The WFCAM multiwavelength Variable Star Catalog[★]

C. E. Ferreira Lopes¹, I. Dékány^{2,3}, M. Catelan^{2,3}, N. J. G. Cross⁴, R. Angeloni⁵, I. C. Leão¹, and J. R. De Medeiros¹

¹ Departamento de Física Teórica e Experimental, Universidade Federal do Rio Grande do Norte, 59072-970 Natal, RN, Brazil
e-mail: ferreiralopes1011@gmail.com

² Instituto de Astrofísica, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, 782-0436 Macul, Santiago, Chile

³ Millennium Institute of Astrophysics, Santiago, Chile

⁴ SUPA (Scottish Universities Physics Alliance) Wide-Field Astronomy Unit, Institute for Astronomy, School of Physics and Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK

⁵ Gemini Observatory, Colina El Pino, Casilla 603, La Serena, Chile

Received 11 March 2014 / Accepted 12 August 2014

ABSTRACT

Context. Stellar variability in the near-infrared (NIR) remains largely unexplored. The exploitation of public science archives with data-mining methods offers a perspective for a time-domain exploration of the NIR sky.

Aims. We perform a comprehensive search for stellar variability using the optical-NIR multiband photometric data in the public Calibration Database of the WFCAM Science Archive (WSA), with the aim of contributing to the general census of variable stars and of extending the current scarce inventory of accurate NIR light curves for a number of variable star classes.

Methods. Standard data-mining methods were applied to extract and fine-tune time-series data from the WSA. We introduced new variability indices designed for multiband data with correlated sampling, and applied them for preselecting variable star candidates, i.e., light curves that are dominated by correlated variations, from noise-dominated ones. Preselection criteria were established by robust numerical tests for evaluating the response of variability indices to the colored noise characteristic of the data. We performed a period search using the string-length minimization method on an initial catalog of 6551 variable star candidates preselected by variability indices. Further frequency analysis was performed on positive candidates using three additional methods in combination, in order to cope with aliasing.

Results. We find 275 periodic variable stars and an additional 44 objects with suspected variability with uncertain periods or apparently aperiodic variation. Only 44 of these objects had been previously known, including 11 RR Lyrae stars on the outskirts of the globular cluster M3 (NGC 5272). We provide a preliminary classification of the new variable stars that have well-measured light curves, but the variability types of a large number of objects remain ambiguous. We classify most of the new variables as contact binary stars, but we also find several pulsating stars, among which 34 are probably new field RR Lyrae, and 3 are likely Cepheids. We also identify 32 highly reddened variable objects close to previously known dark nebulae, suggesting that these are embedded young stellar objects. We publish our results and all light curve data as the WFCAM Variable Star Catalog.

Key words. stars: variables: general – infrared: stars – dust, extinction – stars: formation – techniques: photometric – methods: analytical

1. Introduction

Time-varying celestial phenomena in general represent one of the most substantial sources of astrophysical information, and their study has led to many fundamental discoveries in modern astronomy. Pulsating stars provide insight into stellar interiors through asteroseismology (e.g., [Handler 2012](#)) and serve as standard candles (e.g., [Walker 2012](#)), eclipsing binaries allow us to determine the most accurate stellar masses and radii (e.g., [Clausen et al. 2008](#)), and supernovae provide important means for estimating cosmological distances and probing the large-scale structure of our Universe (e.g., [Riess et al. 1998](#); [Tonry et al. 2003](#)) – just to mention some classical scopes among the countless aspects of time-domain astronomy.

The ever-growing interest in various astronomical time-series data, as well as the tremendous development in astronomical instrumentation and automation during the past two decades, have been giving rise to several time-domain surveys

of increasing scale. Wide-field shallow optical imaging surveys using small, dedicated telescope systems have been scanning the sky since the early 2000s with aims ranging from comprehensive stellar variability searches to exoplanet hunting, such as the Northern Sky Variability Survey (NSVS, [Hoffman et al. 2009](#)), All Sky Automated Survey (ASAS, [Pojmanski 2002](#)), Wide Angle Search for Planets (WASP, [Pollacco et al. 2006](#)), and the Hungarian Automated Telescope Network (HATnet, [Bakos et al. 2004](#)). The interest in transient events like microlensing have led to deeper, higher-resolution photometric campaigns such as the Massive Astrophysical Compact Halo Objects Survey (MACHO, [Alcock et al. 1993](#)) and the Optical Gravitational Lensing Experiment (OGLE, [Udalski 2003](#)), and to some ultra-wide surveys such as the Palomar Transient Factory (PTF, [Law et al. 2009](#)) or the Catalina Real-time Transient Survey ([Drake et al. 2009](#)). The development of imaging technology is leading toward deep all-sky surveys with seeing-limited resolution, and Pan-STARRS ([Kaiser et al. 2002](#)), representing the first generation of such programs, is already operational. In the near future, even more ambitious programs, such as the Large Synoptic Survey Telescope (LSST, [Krabbendam & Sweeney 2010](#)) and

[★] Tables 4–6 are only available at the CDS via anonymous ftp to [cdsarc.u-strasbg.fr](ftp://cdsarc.u-strasbg.fr) (130.79.128.5) or via <http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/573/A100>

Gaia (Perryman 2005), are planned to start monitoring the optical sky. These synoptic surveys are expected to be the new powerhouses of modern astronomy by providing a high data flow for a wide range of science applications, from the study of transients to understanding the dynamics of the Milky Way galaxy.

While optical synoptic surveys are getting wider and deeper, extending the systematic exploration of the variable sky toward other wavelengths, such as the infrared, is also indispensable, not only for a more complete understanding of the observed phenomena, but also for overcoming the problem of interstellar extinction. Infrared time-series photometry had long been constrained mostly to follow-up observations of known objects, but in recent years near-infrared (NIR) imagers have been converging to optical CCDs in terms of resolution and performance, hence wide-field surveys became feasible with some NIR instruments, such as the Wide-Field near-IR CAMera (WFCAM; Casali et al. 2007) on the 3.8 m United Kingdom Infrared Telescope (UKIRT), and the VISTA InfraRed CAMera (VIRCAM; Dalton et al. 2006) on the 4.1 m Visible and Infrared Survey Telescope for Astronomy (VISTA), both hosting a variety of deep Galactic and extragalactic surveys. Nevertheless there are only a handful of wide-field NIR time-domain surveys that have more than a handful of observational epochs, with only the VISTA Variables in the Vía Láctea survey (VVV; Minniti et al. 2010) being comparable to the optical ones in terms of both areal and time-domain coverage (e.g., Arnaboldi et al. 2007, 2012).

Besides the results from the large-scale surveys, valuable time-series data are also being generated by a variety of observational programs as sideproducts. Such data can well be quite heterogeneous and not always be complete completeness, since the original design of the data acquisition might have served various specific purposes. Nevertheless, multi-epoch data accumulated over a period of time from various observational programs using the same facility can hold vast unexploited information and represent a potential treasure trove for time-domain astronomy (see, e.g., the TAROT variable star catalog by Damerdjian et al. 2007). An increasing number of observatories and projects realize the importance of standardizing their archives and incorporating them into the Virtual Observatory, allowing the community to exploit their data further once their proprietary periods have expired. There is already a wealth of fully public science-ready synoptic data from state-of-the-art instruments, which are accessible by standard data mining tools, waiting for analysis.

