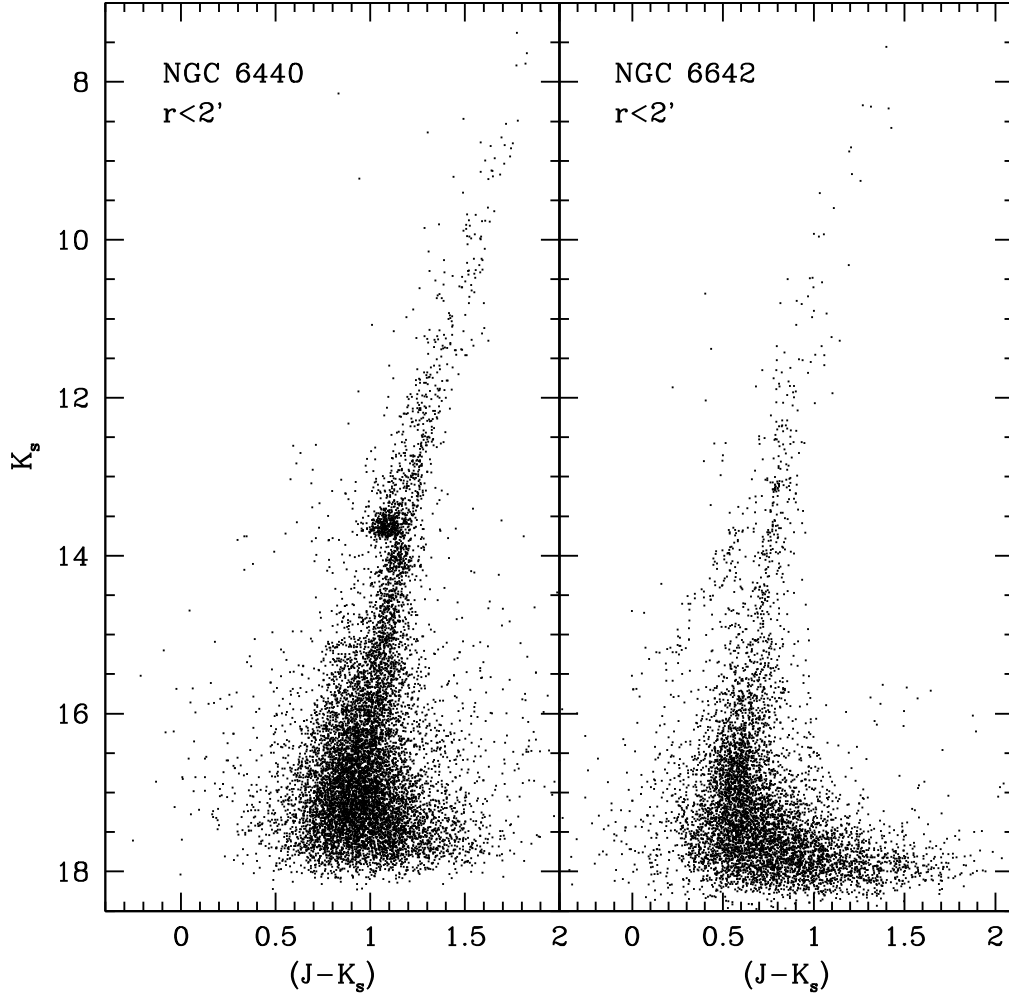
**Figure 3.**—  $JHK_S$  color-composite image showing the newly discovered globular cluster VVV CL001 (framed in a white rectangle) in the vicinity of the previously known globular cluster UKS 1. From Minniti et al. (2011).

typical variable stars that will be detected in the course of the VVV Survey. Therefore, these successive observations carry information about the internal scatter of each individual photometric time-series. The ratio  $d$  between the global rms scatter of a light curve and the internal scatter measured in the above way provides a good means for determining the probability of a star being variable within the photometric accuracy. Objects with  $d$  values deviating substantially from unit are good variable star candidates (see headers of Fig. 2).

### 3.2. Globular Star Clusters

The census of Galactic globular clusters is still far from being complete, because distant objects can easily hide behind/amongst the very crowded and highly reddened stellar fields in the direction of the Galactic bulge and disk. Indeed, the asymmetry of the spatial distribution of known globular clusters around the Galactic center clearly indicates that the census of globular clusters towards the inner Milky Way is likely incomplete. In this sense, Ivanov et al. (2005) estimated that there should be  $10 \pm 3$  such clusters in the more central regions of the Galaxy. The advent of the new generation of extensive, wide-field surveys such as SDSS (Abazajian et al. 2009), 2MASS (Skrutskie et al. 2006), and GLIMPSE (Benjamin et al. 2003) permitted the detection of several new Galactic globular clusters. The December 2010 compilation of the Harris (1996) catalog included seven new globulars not present in the February 2003 version, but six more objects have been proposed in the last years: SDSSJ1257+3419 (Sakamoto &





**Figure 4.**— CMDs of the inner 2' region of NGC 6440 and NGC 6626, two of the clusters sampled in the VVV Survey, representative of the metal-rich/moderately metal-poor globular cluster populations observed in the direction of the inner Galaxy.

Hasegawa 2006), FSR584 (Bica et al. 2007), FSR1767 (Bonatto et al. 2007), FSR190 (Froebrich et al. 2008), and Pfeiderer 2 (Ortolani et al. 2009).

As far as detection of new globular cluster candidates goes, the VVV Survey should naturally play an important role. This has already been confirmed in practice, with the discovery of one new globular cluster, VVV CL001, reported on in Minniti et al. (2011), and an additional three cluster candidates studied by Moni Bidin et al. (2011), at least one of which is very likely a previously unknown globular cluster. A  $JHK_S$  color-composite image of VVV CL001 is shown in Figure 3, where one can see its close proximity to the previously known globular UKS 1.

In addition to new globular clusters, the VVV Survey is also providing an unprecedented deep, wide-field view of the 37 previously known globular clusters located in

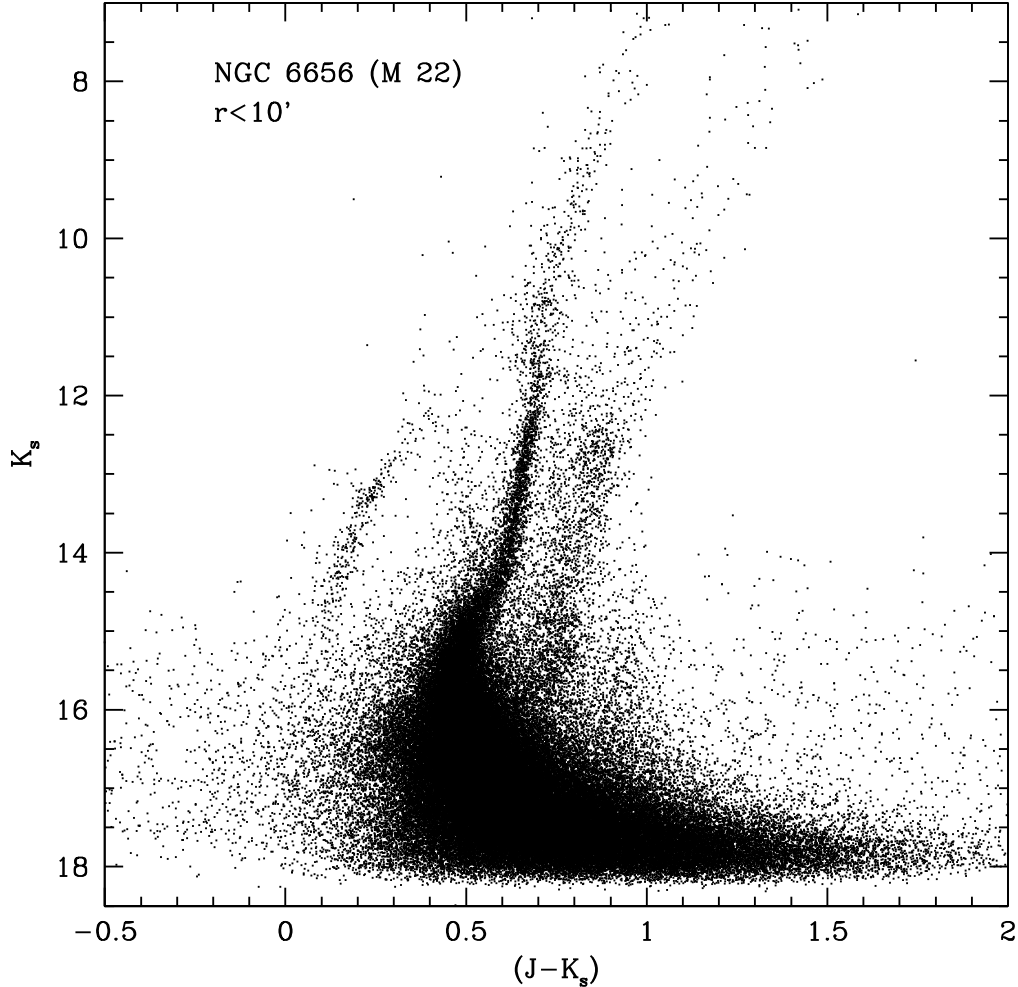
its observed fields. To illustrate the quality of the derived near-IR CMDs, in Figure 4 we show  $K_S$ ,  $(J - K_S)$  CMDs for the innermost  $2'$  regions of a metal-rich (NGC 6440, with  $[\text{Fe}/\text{H}] = -0.36$ ) and a moderately metal-poor (NGC 6642, with  $[\text{Fe}/\text{H}] = -1.26$ ) globular cluster. In Figure 5 we show a similar CMD, for the massive, metal-poor globular cluster M22 (NGC 6656), with  $[\text{Fe}/\text{H}] = -1.70$ , where a prominent blue horizontal branch extension is clearly observed, with a hint that it is far from being uniformly populated. Also, to illustrate the far superior handle on differential extinction that is provided by the VVV near-IR data, in Figure 6 we compare optical and near-IR CMDs of the globular cluster NGC 6553, with  $[\text{Fe}/\text{H}] = -0.18$ . (Metallicities are from the Dec. 2010 version of the Harris 1996 catalog.) In fact, one of the main advantages of the VVV survey is that we can sample and study every cluster present in the surveyed region up to its tidal radius – and indeed beyond. This will allow us to more accurately define the structural parameters of these clusters, and to study their stellar population (or populations) out to regions not usually observed, due to the smaller fields of view of most instruments.