This paper is based on public time-domain data from the WFCAM Science Archive (WSA; Hambly et al. 2008). WFCAM was designed to be capable of carrying out ambitious large-scale survey programs such as the UKIRT Infrared Deep Sky Surveys (UKIDSS; Lawrence et al. 2007). The detector consists of four arrays of 2048×2048 pixels, arranged in a non-contiguous square pattern, providing a resolution of $0.4''$ per pixel. A contiguous image with an areal coverage of 0.78 deg^2 can be constructed by a mosaic of four consecutive offset pointings. Other properties of the instrument are discussed in full detail by Casali et al. (2007). The standard UKIRT photometric system consists of the MKO J , H , and K bands (see Simons & Tokunaga 2002; Tokunaga et al. 2002) complemented with the Z and Y filters. In addition, three narrow-band filters (H_2 , Bry , and nbj) are also available (see Hewett et al. 2006 for further details).

All WFCAM data produced by the UKIDSS surveys and other smaller campaigns and PI projects employing the instrument are processed by the VISTA Data Flow System (VDFS; Emerson et al. 2004). VDFS was designed to initially

handle UKIDSS, and includes both pipeline processing at the Cambridge Astronomical Survey Unit (CASU¹) and further pipeline processing of the CASU products and digital curation at the WSA. Data for standard photometric zero-point calibration of and color term determination for the WFCAM filters are taken regularly and are also fully processed and archived by the VDFS in the same way as the survey data. During the several years of operation of WFCAM, a large amount of fully processed, high-quality, multiband photometric data have been collected during the observations of calibration fields, which are publicly available at the WSA. Since the majority of these fields have been visited several times over many years, these data provide an excellent opportunity for studying variable stars in the NIR.

In this paper, we perform a comprehensive stellar variability analysis of the public WFCAM Calibration (WFCAMCAL) Database (release 08B) and present the photometric data and characteristics of the identified variable stars as the WFCAM Variable Star Catalog, version 1 (WVSC1). The paper is structured as follows. In Sect. 2, we present the database and its characteristics from a data miner's point of view and describe the primary source selection procedure. In Sect. 3, we introduce and discuss a set of variability indices designed for synoptic data with correlated sampling, and we employ these indices as a first selection of candidate variable sources in the WFCAM data. We present a frequency analysis of the candidate variable sources in Sect. 4. In Sect. 5, we present the WVSC1 by describing its structure and discussing its variable star content. Finally, in Sect. 6, we draw our conclusions and discuss some future perspectives.

2. The WFCAMCAL database

WFCAM on-sky image data are pipeline-processed by the VDFS, which incorporates all data reduction steps from image processing to photometry. The various processing steps are described fully by Emerson et al. (2004), and accordingly we limit ourselves to discussing some key aspects. Individual exposures go through the standard preprocessing steps, such as flat fielding and dark subtraction. As is common practice in the NIR, science frames are composed of a set of dithered images, i.e. subsequent exposures taken in step patterns of several arcseconds, in order to allow for the removal of bad detector pixels and to image the structure of the atmospheric IR foreground for its subtraction before stacking the dither sequence. These so-called *detector frame* stacks are the final image products of calibration field observations, but we note that some WFCAM survey images are combined further to form contiguous mosaic images also known as *tiles*. Deep stack images are also produced, in order to push the limiting magnitudes and measure the fluxes of faint sources.

Further steps in the reduction include source extraction and aperture photometry using a set of small circular apertures with radii of 0.5 , $\sqrt{2}/2$, 1 , $\sqrt{2}$, and 2 arcsec, in order to maximize the signal-to-noise ratio and remedy systematics due to source crowding (Cross et al. 2009). Flux loss in the wings of the point spread functions (PSFs) is corrected for Irwin et al. (2004). Each identified source of the catalogs is classified based on the shape of its PSF, and various quality flags are also assigned. Astrometric calibration is performed using 2MASS (Skrutskie et al. 2006) stars as reference, and positions have a typical accuracy of $0.1''$. Magnitudes are corrected for distortion effects and are zero-point-calibrated on a frame-by-frame basis using

¹ <http://casu.ast.cam.ac.uk/surveys-projects/wfcam/technical/photometry>

Fig. 1. Celestial distribution of our initial database of stellar sources from the WFCAMCAL08B archive.

a many of secondary standard stars from the 2MASS catalog (Skrutskie et al. 2006). This involves color corrections from the 2MASS *JHK_s* to WFCAM *ZYJHK*. The calibration process is described in detail in Hodgkin et al. (2009). Many calibration fields (which form the basis of this paper) are also observed repeatedly under photometric conditions, in order to provide standard star measurements for the calibrations of the *Z*, *Y*, and narrow-band magnitudes, as well as to provide data for the color-term determination for transforming 2MASS magnitudes into the WFCAM photometric system. The accuracy of the magnitude zero-points of these fields is a few percent. We note that all magnitude data in the WSA and in our study alike are on the WFCAM magnitude system.

The calibrated source catalogs undergo further curation steps at the WSA, including quality control, source merging, and both internal and external cross-matching. Enhanced image products (e.g., deep stacks) are also created. Final science-ready data are ingested into a relational database, which offers various server-side data management tools. Data can be queried by using the Structured Query Language (SQL). The design of the WSA, the details of the data curation procedures and the layout of the database are described in great detail in Hambly et al. (2008) and Cross et al. (2009). In the following, we highlight some other properties of the data in the WFCAMCAL archive, with an emphasis on some important details for the time-series analysis.

2.1. Data characteristics

The WFCAMCAL archive’s data release 08B contains data from 52 individual pointings from both the northern and southern hemispheres, spread over nearly half of the sky. These calibration fields are located between declinations of $+59^{\circ}62'$ and $-24^{\circ}73'$, distributed over the full range in right ascension in order to provide standard star data year-round. The positions of the pointings are shown Fig. 1. Each pointing consists of a non-contiguous area covered by the four WFCAM chips, covering ~ 0.05 square degrees each. The total area covered by all standard star fields is ~ 10.4 square degrees.

The majority of the fields are observed repeatedly, with a rather irregular sampling. A field is usually visited at most a few times during a night, with the visits separated by up to a few hours. During each visit, the field is usually (but not necessarily) observed through the *JHK* or the *ZYJHK* filter set, and

Fig. 2. Spectral window of a typical well-sampled *K*-band light curve from the WFCAMCAL database, showing two different frequency ranges.

occasionally through the narrow-band filters as well, all within a few minutes. A certain field is usually observed again within a few days, although longer time gaps are common, and of course large seasonal gaps are also present in the data set. The total baseline of the time series is largely field-dependent and varies from a few months up to three years. The time sampling (cadence) in a single passband can be considered to be quasi-stochastic with rather irregular gaps, which is a favorable scenario for detecting periodic signals over a wide range of periods. Figure 2 shows the spectral window of a typical one among the well-sampled *K*-band light curves in the database, showing that only one-day and one-year aliases are significant. However, the sampling in different passbands is very strongly correlated.

The VDFS performs an accurate and deep source extraction procedure on deep stack images, which, in the case of the WFCAMCAL database, is merged from the seven best-seeing frames to reduce any problems with blending (Cross et al. 2009). This curation step gives rise to database entities known as *Sources*. Positional cross-matches between these *Sources* and detections from different single detector frame stacks are then performed for each frame set (i.e., images with the same pointing). The prescriptions for this cross-matching procedure and its limitations are given by Hambly et al. (2008, see their Sect. 3.4.2). Each *Source* of the database contains only partial information from a set of source detections positionally matched with each other, but is related to a set of pointers that track all the attributes of the corresponding detections. We note that after cross-matching, the VDFS also recalibrates the individual epochs after comparison to the deep stack to improve the light curves. The resulting source-merged catalog is stored in WSA’s relational database, which in the case of WFCAMCAL data, follows a synoptic database model designed for observations with correlated sampling (Cross et al. 2009). Since the observations of the standard star fields through different filters are always taken within a time interval that is several orders of magnitude shorter than the interval between two successive observation batches (i.e., visits of the same field, see above), they can be considered to be virtually at the same epoch. Individual source detections in different filters for each epoch (i.e., filter sequence) are therefore merged and stored in a database entity called synoptic source, and these are linked to *Sources* (see above). Various metadata of the time series formed by the detections associated to each *Source* are also provided by WSA as the attributes of the

Table 1. Morphological classification of sources in the WFCAMCAL archive.