### 3.3. Open Clusters and Associations

Estimates indicate that the Galaxy presently hosts 35,000 or more star clusters (Bonatto et al. 2006; Portegies Zwart et al. 2010). However, only about 2500 open clusters have been identified, which constitute a sample affected by several well-known selection effects. Moreover, less than a half of these clusters have actually been studied, and this subset suffers from further selection biases. In this sense, the VVV Survey will not only lead to the discovery of a vast number of previously unknown open clusters and associations, but also – and importantly – allow systematic follow-up studies of their properties to be carried out.

The first efforts by our team members to detect previously unknown open clusters and associations in the VVV Survey areas has led to the discovery of 96 candidates, which are reported on in Borissova et al. (2011). The positions of these candidates are shown in Figure 7. Most of the new candidate clusters are faint and compact, are highly reddened, and likely younger than about 5 Myr. Composite color images of several of these candidates are shown in Figure 8. As can be seen from Figure 9, where 2MASS and VVV images of the cluster candidate VVV CL062 are compared, detection of the new candidates has only become possible due to the much superior resolution of the VVV images, compared with 2MASS. Another such comparison is given in Figure 10, where  $K_S$ -band images of the ionizing star cluster RCW79 taken with 2MASS (left panel), SINFONI@VLT (middle panel), and VVV (right panel) are compared.

For one of our new cluster candidates, VVV CL036, a comparison between images taken in the  $ZYJHK_S$  filters is shown in Figure 11, clearly showing that the cluster only stands out in the near-IR. Even so, proper analysis of each cluster’s CMD branches and derivation of the physical parameters requires careful statistical decon-

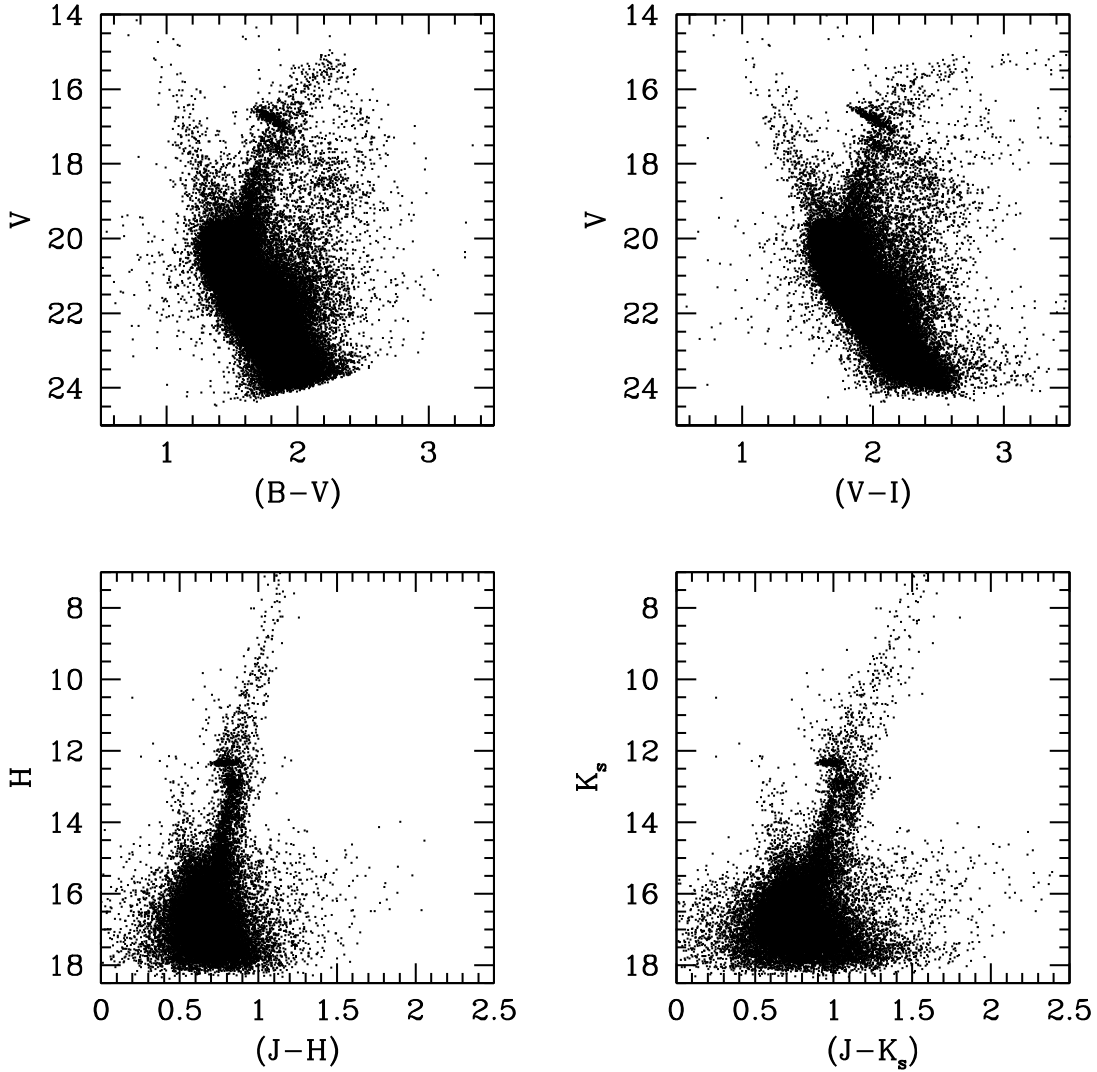


**Figure 5.**— CMD of the inner  $10'$  region of NGC 6656 (M22). Note the prominent, seemingly non-uniformly populated blue horizontal branch. A prominent red giant branch and red horizontal branch associated with the Galactic bulge (in the background) can also be seen.

tamination from the presence of field stars. In Figure 12 we illustrate the procedure adopted in Borissova et al. (2011) for the field star decontamination, which leads to excellent results, with the actual cluster population clearly standing out from the foreground/background field.

### 3.4. Reddening-Free Indices

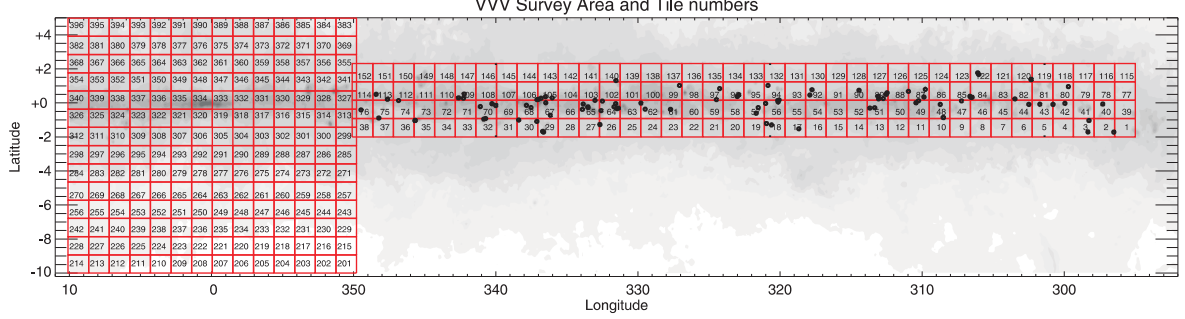
Even in the near-IR filters used in the VVV Survey, differential reddening effects can often complicate the analysis of cluster CMDs. As an example, in Figure 13 we show the near-IR CMD that we have derived for the moderately metal-poor ( $[\text{Fe}/\text{H}] = -1.41$ ) globular cluster Terzan 4, whose reddening is, according to the Harris (1996) catalog



**Figure 6.**— Optical and near-IR CMDs of the inner  $2.5'$  region of the globular cluster NGC 6553, obtained in the visual with ACS@HST and IMACS@Magellan (*top row*), and in the infrared with 2MASS and VVV data (*bottom row*). (2MASS data are used only to supplement the bright, saturated end of the VVV  $JHK_S$  data – i.e.,  $K_S \lesssim 11$ .) The effects caused by absolute and differential extinction are greatly reduced at infrared wavelengths, which allows a better comparison of the observational CMDs with theoretical models, thus leading to more accurate values of derived physical properties, such as photometric metallicity and age.

(Dec. 2010 version) as high as  $E(B-V) = 2.0$ . Clearly, in spite of using near-IR bandpasses, the cluster sequences are still severely distorted by differential reddening.

The fact that the VVV Survey utilizes a set of 5 different broadband filters allows us to explore the possibility of devising reddening-free indices, which can be very important for the studies of regions of high extinction (as is normally the case for the bulge and inner disk regions probed by this survey). To obtain such a set of



**Figure 7.**— As in Figure 1, but with the positions of the new star cluster candidates marked as dots. From Borissova et al. (2011).

reddening-free indices, we proceed as follows.

First, we assume a standard extinction law, with  $R = 3.09$  (Rieke & Lebofsky 1985). We then adopt the effective wavelengths  $\lambda_{\text{eff}}$  for the  $ZYJHK_S$  filters that are listed in Table 2. Based on these effective wavelengths, and using the Cardelli et al. (1989) extinction law,<sup>4</sup> we obtain the relative extinctions for these filters that are listed in Table 2.

These immediately lead to the following, reddening-free indices (among others):

$$m_1 \equiv K_S - 1.08 (Z - Y), \quad (1)$$

$$m_2 \equiv H - 1.13 (J - K_S), \quad (2)$$

$$m_3 \equiv J - 1.03 (Y - K_S), \quad (3)$$

$$m_4 \equiv K_S - 1.22 (J - H), \quad (4)$$

which can all be used as “pseudo-magnitudes,” as reddening-free replacements for  $J$ ,  $H$ , and  $K_S$ ; and

$$c_1 \equiv (Y - J) - 1.14 (J - H), \quad (5)$$

$$c_2 \equiv (Z - Y) - 0.99 (Y - J), \quad (6)$$

$$c_3 \equiv (J - H) - 1.47 (H - K_S), \quad (7)$$

$$c_4 \equiv (J - K_S) - 1.50 (Z - Y), \quad (8)$$

which can all be used as “pseudo-colors,” or reddening-free replacements for such colors as  $(Y - J)$ ,  $(Z - Y)$ ,  $(J - H)$ ,  $(H - K_S)$ , and  $(J - K_S)$ .

<sup>4</sup> See <http://www.mcho.mcmaster.ca/JAVA/Acurve.html>.

**TABLE 2**  
**Relative Extinctions for  $ZYJHK_S$**   
**Filters (VISTA Set)**

Filter	$\lambda_{\text{eff}}$ ( $\mu\text{m}$ )	$A_X/A_V$	$A_X/E(B-V)$
<i>Z</i>	0.878	0.499	1.542
<i>Y</i>	1.021	0.390	1.206
<i>J</i>	1.254	0.280	0.866
<i>H</i>	1.646	0.184	0.567
<i>K<sub>S</sub></i>	2.149	0.118	0.364

As a note of caution, the reader should also bear in mind that, when using such indices, the errors propagate accordingly. In addition, the intrinsic range in values (in mag units) covered by different CMD features can be significantly reduced, with respect to that seen in the original CMDs. Therefore, in order for clear-cut sequences to show up in such reddening-free diagrams, the internal precision must be significantly higher than what is needed for similar such sequences to be seen in the original CMDs. Note also that errors in the calibrations, particularly in the color terms, may also contribute to smearing up features that might otherwise show up well in such reddening-free diagrams. Last but not least, it is important to bear in mind that eqs. (1) through (7) were obtained under the assumption of a standard extinction law with  $R = 3.09$ . For applications to regions with different  $R$  values, these reddening-free indices should be redefined accordingly.

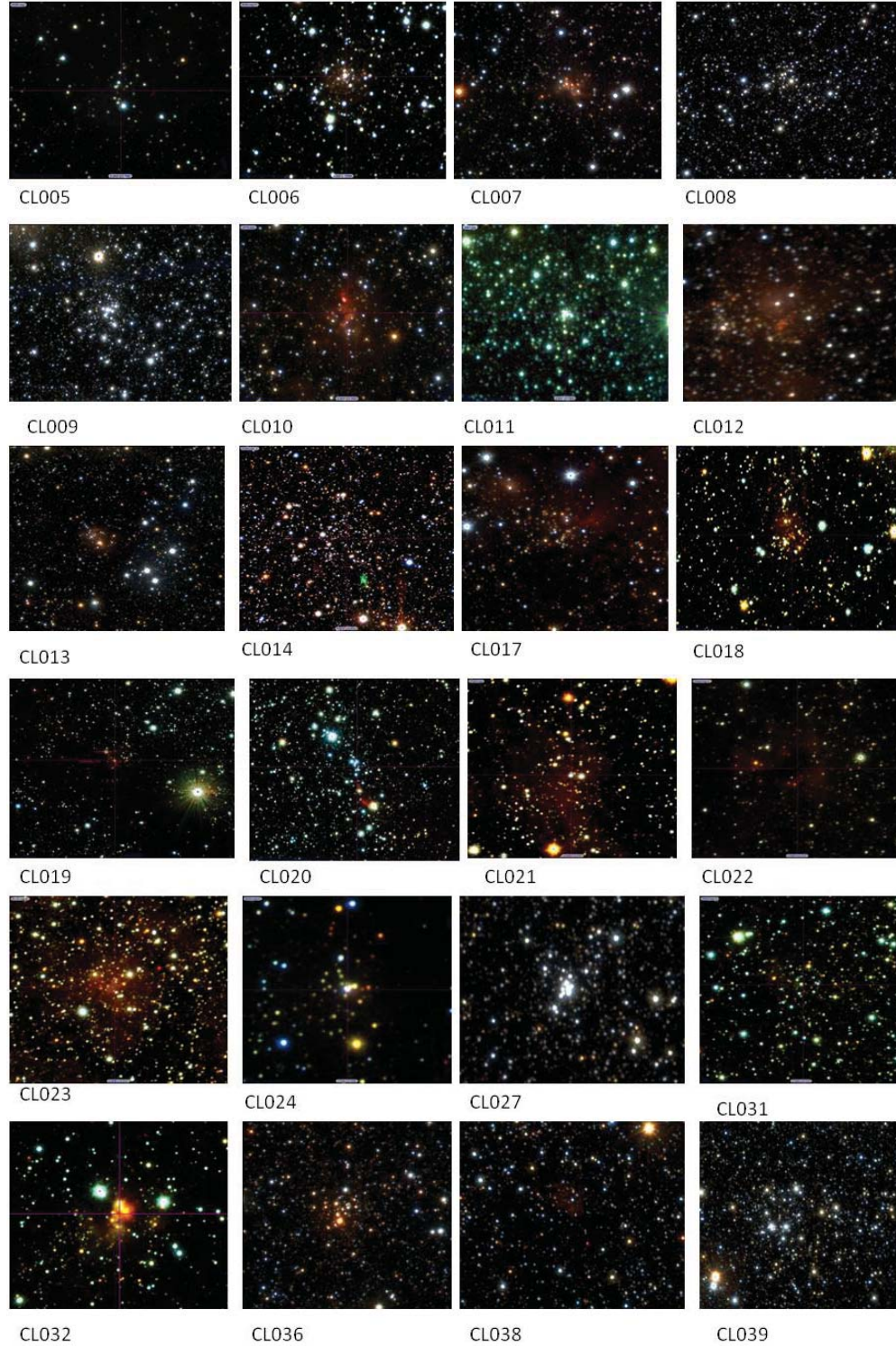
## 4. The VVV Survey Templates Project

As already pointed out (§1), the VVV ESO Public Survey will produce a catalogue and light curves for an estimated  $\sim 10^6$  variable stars, most of which will be previously unknown. Obviously, this very large number of light curves cannot be classified using traditional methods, particularly visual inspection. Instead, an automated classification scheme must be developed in order to properly deal with the classification of these variable stars. This is the ultimate goal of the VVV Templates Project.<sup>5</sup>