Class	Description	Number of sources
-1	Star	318 995
-2	Probable star	33 188
+1	Galaxy	205 352
-3	Probable galaxy	2345
0	Noise	10 377

Table 2. Number of sources, average total baseline ($\langle T_{\text{tot}} \rangle$, in days), and average number of epochs ($\langle N_{\text{ep}} \rangle$) in each filter, in our initial database.

Filter	N	$\langle T_{\text{tot}} \rangle$	$\langle N_{\text{ep}} \rangle$
Z	204 245	1033	51
Y	212 334	1061	56
J	212 717	1058	58
H	212 169	1074	61
K	201 645	1065	60
H ₂	110 587	429	4
B _Y	130 307	258	2
nbj	68 048	120	3

variability table, providing the best aperture for a given source, together with basic statistics on the magnitude distribution, time sampling, etc. In passing we note that this table also provides a basic assessment on the probability that a source is variable in each band, based on the comparison of the rms scatter of the time series to its expected value from a simple noise model (see Cross et al. 2009). Although these pieces of information can provide useful guidance for the user, we opted not to use these attributes as criteria in our source selection, since our aim is to perform a more profound variability search in the data.

2.2. Initial source selection

Data were retrieved in a two-step procedure via WSA's freeform SQL query facility². We queried all *Sources* that were classified as a star or probable star based on the PSF statistics of the merged source, having at least ten unflagged epochs in either of the eight filter passbands.

Our selection resulted in an initial database of 216 722 sources containing their identifiers and general attributes such as positions, basic flux, and sampling statistics. Figure 3 shows histograms of the queried light curves and the average total baseline length as a function of the number of epochs, while Tables 1 and 2 summarize some basic statistics of the queried data. The sampling of the time domain is rather heterogeneous, thus we must emphasize that the WFCAMCAL archive should not be considered as a synoptic survey database, for its completeness is highly varying from field to field; accordingly, care must be taken with any statistical interpretation based on these data. However, there are a large number of light curves with very good sampling that have ~ 100 data points in three to band broad-band filters, with a total baseline of up to 3 years. Since the number of data points in the narrow-band filters is generally very low, we do not use these measurements for the variability search, but they are provided (if available) for the detected variable sources in our catalog.

With the result of our first selection query at hand, we queried the complete light curves of each source by linking their

Fig. 3. Number of sources (N_s) as a function of the number of epochs (N_{ep}) for the queried J- (left) and Z-band (right) light curves with a bin size of 10 days. The distribution corresponding to the other broad bands are very similar. The solid curves show the average total baseline length as a function of the number of epochs for the queried light curves in the same filters with a bin width of 2 days.

entries in the *Source* table to the attributes of the *Synoptic Source* table. This procedure requires the generation of a temporary SQL table provided by the user containing the identifiers of the *Sources* from the first query. We note that some data, such as Julian Dates and zero-point errors, are stored in different database objects like the *Multiframe* table, which require the execution of further table joins. We selected magnitudes with sufficiently small error bits (< 128), rejecting highly saturated measurements and data affected by various severe defects (e.g., bad pixels, poor flat-field region, detection close to frame boundary, etc.).

3. Broad selection by variability indices

A key characteristic of the time-series data in the WFCAMCAL database is that, owing to the observation strategy, their sampling is strongly correlated in the different filters (see Sect. 2.1). The characteristic time lag between data points in different filters in an observing sequence (i.e., a standard visit of a calibration field) is not longer than a few minutes, which is orders of magnitude shorter than the typical time gap between such batches of data and also than the time scales of stellar variability at amplitudes that are typically recoverable by the present sampling and accuracy. This property of the data enables us to search for stellar variability through correlations between the temporal flux changes at different wavelengths.

In our approach to stellar variability searching, we make the following general assumptions: (i) intrinsic stellar variability is typically identifiable in a wavelength range that is wider than our broad-band filters (i.e., in more than one filter); (ii) it is sufficiently phase-locked at two close wavelengths, thus flux variations in neighboring wavebands will be correlated; and (iii) non-intrinsic variations will be typically stochastic. In point (iii) we also implicitly assume that wavelength-correlated systematics of instrumental and atmospheric origin, or due to possible data reduction anomalies, have amplitudes that are small to be enough comparable to those provided by stochastic variations. The most important deviation from this assumption for our data is due to

² http://surveys.roe.ac.uk:8080/wsa/SQL_form.jsp

the temporal saturation of bright objects, which can, however, be easily distinguished from stellar variability.

A commonly used approach to identifying stellar variability through correlations in flux changes is by the Welch-Stetson index (I_{WS} ; Welch & Stetson 1993; Stetson 1996). The idea behind this index is to separate stochastic variations from systematic trends by measuring the correlations in the deviations from the mean of data points that are located sufficiently close in time. It is defined as

$$I_{WS} = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n (\delta b_i \delta v_i)}, \quad (1)$$

where

$$\delta b_i = \frac{b_i - \bar{b}}{\sigma_{b,i}}, \quad \delta v_i = \frac{v_i - \bar{v}}{\sigma_{v,i}}, \quad (2)$$

with b_i , v_i , and $\sigma_{b,i}$, $\sigma_{v,i}$ denoting magnitudes and their errors (the latter computed following the prescriptions by Stetson 1981), respectively, taken on a time scale that is much shorter than that of the intrinsic stellar variations of interest. These magnitudes and errors can represent either successive monochromatic measurements or data points taken in two different filters. We also note that $\bar{b}$, $\bar{v}$ are weighted averages. Finally, n denotes the number of epochs, by which we also refer to the number of short time slots that contain subsequently taken measurements in the case of non-simultaneously observed data with strongly correlated sampling. The I_{WS} index is found to be significantly more sensitive than the “traditional” χ^2 -test for single variance, which uses the magnitude-rms scatter distribution of the data as a predictor (see, e.g., Pojmanski 2002).

3.1. Extension of I_{WS} to multiband data sets

According to the definition given in Eq. (1), I_{WS} is limited to pairwise comparisons of fluxes. When it is applied its application to panchromatic data with measurements taken in several wavebands with correlated sampling, one would need to define some conversion of the pairwise indices into a single quantity. To accomplish this, we introduce the following modification to I_{WS} , for quantifying panchromatic flux correlations (pfc):

$$I_{pfc}^{(2)} = \sqrt{\frac{(n_2 - 2)!}{n_2!}} \sum_{i=1}^n \left\{ \sum_{j=1}^{m-1} \left[\sum_{k=j+1}^m (\delta u_{ij} \delta u_{ik}) \right] \right\}, \quad (3)$$

where n is the number of epochs, m is the number of wavebands, u_{ij} are the flux measurements, δu_{ij} is defined by Eq. (2) for each filter, and $n_2 = n \cdot m!/[2!(m-2)!]$. We note that, for $m = 2$, $I_{pfc}^{(2)} = I_{WS}$.