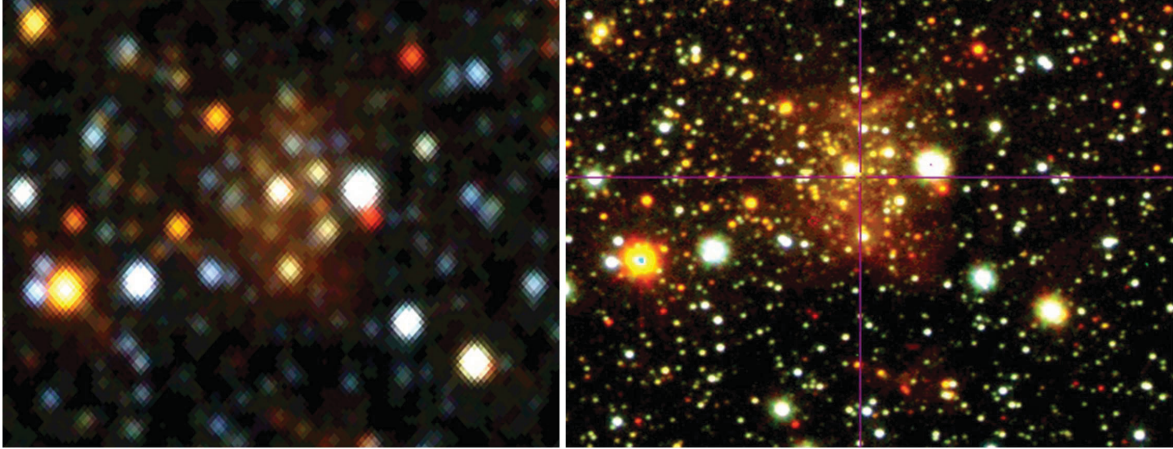
### 4.1. The Need for Near-IR Light Curve Templates

Unlike previous variability surveys, the VVV Survey is carried out in the near-IR. Despite several fundamental advantages, mostly due to the ability to probe deeper

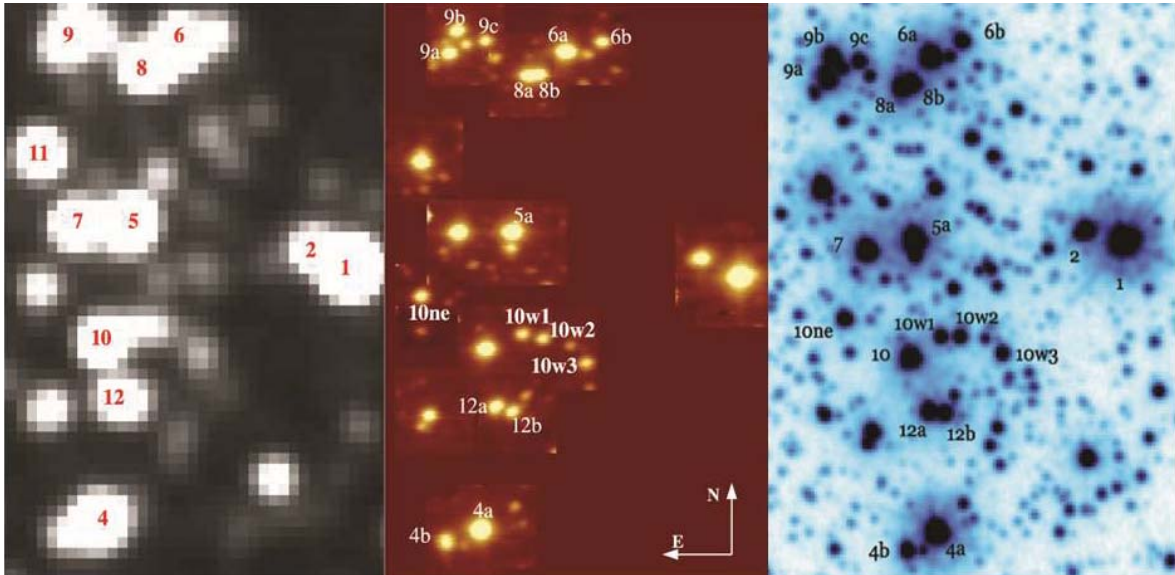
<sup>5</sup> <http://www.vvvtemplates.org>



**Figure 8.**—  $JHK_S$  true-color images of some of the new cluster candidates from the VVV Survey. The coordinates of the new candidates are given in Borissova et al. (2011).



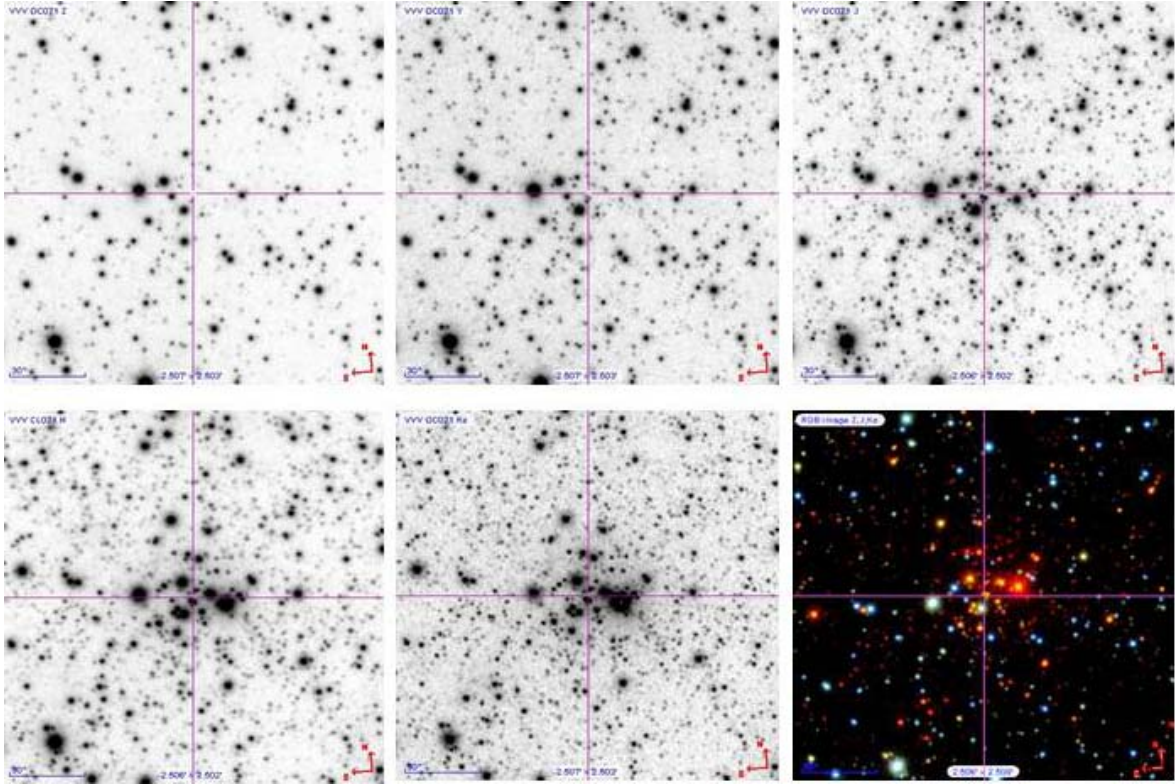
**Figure 9.**— 2MASS (*left*) and VVV (*right*) composite  $JHK_S$  color images of the cluster candidate VVV CL062 (Borissova et al. 2011). The field of view is approximately  $2.2' \times 1.8'$ . Note the far superior resolution of the VVV image.



**Figure 10.**— 2MASS (*left*), SINFONI@VLT (*middle*) and VVV (*right*)  $K_S$ -band images of the ionizing star cluster RCW79. The left and middle panels, as well as the identification of individual sources, are taken from Martins et al. (2010). Note that most of the sources resolved with SINFONI are also resolved in the VVV image.

into the heavily reddened regions of the Galactic bulge and plane, the use of this spectral region also presents us with important challenges. In particular, the high-quality templates that are needed for training the automated variable star classification algorithms are not available. This is very different from the situation of variability surveys carried out in the visible, for which vast quantities of high-quality template light curves are available (see, e.g. Debosscher et al. 2007; Blomme et al. 2011; Dubath





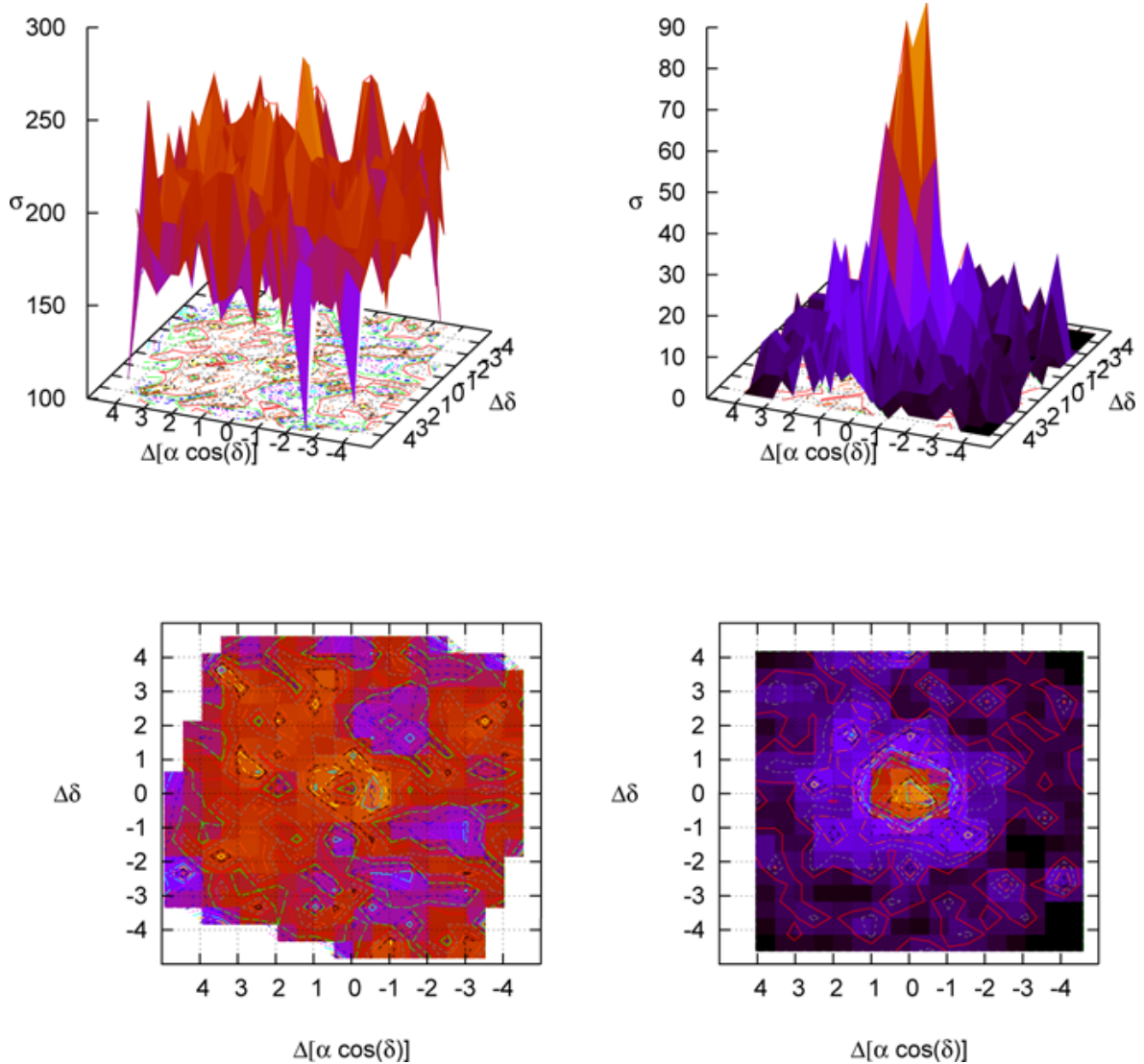
**Figure 11.**— VVV  $ZYJHK_S$  images and  $ZJK_S$  true-color image of VVV CL036. The field of view is  $2.5' \times 2.5'$ . North is up and East is to the left. From Borissova et al. (2011).

et al. 2011).

To properly illustrate the importance of this point, one should consider that not only is the currently available number of near-IR templates inadequate for our purposes (since they do not properly sample the large variety of light curve shapes encountered for many of the most important variability classes). Also, many such classes have not yet been observed in a sufficiently extensive way in the near-IR, so that good light curves are entirely lacking for these classes. Therefore, in order to properly classify the  $\sim 10^6$  light curves produced by the VVV Survey, we need to build our own template light curves.

#### 4.2. Current Status and Future Perspectives

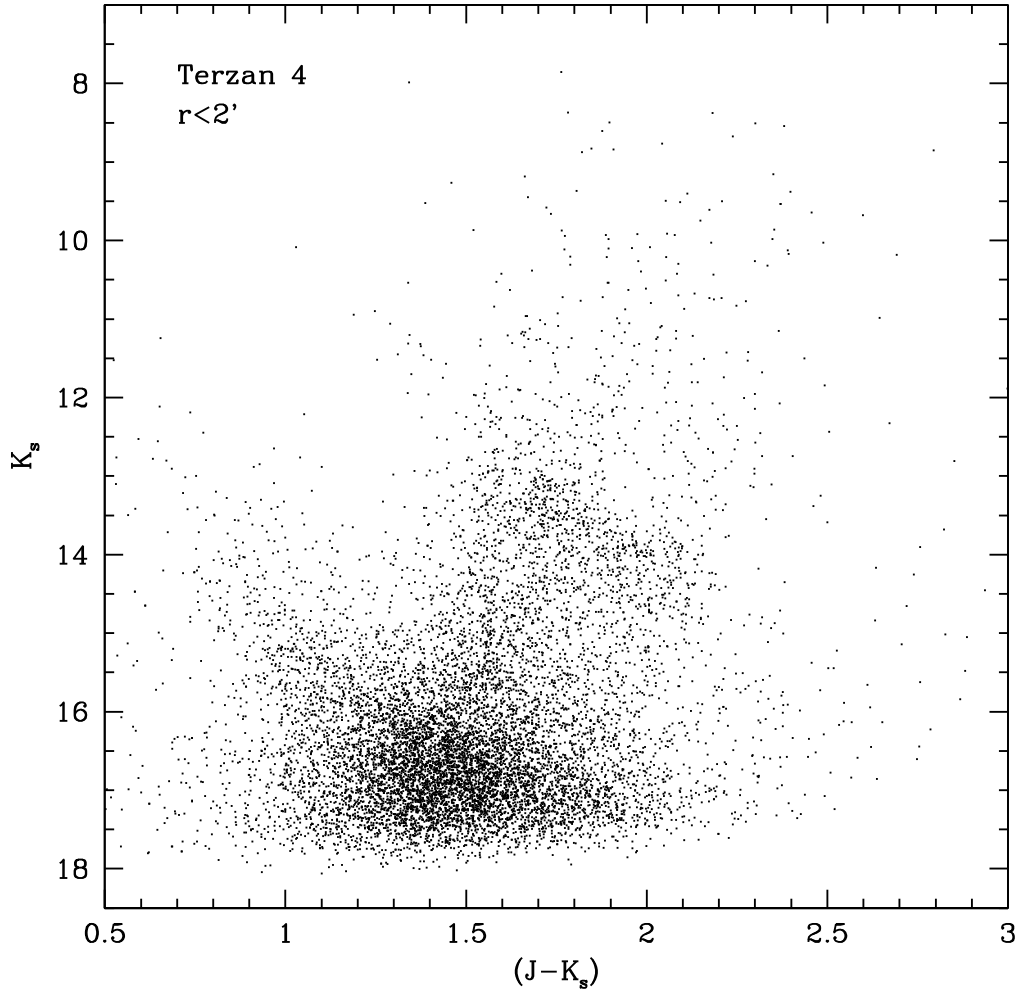
In the more general framework of the VVV Survey, the VVV Templates Project has turned out to be a large observational effort in its own right, aimed at creating the first database on stellar variability in the near-IR, i.e. producing a large database of well-defined, high-quality, near-IR light curves for variable stars belonging to different variability classes. By monitoring hundreds of (optically well-studied) variable stars in the  $JHK_S$  bands, its primary goal is thus to provide a statistically significant training



**Figure 12.**— Stellar surface density  $\sigma$  (stars arcmin $^{-2}$ ) map for VVV CL036. On the left is shown the radial density profile produced with the raw photometry, whereas on the right the same field is shown after statistical decontamination. From Borissova et al. (2011).

set for the automated classification of VVV light curves (see §4.3).