While Eq. (2) still measures variability through pairwise correlations of relative fluxes, we can generalize it for quasi-simultaneously measured batches of three data points (in case of $m \geq 3$) by

$$I_{pfc}^{(3)} = \sqrt{\frac{(n_3 - 3)!}{n_3!}} \sum_{i=1}^n \left\{ \sum_{j=1}^{m-2} \left[\sum_{k=j+1}^{m-1} \left(\sum_{l=k+1}^m \Lambda_{ijkl}^{(3)} |\delta u_{ij} \delta u_{ik} \delta u_{il}| \right) \right] \right\}, \quad (4)$$

where $n_3 = n \cdot m!/[3!(m-3)!]$, and the Λ factor is defined as

$$\Lambda_{ijkl}^{(3)} = \begin{cases} +1 & \text{if } \delta u_{ij} > 0, \delta u_{ik} > 0, \delta u_{il} > 0; \\ +1 & \text{if } \delta u_{ij} < 0, \delta u_{ik} < 0, \delta u_{il} < 0; \\ -1 & \text{otherwise.} \end{cases} \quad (5)$$

We note that we introduced $\Lambda^{(3)}$ to give the proper sign to the product in Eq. (4).

Finally, in the case of quasi-simultaneously measured batches of s data points (for $m \geq s$), our variability index takes the following general form:

$$I_{pfc}^{(s)} = \sqrt{\frac{(n_s - s)!}{n_s!}} \sum_{i=1}^n \left[\sum_{j_1=1}^{m-(s-1)} \cdots \left(\sum_{j_s=j_{(s-1)+1}}^m \Lambda_{ij_1 \cdots j_s}^{(s)} |\delta u_{ij_1} \cdots \delta u_{ij_s}| \right) \right], \quad (6)$$

where $n_s = n \cdot m!/[s!(m-s)!]$, and the $\Lambda^{(s)}$ correction factor is

$$\Lambda_{ij_1 \cdots j_s}^{(s)} = \begin{cases} +1 & \text{if } \delta u_{ij_1} > 0, \dots, \delta u_{ij_s} > 0; \\ +1 & \text{if } \delta u_{ij_1} < 0, \dots, \delta u_{ij_s} < 0; \\ -1 & \text{otherwise.} \end{cases} \quad (7)$$

Clearly, these indices set increasingly strict constraints on the presence of variability with increasing order s .

3.2. An alternative variability index

While the $I_{pfc}^{(s)}$ index is quite robust, in the sense that it is weighted with the individual errors, it can be insensitive to true variable stars for one or several substantially outlying data points when incorrect error estimates are present. Indeed, this index may even introduce false variability candidates if these outliers are correlated in two or more bands. Although such situations might seem rare, NIR data in particular can present us with these cases quite frequently, particularly in the case of bright stars. Since the sky foreground emitted by the atmosphere is highly variable in the NIR, it causes a highly time-varying saturation limit, which can affect large parts of otherwise highly accurate time-series data for bright stars with substantial outliers having very small formal error estimates. In case of correlated sampling, these outliers will probably be correlated between different filters, leading to a spurious impact upon the $I_{pfc}^{(s)}$ index.

To alleviate the effect of such anomalous outliers, we introduce an alternative variability index that is similar to $I_{pfc}^{(s)}$, but does not depend on the actual value of the flux deviations from the mean. This is simply obtained by keeping only the Λ function (Eq. (7)) in the sum that appears in Eq. (6). Thus, the flux-independent (fi) version of $I_{pfc}^{(2)}$ is defined as

$$2 \cdot I_{fi}^{(2)} - 1 = \frac{1}{n_2} \sum_{i=1}^n \left[\sum_{j=1}^{m-1} \left(\sum_{k=j+1}^m \Lambda_{ijk}^{(2)} \right) \right], \quad (8)$$

with $\Lambda_{ijk}^{(2)}$ defined as

$$\Lambda_{ijk}^{(2)} = \begin{cases} +1 & \text{if } \delta u_{ij} > 0, \delta u_{ik} > 0; \\ +1 & \text{if } \delta u_{ij} < 0, \delta u_{ik} < 0; \\ -1 & \text{otherwise.} \end{cases} \quad (9)$$

In this expression, the δu 's have the same meaning as before (e.g., Eq. (4)).

The righthand side of Eq. (8) thus gives the difference between the number of positive and negative terms in $I_{pfc}^{(s)}$, so it can only take a number of discrete values (depending on the values of n and m). We note that, with the coefficients included on the lefthand side of Eq. (8), $I_{fi}^{(2)}$ will always have absolute values between 0 and 1, analogously to a probability measure.

For higher orders s , $I_{\text{fi}}^{(s)}$ is defined similarly to Eq. (6):

$$2 \cdot I_{\text{fi}}^{(s)} - 1 = \frac{1}{n_s} \sum_{i=1}^n \left[\sum_{j_1=1}^{m-(s-1)} \cdots \left(\sum_{j_s=j_{(s-1)}+1}^m \Lambda_{ij_1 \dots j_s}^{(s)} \right) \right], \quad (10)$$

with $0 \leq I_{\text{fi}}^{(s)} \leq +1$, and where $\Lambda_{ij_1 \dots j_s}^{(s)}$ is defined as in Eq. (7).

Thus, the index $I_{\text{fi}}^{(s)}$ represents a problem of combining signals whose function $\Lambda_{ij_1 \dots j_s}^{(s)}$ (Eq. (7)) assumes the values $+1$ or -1 , depending on the sign of the correlation. The total number of possible combinations of signals in a set of s measurements, so that each group of signals is different from the other, is given by

$$A_s = s^2. \quad (11)$$

For the index $I_{\text{fi}}^{(2)}$, for instance, we have four possible configurations, namely $(++, +-, -+, --)$. According to elementary probability theory, in the case of statistically independent events, the probability that a given event will occur is obtained by dividing the number of events of the given type by the total number of possible events. For the $I_{\text{fi}}^{(s)}$ index of any order, the desired events will be those in which all signals are either positive or negative, so that, irrespective of the value of s , the number of events desired always equals two, and the number of possible events is given by Eq. (11). In this way, the general expression that determines the probability value of a random event leading to a positive $I_{\text{fi}}^{(s)}$ index is given by

$$P_s = \frac{2}{s^2}, \quad (12)$$

given that there are s^2 events in total, but only two produce a positive $I_{\text{fi}}^{(s)}$ value.

3.3. Numerical tests

To establish robust criteria for selecting variable star candidates, we evaluated the responses of the $I_{\text{pfc}}^{(s)}$ and $I_{\text{fi}}^{(s)}$ indices to statistical fluctuations. We generated a large number of test time-series sequences by shuffling the times (“bootstrapping”) of the WFCAMCAL data proper. By following this approach, we are able to keep part of the correlated nature of the noise intrinsic to the data, as opposed to numerical tests based on pure Gaussian noise. Figure 4 shows the histogram of the number of test time-series sequences as a function of the n_s number of terms appearing in the variability indices (see Eqs. (6), (10)) for various s values. In total, 200 000 realizations were performed to our tests.