The astronomical community as a whole has been very supportive, with the end result that we have secured time for this project using several IR facilities across the globe, as listed in Table 3. The availability of telescope time is key to ensure that we produce a significant number of template-grade light curves in time for classification of the VVV light curves, i.e. within the next 3 years or so, before the bulk of the VVV variability campaign is carried out (Minniti et al. 2010; Saito et al. 2010). In order to ensure a homogeneous observational strategy and to optimize the use of the awarded time, each telescope/instrument combination is used to build template light curves for

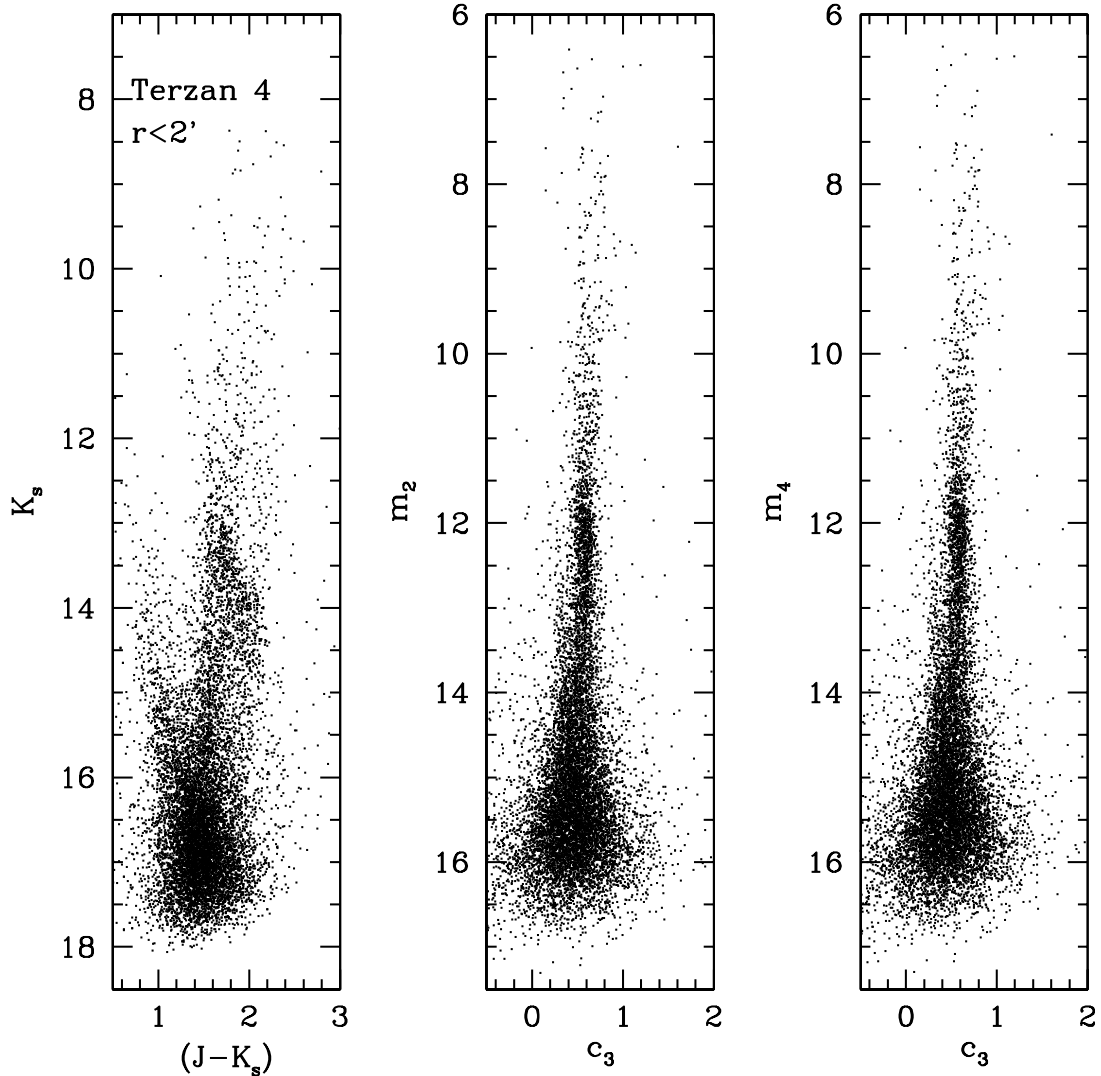


**Figure 13.**— CMD of the inner  $2'$  region of Terzan 4. Even though the effects of the interstellar extinction are much reduced in the IR when compared with the optical, the high amounts of dust and gas in the direction of the inner Galaxy can still affect the quality of the CMDs of some of the surveyed globular clusters. We plan to use techniques to decontaminate CMDs such as this one from differential extinction effects.

at most a very few specific variability classes, as can also be seen from Table 3.

Importantly, the scientific return of the VVV template light curves will not be restricted to the automated classification of the  $\sim 10^6$  light curves produced by the VVV Survey. Rather, these light curves will provide us with a unique opportunity to expand our knowledge of the stellar variability phenomenon per se, by tackling the comparatively ill-explored near-IR regime.

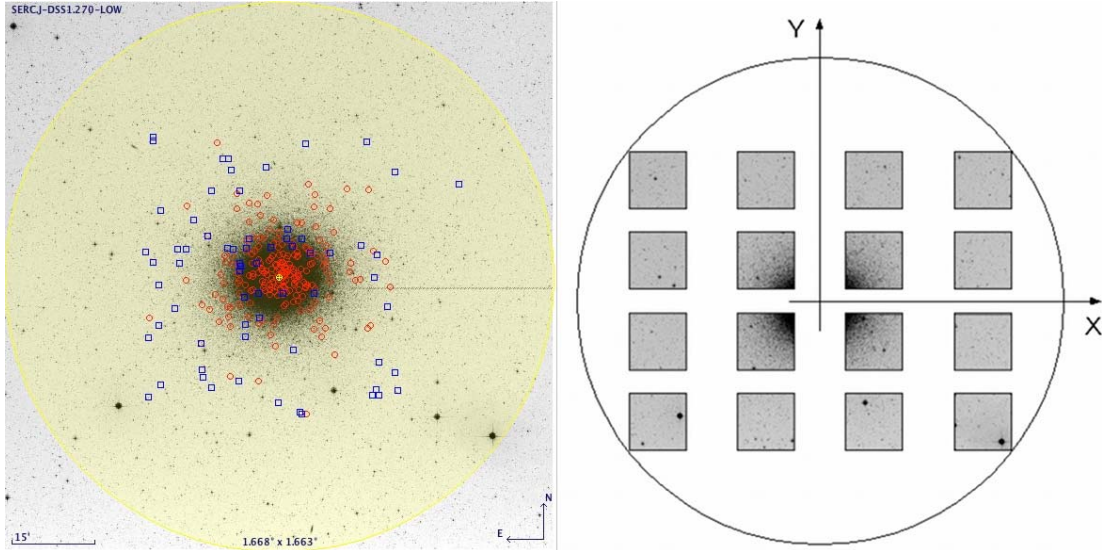
Consider, as an example, our upcoming monitoring, using the VISTA telescope, of the variable star content of  $\omega$  Centauri (Fig. 15). Such a project will lead to high-quality light curves for more than 250 variables, greatly exceeding, both in quantity and in completeness of each individual light curve, what was achieved in previous near-



**Figure 14.**— Same as in Figure 13, but using our newly-defined, reddening-free indices  $m_2$  (eq. 2),  $m_4$  (eq. 4), and  $c_3$  (eq. 7). The *left panel* shows the original  $K_S$ ,  $(J - K_S)$  CMD, whereas the *middle panel* shows the  $m_2$ ,  $c_3$  CMD, and the *right panel* shows the  $m_4$ ,  $c_3$  CMD. Note the remarkable tightening of the cluster’s sequences, when the reddening-free indices are used.

IR studies of the cluster (Del Principe et al. 2006). In particular, since  $\omega$  Cen hosts the largest known population of SX Phoenicis stars, some 75 members, we will be in a position to directly derive, for the first time, a period-luminosity relation for SX Phe stars in the near-IR (Freyhammer & Sterken 2003; Weldrake et al. 2007). In like vein, we will also improve the calibration of the RR Lyrae period-luminosity relation in the near-IR, given that we will obtain, for the first time, complete light curves for most of the 170 RR Lyrae variables in the cluster.

Similarly, we have recently started to monitor a series of open clusters known to host sizeable populations of  $\delta$  Scuti stars. In addition to high-quality near-IR templates,



**Figure 15.**— Upcoming VISTA observations of the globular cluster  $\omega$  Cen. *Left panel:* An example of variable star distribution across the cluster field: red circles mark the positions of known RR Lyrae stars, while blue squares mark the positions of known eclipsing binaries. *Right panel:* VIRCAM@VISTA’s 16 detectors, with  $\omega$  Cen at the center of the focal plane.