Figure 5 shows the distribution of the variability indices as a function of the average apparent brightness of the sources for both the bootstrapped and the corresponding original data. The excess of high values for bright sources, which is particularly evident for $s = 2$, is due to temporal saturation, which usually happens in more than one waveband at the same time. In this bright regime, as expected (Sect. 3.2), saturation affects the $I_{\text{pfc}}^{(s)}$ index much more frequently than it does $I_{\text{fi}}^{(s)}$. At the faint end, however, a number of effects can also be recognized. First, the quantized nature of the $I_{\text{fi}}^{(s)}$ index (as opposed to the $I_{\text{pfc}}^{(s)}$ index; see Sect. 3) becomes pronounced in the distribution, owing to the higher relative frequency of sources with but a few epochs (typically $n_s \lesssim 20$) among the faint stars. (These are often not detected if the atmospheric foreground flux is too high.) Second, there is also an excess of sources with high index values at the

Fig. 4. Number of combinations n_s in the simulations described in the text. Since the $I_{\text{fi}}^{(s)}$ index is computed in a manner similar to that used to compute the $I_{\text{pfc}}^{(s)}$ index, the number of combinations used is the same for both.

faint end, which is due to the much lesser amount of data for these fainter sources, which makes the variability indices more sensitive to statistical fluctuations and systematics. The general trend in the distributions of the $I_{\text{pfc}}^{(s)}$ and $I_{\text{fi}}^{(s)}$ are also rather different. Because it is sensitive only to the consistency in the direction of the flux changes, the $I_{\text{fi}}^{(s)}$ is equally responsive to contaminating systematics of different amplitudes. This is reflected by both the slight increase in the main locus of the distribution of this index as a function of magnitude toward the bright end (see Fig. 5, lower panels) and the enhanced (false) response in the faint end, as discussed earlier. These are both caused by the increasing dominance of correlated noise over photon noise. We note that this is not shown by the bootstrapped data at the bright end, since the points of distinct *pawprint* batches were shuffled, and thus the correlations between them were mostly lost, while this is alleviated by the small number of points per light curve at the faint end.

Also noteworthy is that the simulated $I_{\text{fi}}^{(s)}$ distributions that are shown in Fig. 5 are centered on the theoretically expected values (see Sect. 3), as given by Eq. (12), namely 0.5 in the case of $s = 2$, 0.25 in the case of $s = 3$, and 0.125 for $s = 4$. Moreover, the simulation results for the $I_{\text{pfc}}^{(s)}$ index show that the higher the order of the index, the lower the dispersion around $I_{\text{pfc}}^{(s)} = 0$; however, a pronounced level of scatter remains at both the bright and faint ends of the distribution. In addition, there is a symmetry between scatterings for the index $I_{\text{pfc}}^{(2)}$ and asymmetric scattering

Fig. 5. Variability indices versus the K -band magnitude, for the simulations (*top two rows*) and the actual data (*bottom two rows*), for three values of s , namely 2 (*left*), 3 (*middle*), and 4 (*right*).

for $I_{\text{pfc}}^{(3)}$ and $I_{\text{pfc}}^{(4)}$, both of which primarily scatter toward negative values, particularly at the bright end. This happens because the number of possible configurations of signs that lead to a $\Lambda = +1$ (Eq. (7)) is always equal to two, whereas the number of possible configurations that lead to a $\Lambda = -1$ is instead given by $s^2 - 2$. In this way, for $s = 2$, we have a symmetrical distribution of configurations of signs, whereas for $s > 2$ the asymmetry in relation to negative values grows as $s^2 - 2$, as seen in the actual simulations.

We estimated the significance levels of our variability indices based on their cumulative number density distributions obtained from our bootstrapped data set. In other words, we used our bootstrap simulations to set the signals apart that are compatible with pure random noise from the signals that indicate correlated

variability in the different bandpasses. We expect these estimates to be highly dependent on the number of epochs n and the brightness (see above); accordingly, we obtained separate estimates for data within narrow ranges of n and K -band average magnitude, with values in between being obtained by interpolation. Figure 6 shows the result. In this figure, we show surfaces, in (index, n_s, K) space that set random noise apart from correlated signals at the 0.5% significance level, according to our two variability indices, namely $I_{\text{pfc}}^{(s)}$ (top row) and $I_{\text{fi}}^{(s)}$ (bottom row). As can be seen from Fig. 6, the corresponding cutoff values of the $I_{\text{fi}}^{(s)}$ variability index show a relatively weak dependence on the magnitude and a strong dependence on n_s , whereas the $I_{\text{pfc}}^{(s)}$ index depends strongly on both of these quantities. This is mainly attributed to the fact that the mean values of the formal errors

Fig. 6. As in Fig. 5, but for the cutoff surfaces used to select our target list, and also adding n_i as an independent variable. These surfaces set apart, at the 0.5% significance level, instances that are compatible with random noise (low values of the indices) from those that are compatible with coherent signal being present in the different passbands (high values of the indices).

(estimated from photon noise only), to which $I_{\text{pfc}}^{(s)}$ is very sensitive, increasingly underestimate the global scatter of the light curves toward lower magnitudes in this data set.

3.4. Initial selection of variable star candidates

We next computed the variability indices introduced in Sect. 3 for all sources in our initial database. As many as 99.3% of these sources had quasi-simultaneous observations in at least two filters, while 91.3% and 81.3% of them were observed in at least three and four filters, respectively. For each variability index, we selected all sources above its empirical, sampling, and magnitude-dependent 0.5% significance level described in Sect. 3.3. We considered a source as a candidate variable star if either value of its variability indices was significant. This procedure resulted in our initial catalog of 6651 candidate variable stars.

4. Frequency analysis

There are many methods available for computing periods in unevenly spaced time-series data, based mainly on Fourier analysis, information theory, and statistical techniques, among others (see, e.g., [Templeton 2004](#), for a review). Some of these methods are based on the fact that the phase diagram of the light curve (also known as simply “phased light curve”) is smoothest when it is visualized using its real period. They transform the set of data by folding within the phase interval $0 \leq \varphi(i) < 1$, which is

defined by the following expression:

$$\varphi(i) = \frac{t(i) - t_0}{P} - \text{INT} \left[\frac{t(i) - t_0}{P} \right], \quad (13)$$

where t is the time, t_0 is the time origin, P is a test period, and INT denotes the integer function. Because of the presence of observational gaps, it often happens that computed phases can have the same numerical values for several different test periods. Many periods are spurious because they correspond to gaps that may be present in the data (e.g., daytime, seasons, etc.; [Lafler & Kinman 1965](#)). If P is the true period and P_{gaps} the period of gaps ([Damerdjian et al. 2007](#)), spurious periods are given by

$$P_{\text{spur}}^{-1} = P^{-1} \pm k P_{\text{gaps}}^{-1}, \quad (14)$$

where $k \in \mathbb{N}$. In some cases, a spurious period could be ranked as the best period.

We searched for periodic variable stars in our initial catalog (see Sect. 3.4) by applying various frequency analysis methods in combination. To search for the best period (or, equivalently, frequency), many methods require that a search range (given by limiting frequencies f_0 and f_N at the low- and high-frequency ends, respectively) and resolution Δf be specified first. Since we are dealing with data sets with various different time spans, we adapted the f_0 low-frequency limits to each light curve by $f_0 = 0.5/T_{\text{tot}}$, where T_{tot} is the total baseline of the observations, and we scaled the Δf frequency resolution bas $\Delta f = 0.1/T_{\text{tot}}$. The method described by [Eyer & Bartholdi \(1999\)](#) was used to determine the high-frequency limit f_N .