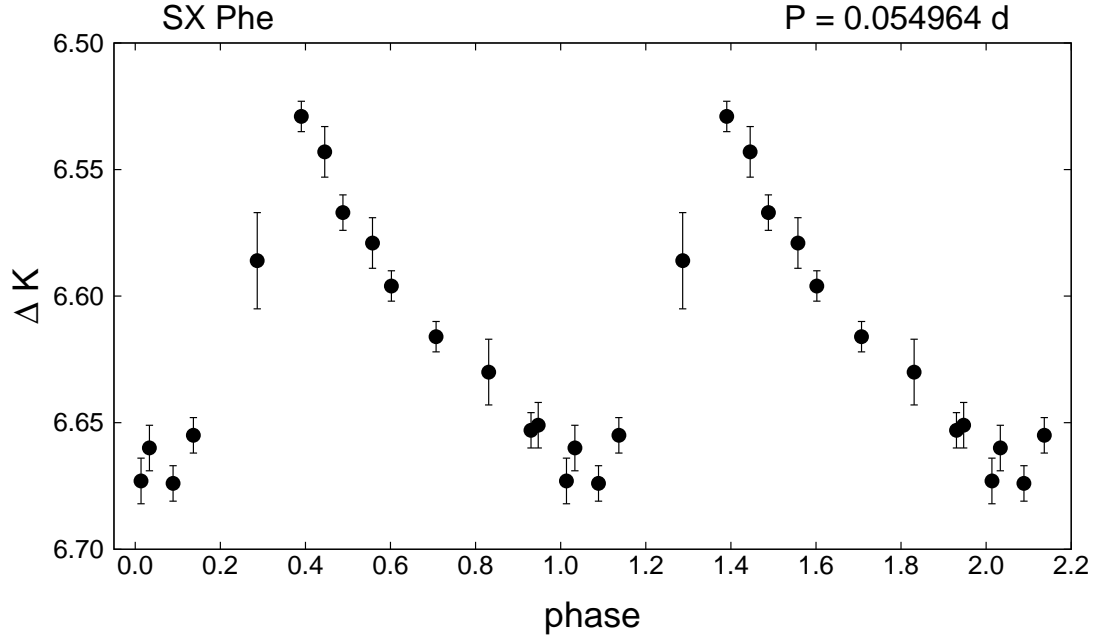
this will also allow us to verify the period-luminosity relation in the near-IR that has been suggested for this important class of variables (King 1990).

Some of the first light curves obtained in the framework of the VVV Templates Project are shown in Figure 16 ( $K$ -band light curve for SX Phe, the prototype of its class) and Figure 17 ( $K$ -band light curve for WY Sco, a classical Cepheid). The SX Phe light curve was obtained in the course of an observing campaign carried out with the SAAO 0.75m photometer in Swinburne, South Africa, whereas the WY Sco light curve has been obtained in the course of our ongoing observations with the REM 0.6m robotic telescope in La Silla, Chile. By the end of our monitoring campaigns for these and other stars, the number of data points per light curve will at least double, as required in order to properly define light curve templates.

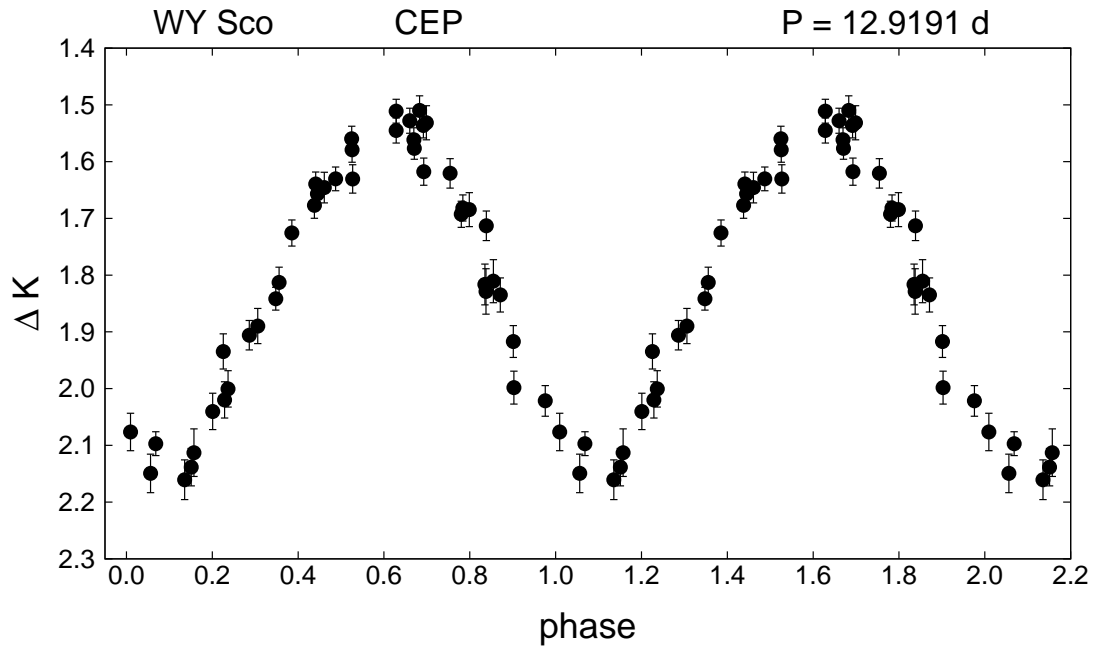
### 4.3. Towards the Automated Classification of VVV Light Curves

In the last few years, there has been an increasing interest towards the application of artificial intelligence algorithms in astronomical research. This has been mainly triggered by the ongoing and future large-scale observational surveys, which are expected to deliver huge amounts of high-quality data, which will need to be promptly analyzed and classified.

An illustrative example of such a need is indeed the VVV Survey, whose final product will be a deep near-IR atlas in five passbands and a catalog with an estimated more than  $10^6$  variable point sources (Minniti et al. 2010; Saito et al. 2010).



**Figure 16.**— *K*-band light curve of SX Phe, taken at the SAAO 0.75m telescope in Sept. 2010. Note the overall amplitude of  $\approx 0.15$  mag.



**Figure 17.**— *K*-band light curve of the classical Cepheid WY Sco, currently being observed with the REM 0.6m telescope in the framework of the VVV Templates Project.

**TABLE 3**  
**VVV Templates Project: Observing Time at Different Observatories**

Observatory	Telescope	IR Camera	FoV	Awarded Time	Target
ESO Paranal	VISTA 4m	VIRCAM	60' × 90'	34 hours	$\omega$ Centauri
	CTIO Blanco 4m	NEWFIRM	28' × 28'	3 nights	NGC 3293/4755/6231
	KASI BOAO 1.8m	KASINICS	3.6' × 3.6'	6 weeks	CVs, $\delta$ Scutis
	IAC TCS 1.4m	CAIN-III	4.2' × 4.2'	36 nights	NGC 1817/7062
	SAAO IRSF 1.4m	SIRIUS	7.7' × 7.7'	3 weeks	M62, NGC 1851/6134
	CTIO SMARTS 1.3m	ANDICAM	2.4' × 2.4'	90h/sem	field RR Lyrae
	SAAO 0.75m	MkII phot	–	3 weeks	SX Phe, Ellipsoidal, ...
ESO La Silla	REM	REMIR	10' × 10'	225h/sem	EBs, $\delta$ Scutis
Asiago <sup>a</sup>	Schmidt	Opt. CCD	52' × 36'	36 nights	NGC 1817/7062
Kazakshtan <sup>b</sup>	1m	Opt. CCD	30' × 30'	6 weeks	CVs
OMM <sup>c</sup>	1.6m	CAPAPIR	30' × 30'	8 hours	NGC 7062
OAGH <sup>c</sup>	2.1m	CANANEA	4' × 4'	10 nights	NGC 1817

<sup>a</sup> Simultaneous optical monitoring of TCS targets.

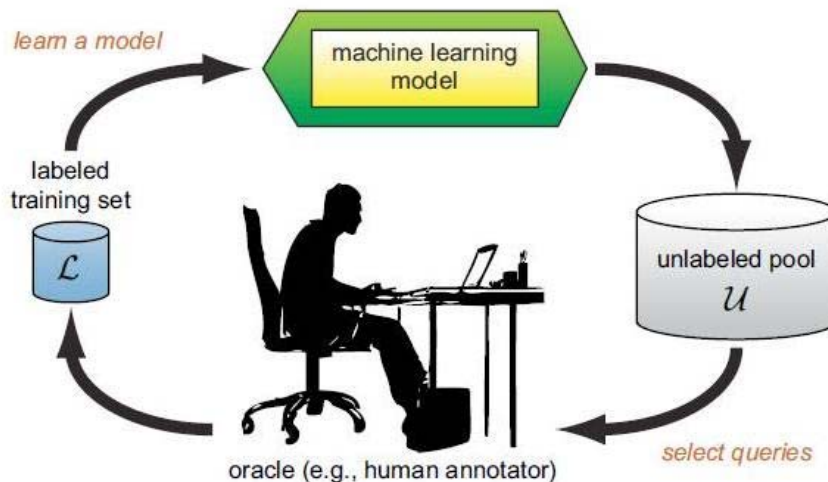
<sup>b</sup> Simultaneous optical monitoring of BOAO targets.

<sup>c</sup> Proposals submitted but still under evaluation by the respective TACs.