Initially, we applied the string-length minimization (SLM; Lafler & Kinman 1965; Stetson 1996) method on the light curves in our initial catalog of candidate variable stars. In this method, the period is found by minimizing the sum of the lengths of the segments joining adjacent points in the phase diagrams (called the “string lengths”) using a series of trial periods, i.e.:

$$\Phi = \frac{\sum_{i=1}^{N-1} w_{i,i+1} |m_{i+1} + m_i|}{\sum_{i=1}^{N-1} w_{i,i+1}}, \quad (15)$$

where

$$w_{i,i+1} = \frac{1}{(\sigma_{i+1}^2 + \sigma_i^2)(\varphi_{i+1} - \varphi_i + \epsilon)}, \quad (16)$$

and φ_i , m_i , and σ_i are phases, magnitudes, and corresponding magnitude uncertainties, respectively, sorted in order of increasing phase according to the trial period. The variable ϵ is a term that is added to reduce the weight of very closely spaced data in phase space, which otherwise could have a weight approaching infinity (Stetson 1996). In our work, we assumed $\epsilon = 1/N$, again following Stetson (1996).

To increase the probability that among the higher peaks of the periodogram derived on the basis of the SLM method the best period is included, we took the following additional steps:

- the SLM periodogram was independently computed for each broadband filter (Y , Z , J , H , and K), based on which all periods that presented a power greater than the 3σ level every band were selected;
- the SLM periodogram was also computed for the chromatic light curve, comprised of the sum of all broadband filters, and again all periods above the 3σ level were selected;
- the results of the previous two steps were then combined, yielding the best periods (i.e., the ones with the highest amplitude peaks) from both types of analysis.

These steps are very important, because the same source can have photometry of high quality in some filters, but not others.

In the second step of the period search, we applied three additional frequency analysis methods, to select the ten best among the periods selected by the SLM method. For this step, we used the “classical” Lomb-Scargle (LS) periodogram (Lomb 1976; Scargle 1982), the generalized Lomb-Scargle (GLS) periodogram (Zechmeister & Küster 2009), and the phase dispersion minimization (PDM; Stellingwerf 1978) methods. We included both LS and GLS because the latter gives unbiased estimates of the standard deviation of the measurements only when the photometric errors are good, easily leading to spurious results otherwise. The PDM method was included due to its high sensitivity to highly non-sinusoidal variations with alternating peaks, as in the case of the light curves of eclipsing binary systems.

We proceeded with our analysis as follows:

- for all light curves that were preselected by the SLM method, we independently computed the periodogram using each of these methods and for each filter separately;
- after inverting the spectra from the PDM and SLM methods, the periodograms for each filter were then normalized by the maximum power;
- a single, normalized power spectrum was then obtained for each method by averaging over all filters the results obtained in the previous step for each filter independently;

Table 3. Number of variable star candidates selected by the different variability indices at different significance levels before and after (in parentheses) visual inspection.

Sign. level	$I_{\text{pfc}}^{(2)}$	$I_{\text{pfc}}^{(3)}$	$I_{\text{n}}^{(2)}$	$I_{\text{n}}^{(3)}$
0.5%	4598(219)	3574(192)	1292(242)	1045(242)
0.1%	1676(145)	1337(141)	466(186)	398(190)

- finally, we obtained a ranked list of the best periods for each method.

The best periods should in principle be those that ranked highest in all four methods. However, not all methods rank periods in the same way. Therefore, to choose of periods in a more objective fashion, we computed a “super rank” index as the sum of the ranks provided by each of these five methods. We then selected the ten best periods as those ten most highly ranked, according to this new index.

Finally, in order to select the very best period, we use the χ^2 test, in which Fourier coefficients (a_0, a_i, b_i) are derived that fit the data according to the following expression:

$$f(i) = a_0 + \sum_{i=1}^n \{a_i \sin [2k\pi\varphi(i)] + b_i \cos [2k\pi\varphi(i)]\}, \quad (17)$$

where $k = 1/P$. We used the Levenberg-Marquardt (Levenberg 1944; Marquardt 1963) method to find the solutions and employed the statistical F-test to evaluate the results. We limited the amplitude of the fit to the observed magnitude range ($m_{\text{max}} - m_{\text{min}}$) and kept the number of harmonics to $n \leq 20$. The period that produced the lowest reduced χ^2 value was selected as the true main period. Finally, we visually inspected the phase diagrams, which narrowed down our selection to our *final catalog* containing 275 periodic variable stars, as described in more detail in the next section.

5. Results and discussions

5.1. Catalog of periodic variable stars (C1)

Proceeding as described in the previous section, we have thus obtained our final sample of 275 clearly periodic variable stars (C1). Their coordinates, periods, mean magnitudes, and the number of epochs in each filter are listed in Table 4. We also provide preliminary classifications for most of the newly discovered variables. The variability types of these sources were assigned by visual inspection of their phase diagrams, and are also listed in Table 4, using the nomenclature of the General Catalog of Variable Stars³ (GCVS, Samus 1977). Phase-folded light curves for some of these objects are shown in Fig. 7.

5.2. Catalog of aperiodic variable stars (C2)

We also searched for those stars that, in spite of showing reasonably coherent light curves, do not show a clear main periodicity in the WFCAMCAL data, either because their variations are intrinsically aperiodic or because they have such long periods that these data were insufficient for deriving them. To identify such sources, we relied only on the variability indices, requiring that these have highly significant values, indicating the presence of

³ A full description of the nomenclature can be found at <http://www.sai.msu.ru/gcvs/gcvs/iii/vartype.txt>

Fig. 7. Phase-folded light curves of selected objects from our catalog of periodic variables (C1), showing data in all five broadband filters of WFCAM. ID's, types, and periods for each object are shown in the headers. The object WVSC-208 was only detected in the Z and Y bands.

Fig. 8. Distribution of I_{fi} versus I_{pfc} variability indices, for orders 2 (left) and 3 (right). The C1 and C2 sources are indicated by red and green circles, respectively.

correlated variations in the different WFCAM bandpasses. First, we selected sources by using a strong cutoff at the sampling- and magnitude-dependent 0.1% significance level, equivalently to the procedure described in Sect. 3.4. Table 3 shows the number of selected sources based on each of the cutoff surfaces shown in Fig. 5, before and after visual inspection. The number of sources selected by the $I_{fi}^{(s)}$ index is less than a third of the sources selected using the $I_{pfc}^{(s)}$ index, and at the same time, it includes 80% of all sources in C1 (Sect. 5.1). This high efficiency at a relatively low false alarm rate favored the application of the $I_{fi}^{(s)}$ index alone for selecting aperiodic variable candidates, which was followed by a visual inspection in order to reject likely false candidates.

Our procedure has led to an additional 44 sources, comprising our catalog of semi-regular or aperiodic variable stars and stars with uncertain periods (C2), shown in Table 5. The selected variable star candidates, including both periodic (C1) and non-periodic (C2) sources, are shown in the $[I_{fi}^{(s)}, I_{pfc}^{(s)}]$ plane (for $s = 2$ and 3) in Fig. 8. This figure clearly shows that the $I_{fi}^{(s)}$ index is significantly more powerful, as far as distinguishing the variability in the WFCAM data is concerned, compared with the $I_{pfc}^{(s)}$ index, since there is much less overlap between variable and non-variable sources in the former than the latter. As a word of caution, we note that the C2 catalog could still contain spurious sources, mainly due to sources that show correlated seasonal variations and/or correlated noise variations. Follow-up studies of these sources is thus strongly recommended, before conclusively establishing their variability status.

5.3. Cross-identifications

As the final step in our analysis, we performed a systematic cross-check of the sample of 319 sources in C1 and C2 catalogs to identify previously known sources and complement our catalog with data already in the literature. Among the cross-checked catalogs, one finds the SIMBAD database, the latest version of the General Catalog of Variable Stars (Samus et al. 2012), the AAVSO International Variable Star Index (VSX v1.1, now including 284 893 variable stars; Watson et al. 2014), the New Catalog of Suspected Variable Stars (Kazarovets et al. 1998),

and the Northern Sky Variability Survey (NSVS; Hoffman et al. 2009) catalog, among many other databases incorporated in the International Virtual Observatory Alliance (IVOA), using the Astrogrid facility⁴.