Since the majority of these surveys will deal with temporal series, it is natural that most of the effort for developing, adapting and testing such automated classification algorithms has concentrated on the field of variable stars. The reference work in machine learning applied to variable star classification can be considered the paper by Debosscher et al. (2007), in which the authors explored several classification techniques, quantitatively comparing performance (e.g., in terms of computational time) and final results (e.g., in terms of accuracy). More recently, a few other studies have focused on specific methodologies, with the implicit goal of finding the best compromise between robustness and speed (e.g., Blomme et al. 2011; Dubath et al. 2011; Richards et al. 2011).

Irrespective of the specific algorithm, the general idea behind these “supervised machine learning” methodologies is to create a function, *the classifier*, that is able to infer the most probable *label* for an object – in our case, the variability class to which an unclassified variable star belongs to – on the basis of what is learned by the analysis of the *inputs* – in our case, light curve features, including, for instance, periods, Fourier decomposition parameters, colors, etc. – from a *training set*. The latter, in our case, is a collection of high-quality light curves of independently classified variable stars – the *templates*.

In this sense, in Pichara et al. (2011) we show that a significant improvement in the automated classification of astronomical data can be achieved from an optimization of the way *the training set itself* is selected. This is of crucial importance for our purposes, given that the many hundreds of light curve templates that optical automated



**Figure 18.**— The pool-based active learning cycle (adapted from Settles 2010).

classification studies can rely upon (see, e.g., Table 7 in Richards et al. 2011) are not yet available in the near-IR, and will likely – despite our best efforts – not be available in comparable numbers in the near future. Given the availability of such huge template datasets, the precise way in which the training set itself is optimized has never been a major concern in optical studies, but is instead an important aspect of our near-IR automated classification work that must be carefully addressed.

In fact, there are concrete cases in which it is convenient to keep the training set as small as possible, while still retaining its maximum information content. For example, one may need to classify *as quickly as possible* huge amounts of data obtained in “real-time,” such as the 30 TB of data produced every night by LSST (Ivezic et al. 2008). We might also be in the situation for which gathering a training set may be far from straightforward, as in the case of the VVV Survey. In all these cases, an active learning approach can be the key to the problem.

Active learning (Fig. 18) is a form of supervised machine learning in which the learning algorithm is able to interactively query the user (or some other information source). The key hypothesis is that if the learning algorithm is allowed to choose the data from which it learns, it will perform better with less training. In other words, the active learner aims to achieve high accuracy using as few (in our case) template light curves as possible, thereby minimizing the cost of obtaining observational data. Active learning is thus fully justified in many modern machine learning problems where unlabeled data may be abundant (in our case, e.g., the  $10^9$  point sources in the VVV Survey catalog), but labels (i.e., template light curves) are scarce or expensive to obtain. This is the technique that we eventually plan to implement, in order to properly carry out the automated classification of VVV light curves (Pichara et al. 2011).



## 5. Summary

The VVV Survey of the Galactic bulge and inner disk, which has recently started its second year of observations, represents a treasure trove for many different types of studies. Its final data products will include a deep, multi-band catalog containing of order  $10^9$  point sources in the  $ZYJHK_S$  filters, in addition to near-IR light curves for an estimated  $10^6$  variable stars in the surveyed fields. In this contribution, we have provided an overview of the project's main scientific goals and present status. We have also highlighted some of the first scientific results that have been obtained on the basis of its first year of operation, including the discovery of a number of previously unknown globular clusters, open clusters, and stellar associations. Deep color-magnitude diagrams for a few selected systems were presented, and a set of reddening-free indices for analysis of heavily differentially obscured regions introduced.  $K_S$ -band light curves for a few variable stars in the VISTA science verification field were also presented. Finally, we provided an overview of the VVV Templates Project, whose aim is to provide high-quality light curves for the automated classification of VVV near-IR light curves. In this framework, we also argued that active learning may represent an ideal technique for classifying the VVV light curves on the basis of the template set that we are currently building.

**Acknowledgments.** The VVV Survey is supported by the European Southern Observatory, the BASAL Center for Astrophysics and Associated Technologies (PFB-06), the FONDAF Center for Astrophysics (15010003), and the Chilean Ministry for the Economy, Development, and Tourism's Programa Iniciativa Científica Milenio through grant P07-021-F, awarded to The Milky Way Millennium Nucleus. Support for M.C., I.D. and J.A.-G. is also provided by Proyecto Fondecyt Regular #1110326. Support for R.A. is provided by Proyecto Fondecyt #3100029. J.B. acknowledges support from Proyecto Fondecyt Regular #1080086.

## References

- Abazajian, K. N., et al. 2009, *ApJS*, 182, 543  
Alard, C. 2000, *A&AS*, 144, 363  
Alcock, C., et al. 1998, *ApJ*, 492, 190  
Arnaboldi, M., et al. 2010, *The Messenger*, 139, 6  
Benjamin, R. A., et al. 2003, *PASP*, 115, 953  
Bica, E., Bonatto, C., Ortolani, S., & Barbuy, B. 2007, *A&A*, 472, 483  
Blomme, J., et al. 2011, *MNRAS*, in press (arXiv:1101.5038)  
Bonatto, C., Bica, E., Ortolani, S., & Barbuy, B. 2007, *MNRAS*, 381, L45  
Bonatto, C., Kerber, L. O., Bica, E., & Santiago, B. X. 2006, *A&A*, 446, 121  
Borissova, J., et al. 2011, *A&A*, submitted  
Bramich, D. M. 2008, *MNRAS*, 386, L77  
Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, *ApJ*, 345, 245  
Catelan, M. 2009, *Ap&SS*, 320, 261  
Catelan, M., Pritzl, B. J., & Smith, H. A. 2004, *ApJS*, 154, 633  
Dalton, G. B., et al. 2006, *SPIE*, 6269, 62690X

- De Propriis, R., et al. 2011, *ApJ*, 732, L36
- Debusscher, J., Sarro, L. M., Aerts, C., Cuypers, J., Vandebussche, B., Garrido, R., & Solano, E. 2007, *A&A*, 475, 1159
- Del Principe, M., et al. 2006, *Mem. Soc. Astron. Italiana*, 77, 330
- Dubath, P., et al. 2011, *MNRAS*, in press (arXiv:1101.2406)
- Freyhammer, L. M., & Sterken, C. 2003, *Interplay of Periodic, Cyclic and Stochastic Variability in Selected Areas of the H-R Diagram*, 292, 81
- Froebrich, D., Meusinger, H., & Davis, C. J. 2008, *MNRAS*, 383, L45
- Harris, W. E. 1996, *AJ*, 112, 1487
- Irwin, M. J., et al. 2004, *SPIE*, 5493, 411
- Ivanov, V. D., Kurtev, R., & Borissova, J. 2005, *A&A*, 442, 195
- Ivezic, Z., Tyson, J. A., Allsman, R., Andrew, J., Angel, R., & for the LSST Collaboration 2008, arXiv:0805.2366
- Kholopov, P. N., et al. 1998, *Combined General Catalogue of Variable Stars*, 4.1 Ed (II/214A)
- King, J. R. 1990, *PASP*, 102, 658
- Lewis, J. R., Irwin, M., & Bunclark, P. 2010, in *Astronomical Data Analysis Software and Systems XIX*, ASP Conf. Ser., vol. 434, 91
- Martins, F., Pomarès, M., Deharveng, L., Zavagno, A., & Bouret, J. C. 2010, *A&A*, 510, A32
- McWilliam, A., & Zoccali, M. 2010, *ApJ*, 724, 1491
- Minniti, D., et al. 2010, *NewA*, 15, 433
- Minniti, D., et al. 2011, *A&A*, 572, A81
- Moni Bidin, C., et al. 2011, in preparation
- Nataf, D. M., Udalski, A., Gould, A., Fouqué, P., & Stanek, K. Z. 2010, *ApJ*, 721, L28
- Ortolani, S., Bonatto, C., Bica, E., & Barbuy, B. 2009, *AJ*, 138, 889
- Pichara, K., et al. 2011, in preparation
- Portegies Zwart, S. F., McMillan, S. L. W., & Gieles, M. 2010, *ARA&A*, 48, 431
- Richards, J. W., Homrighausen, D., Freeman, P. E., Schafer, C. M., & Poznanski, D. 2011, *MNRAS*, submitted (arXiv:1103.6034)
- Rieke, G. H., & Lebofsky, M. J. 1985, *ApJ*, 288, 618
- Saito, R., et al. 2010, *The Messenger*, 141, 24
- Sakamoto, T., & Hasegawa, T. 2006, *ApJ*, 653, L29
- Settles, B. 2010, *Active Learning Literature Survey*, Computer Sciences Technical Report, Burr Settles, 1648, University of Wisconsin–Madison
- Skrutskie, M. F., et al. 2006, *AJ*, 131, 1163
- Soszyński, I., et al. 2011, *AcA*, 61, 1
- Weldrake, D. T. F., Sackett, P. D., & Bridges, T. J. 2007, *AJ*, 133, 1447