A delicate issue that we face when performing these extensive cross-checks between surveys with such diverse technical properties is their different astrometric accuracy. In this sense, we assumed a positional accuracy of 2'' in the sky coordinates for WFCAM, and then used optimized search radii according to the specific nature of each cross-matched database.

Taking the distribution of our sources across the sky into account, along with the specific nature of the observations that comprise the WFCAMCAL catalog, which were aimed at observing standard star fields and hence tended to avoid very crowded regions, it did not come as a surprise that the cross-checking with variability surveys of the southern sky and Galactic central regions, such as OGLE, MACHO, and ASAS, did not result in any positive match. We extended the search further to other astronomical catalogs of non stellar and/or extragalactic objects (e.g., planetary nebulae, quasars, optical counterparts of GRBs, just to mention a few) and in different spectral bands (from radio to X-rays), but again finding no superpositions with our WFCAM sources.

At the end of this search, we found a total of 44 stars that were already known from previous studies. Among them, 37 sources are included in the VSX catalog, three of which are also GCVS objects (i.e., AM Tau, EH Lyn, UV Vir, which are an Algol-type eclipsing binary, a contact binary, and an ab-type RR Lyrae, respectively). The GCVS also lists five other eclipsing binaries and another RRab Lyrae (HM Vir).

The outer part of the globular cluster M3 (NGC 5272) is also partly covered by the WFCAMCAL pointings. Indeed, among our preselected variable candidates, we were able to recover 11 stars that had already been identified as RR Lyrae stars by Cacciari et al. (2005), Jurcsik et al. (2012), and Watson et al. (2014). Cross-identifications and literature variability types of the previously known sources are given for C1 and C2 in Tables 4 and 5, respectively.

⁴ <http://www.astrogrid.org/>

Fig. 9. Color-color (*left panel*) and color-magnitude (*right panel*) diagrams of the analyzed sample. New objects in C1 (in blue) and C2 (in red) and previously known objects (in green) are shown with colored points. Red arrows indicate the reddening vectors.

5.4. Detection efficiency

To estimate the variable star detection efficiency of our method, we determined whether any variable stars in the WFCAM calibration fields that are known from other catalogs were not detected by us. Similar to the procedure described in Sect. 5.3, we performed positional cross-matches of our initial database of 216 722 light curves with all existing survey catalogs of variable stars incorporated in the IVOA. We found a total of 15 known variables (all of them periodic) that were missed by our search. Thirteen of them were excluded in the first broad selection phase (see Sect. 3) owing to the low values of their variability indices. The properties of these stars are listed in Table 6, and their WFCAM photometry is added to our catalog of periodic variable stars (C1), including a flag that refers to their non-detection as variables by our analysis.

Figure 10 shows four typical examples among the non-detected variables. The non-detection of these sources is either due to the insufficient phase coverage of their magnitude variations by WFCAM data (primarily in the cases of long-periodic eclipsing binaries with very low fractional transit lengths, see Fig. 10, upper panels) or due to saturation (Fig. 10, lower left panel) or to a very low signal-to-noise ratio.

Since the heterogeneity and small number of objects known from other variable star catalogs overlap with the WFCAMCAL fields make them insufficient for a quantitative assessment of our detection efficiency, we performed further tests using synthetic data. To do this we first built a database of noise-free LCs as being harmonic fits (see Eq. (17)) of actual C1 data, in each filter. Next, we generated 10^5 synthetic light curves, following the distributions of periods, amplitudes, and sampling of the final C1 catalog. Then, these synthetic light curves were added to segments of the real light curves of non-variable stars from the WFCAMCAL database. Finally, we applied the same procedures of variability search and period analysis that we discussed in Sects. 3 and 4 on the simulated data.

The result of our test is summarized by Fig. 11, which shows the detection efficiency $E_{\text{det.}} = N_{\text{det.}}/N_{\text{all}}$ as a function of K magnitude using bins of 0.25 mag, where $N_{\text{det.}}$ is the number of detected variables, and N_{all} is the number of all variables, respectively. Based on this test, we estimate an average detection efficiency of 93% over the complete magnitude range of the WFCAM database. Detection rates lower than 90% are only

Fig. 10. Example light curves of 4 previously known variable stars that were not detected in our analysis. ID's, variability types, and periods for each star are shown in the headers.

present in the two extremes of the magnitude range, dominated by saturation in the bright end and low signal-to-noise ratio in the faint end. We note that the lower overall detection efficiency suggested by the catalog cross-matches is due to the biased magnitude distribution of the cross-matched sources toward the bright end of the WFCAM magnitude range.

Fig. 11. Detection efficiency E_{det} vs. K mean magnitude for 10^5 synthetic variables.

5.5. Photometric properties of the variables

Figure 9 shows the variable and non-variable sources on the $(H-K)-(J-H)$ color-color and the $K-(J-K)$ color-magnitude planes. The variable sources cover the entire range of stellar parameter space in these planes. It is also clear from this figure that the area covered by the WFCAM Calibration fields have significant differential reddening, and a great fraction of the variables are reddened sources. This, together with the large aperture of the UKIRT telescope, explains the relatively low number of cross-identifications with previously known objects (Sect. 5.3): the WFCAMCAL database covers a range of faint NIR magnitudes where most of the optical variability surveys cannot penetrate, while most of the known variables in those catalogs are too bright for WFCAM.

We found 32 highly reddened ($Z - K > 3$) variable sources, 17 of which are periodic. We found that the positions of these red sources are strongly clustered around the positions $\alpha = 18^{\text{h}}29.5^{\text{m}}$, $\delta = +1^{\circ}36'$ (25 sources), and $\alpha = 7^{\text{h}}0^{\text{m}}$, $\delta = -4^{\circ}51'$ (8 sources). They are surrounded by a number of dark nebulae previously cataloged by Dobashi (2011) and Dutra & Bica (2002), respectively, suggesting that these sources might be embedded young stellar objects (see Tables 4, 5).

6. Conclusions

In this paper, we have established the WFCAM Variable Star Catalog, based on a detailed analysis of the WFCAMCAL NIR database. Our catalog contains 319 variable point sources, among which are 275 that are clearly periodic, and it includes 44 previously known objects. All catalog entries including multi-band light curve data are available online via the WFCAM Science Archive (WSA)⁵. Our approach to variability analysis included introducing a new, flux-independent variability index that is highly insensitive to the presence of outliers in the time-series data. A cross-matching procedure with previous variable star catalogs was also carried out, and the few sources with previous identification in the literature are noted. These catalogs represent one of the first such resource in the NIR, and thus an important first step toward the interpretation of future, more extensive NIR variability data sets, such as will be provided by the Vista Variables in the Vía Láctea (VVV) Survey in particular

(Catelan et al. 2013). A more detailed analysis of the different classes of variable stars detected in our catalog will be presented in a forthcoming paper.

Acknowledgements. Research activities of the Observational Astronomy Stellar Board of the Universidade Federal do Rio Grande do Norte are supported by continuous grants from the CNPq and FAPERN Brazilian agencies. We also acknowledge financial support from INCT INESpaco/CNPq/MCT. CEFL acknowledges a CAPES graduate fellowship and a CNPq/INESpaco postdoctoral fellowship. Support for C.E.F.L., M.C., and I.D. is provided by the Chilean Ministry for the Economy, Development, and Tourism's Programa Iniciativa Científica Milenio through grant IC 12009, awarded to The Millennium Institute of Astrophysics (MAS); by Proyectos FONDECYT Regulares #1110326 and 1141141; by Proyecto Basal PFB-06/2007; and by Proyecto Anillo de Investigación en Ciencia y Tecnología PIA CONICYT-ACT 86. RA is supported by the Gemini Observatory, which is operated by the Association of Universities for Research in Astronomy, Inc., on behalf of the international Gemini partnership of Argentina, Australia, Brazil, Canada, Chile, and the United States of America. Additional support by project VRI-PUC 25/2011 is also gratefully acknowledged. ICL acknowledges a CNPq/PNPd postdoctoral fellowship.

References

- Alcock, C., Allsman, R. A., Axelrod, T. S., et al. 1993, *Sky Surveys. Protostars to Protogalaxies*, 43, 291
- Arnaboldi, M., Neeser, M. J., Parker, L. C., et al. 2007, *The Messenger*, 127, 28
- Arnaboldi, M., Rejkuba, M., Retzlaff, J., et al. 2012, *The Messenger*, 149, 7
- Bakos, G., Noyes, R. W., Kovács, G., et al. 2004, *PASP*, 116, 266
- Blažko, S. 1907, *Astron. Nachr.*, 175, 325
- Blomme, J., Sarro, L. M., O'Donovan, F. T., et al. 2011, *MNRAS*, 418, 96
- Cacciari, C., Corwin, T. M., & Carney, B. W. 2005, *AJ*, 129, 267
- Casali, M., Adamson, A., Alves de Oliveira, C., et al. 2007, *A&A*, 467, 777
- Catelan, M., Minniti, D., Lucas, P. W., et al. 2013, in *40 Years of Variable Stars: A Celebration of Contributions by Horace A. Smith*, 139 [[arXiv:1310.1996](https://arxiv.org/abs/1310.1996)]
- Cincotta, P. M., Mendez, M., & Nunez, J. A. 1995, *ApJ*, 449, 231
- Clausen, J. V., Torres, G., Bruntt, H., et al. 2008, *A&A*, 487, 1095
- Cross, N. J. G., Collins, R. S., Hambly, N. C., et al. 2009, *MNRAS*, 399, 1730
- Dalton, G. B., Caldwell, M., Ward, A. K., et al. 2006, *Proc. SPIE*, 6269, 62690X-1
- Damerdj, Y., Klotz, A., & Boër, M. 2007, *AJ*, 133, 1470
- Deeming, T. J. 1975, *Ap&SS*, 36, 137
- Distefano, E., Lanzafame, A. C., Lanza, A. F., et al. 2012, *MNRAS*, 421, 2774
- Dobashi, K. 2011, *PASJ*, 63, 1
- Drake, A. J., Djorgovski, S. G., Mahabal, A., et al. 2009, *ApJ*, 696, 870
- Dutra, C. M., & Bica, E. 2002, *A&A*, 383, 631
- Dworetzky, M. M. 1983, *MNRAS*, 203, 917
- Emerson, J. P., Irwin, M. J., Lewis, J., et al. 2004, *Proc. SPIE*, 5493, 401
- Eyer, L. 2006, *Astrophysics of Variable Stars*, ASP Conf. Ser., 349, 15
- Eyer, L., & Bartholdi, P. 1999, *A&AS*, 135, 1
- Fruth, T., Kabath, P., Cabrera, J., et al. 2012, *AJ*, 143, 140
- Hambly, N. C., Collins, R. S., Cross, N. J. G., et al. 2008, *MNRAS*, 384, 637
- Handler, G. 2012, in *Planets, Stars and Stellar Systems* eds. T. D. Oswalt, & M. A. Barstow (Dordrecht: Springer Science+Business Media), Vol. 4
- Hewett, P. C., Warren, S. J., Leggett, S. K., & Hodgkin, S. T. 2006, *MNRAS*, 367, 454
- Hoffman, D. I., Harrison, T. E., & McNamara, B. J. 2009, *AJ*, 138, 466
- Hodgkin, S. T., Irwin, M. J., & Hewett, P. C. 2009, *MNRAS*, 394, 675
- Irwin, M. J., Lewis, J., Hodgkin, S., et al. 2004, *Proc. SPIE*, 5493, 411
- Jurcsik, J., Hajdu, G., Szeidl, B., et al. 2012, *MNRAS*, 419, 2173
- Kaiser, N., Aussel, H., Burke, B. E., et al. 2002, *Proc. SPIE*, 4836, 154
- Kazarovets, E. V., Samus, N. N., & Durevich, O. V. 1998, *IBVS*, 4655, 1
- Krabbendam, V. L., & Sweeney, D. 2010, *Proc. SPIE*, 7733
- Lafler, J., & Kinman, T. D. 1965, *ApJS*, 11, 216
- Law, N. M., Kulkarni, S., Ofek, E., et al. 2009, *BAAS*, 41, 418
- Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, *MNRAS*, 379, 1599
- Levenberg, K. 1944, *Quart. Appl. Math.*, 2, 164
- Lomb, N. R. 1976, *Ap&SS*, 39, 447
- Marquardt, D. W. 1963, *SIAM J. Appl. Math.*, 11, 431
- McCommas, L. P., Yoachim, P., Williams, B. F., et al. 2009, *AJ*, 137, 4707
- Minniti, D., Lucas, P. W., Emerson, J. P., et al. 2010, *New Astron.*, 15, 433

⁵ <http://surveys.roe.ac.uk/wsa/>

- Perryman, M. A. C. 2005, *Astrometry in the Age of the Next Generation of Large Telescopes*, 338, 3
- Pojmanski, G. 2002, *Acta Astron.*, 52, 397
- Pollacco, D. L., Skillen, I., Collier Cameron, A., et al. 2006, *PASP*, 118, 1407
- Riello, M., & Irwin, M. 2008, 2007 ESO Instrument Calibration Workshop, 581
- Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, *AJ*, 116, 1009
- Samus, N. N., Durlevich, O. V., & Kazarovets, R. V. 1997, *Balt. Astron.*, 6, 296
- Samus N. N., Durlevich O. V., Kazarovets, E. V., et al. 2012, *General Catalog of Variable Stars*, VizieR On-line Data Catalog: B/gcvs
- Scargle, J. D. 1982, *ApJ*, 263, 835
- Shin, M.-S., Yi, H., Kim, D.-W., Chang, S.-W., & Byun, Y.-I. 2012, *AJ*, 143, 65
- Siegel, M. H., & Majewski, S. R. 2000, *AJ*, 120, 284
- Simons, D. A., & Tokunaga, A. 2002, *PASP*, 114, 169
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, 131, 1163
- Stellingwerf, R. F. 1978, *ApJ*, 224, 953
- Stetson, P. B. 1981, *AJ*, 86, 1500
- Stetson, P. B. 1996, *PASP*, 108, 851
- Templeton, M. 2004, *JAASO*, 32, 41
- Tokunaga, A. T., Simons, D. A., & Vacca, W. D. 2002, *PASP*, 114, 180
- Tonry, J. L., Schmidt, B. P., Barris, B., et al. 2003, *ApJ*, 594, 1
- Udalski, A. 2003, *AcA*, 53, 291
- Walker, A. R. 2012, *Ap&SS*, 341, 43
- Watson, C., Henden, A. A., & Price, A. 2014, *VizieR Online Data Catalog*, 1, 2027
- Welch, D. L., & Stetson, P. B. 1993, *AJ*, 105, 1813
- Zechmeister, M., & Külrster, M. 2009, *A&A*, 496, 577