

## The *Hubble* Catalog of Variables

K. Sokolovsky<sup>1,2,3,\*</sup>, A. Bonanos<sup>1,\*\*</sup>, P. Gavras<sup>1</sup>, M. Yang<sup>1</sup>, D. Hatzidimitriou<sup>4,1</sup>, M. I. Moretti<sup>5,1</sup>, A. Karamelas<sup>1</sup>, I. Bellas-Velidis<sup>1</sup>, Z. Spetsieri<sup>1,4</sup>, E. Pouliaxis<sup>1,4</sup>, I. Georgantopoulos<sup>1</sup>, V. Charmandaris<sup>1</sup>, K. Tsinganos<sup>1</sup>, N. Laskaris<sup>6</sup>, G. Kakalettris<sup>6</sup>, A. Nota<sup>7,8</sup>, D. Lennon<sup>9</sup>, C. Arviset<sup>9</sup>, B. Whitmore<sup>7</sup>, T. Budavari<sup>10</sup>, R. Downes<sup>7</sup>, S. Lubow<sup>7</sup>, A. Rest<sup>7</sup>, L. Strolger<sup>7</sup>, and R. White<sup>7</sup>

<sup>1</sup>IAASARS, National Observatory of Athens, Vas. Pavlou & I. Metaxa, 15236 Penteli, Greece

<sup>2</sup>Sternberg Astronomical Institute, Moscow State University, Universitetskii pr. 13, 119992 Moscow, Russia

<sup>3</sup>Astro Space Center of Lebedev Physical Institute, Profsoyuznaya Str. 84/32, 117997 Moscow, Russia

<sup>4</sup>Department of Physics, National and Kapodistrian University of Athens, 15771 Ilissia, Greece

<sup>5</sup>INAF-Osservatorio Astronomico di Capodimonte, Salita Moiariello, 16, 80131 Napoli, Italy

<sup>6</sup>Athena Research and Innovation Center, Artemidos 6 & Epidavrou, 15125 Maroussi, Greece

<sup>7</sup>Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

<sup>8</sup>European Space Agency, Research and Scientific Support Department, Baltimore, MD 21218, USA

<sup>9</sup>European Space Astronomy Centre, Camino bajo del Castillo, Urbanizacion Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain

<sup>10</sup>The Johns Hopkins University, Baltimore, MD 21218, USA

**Abstract.** We aim to construct an exceptionally deep ( $V \lesssim 27$ ) catalog of variable objects in selected Galactic and extragalactic fields visited multiple times by the *Hubble Space Telescope* (HST). While HST observations of some of these fields were searched for specific types of variables before (most notably, the extragalactic Cepheids), we attempt a systematic study of the population of variable objects of all types at the magnitude range not easily accessible with ground-based telescopes. The variability timescales that can be probed range from hours to years depending on how often a particular field has been visited. For source extraction and cross-matching of sources between visits we rely on the Hubble Source Catalog which includes  $10^7$  objects detected with WFPC2, ACS, and WFC3 HST instruments. The lightcurves extracted from the HSC are corrected for systematic effects by applying local zero-point corrections and are screened for bad measurements. For each lightcurve we compute variability indices sensitive to a broad range of variability types. The indices characterize the overall lightcurve scatter and smoothness. Candidate variables are selected as having variability index values significantly higher than expected for objects of similar brightness in the given set of observations. The *Hubble* Catalog of Variables will be released in 2018.

## 1 Introduction

The *Hubble Space Telescope* (HST) is great for deep imaging thanks to the unique combination of low sky background, sharp point spread function (PSF) and wide field of view (compared to ground-

\*kirx@noa.gr

\*\*bonanos@astro.noa.gr

based adaptive optic systems; [13]). It is sensitive to ultraviolet light not accessible from the ground. Some fields were imaged by the HST multiple times opening a window to time-domain studies at faint magnitudes and high spatial resolution. High resolution imaging is important to overcome confusion in dense star fields, like the ones found in nearby galaxies and globular cluster cores. A number of dedicated variability studies were conducted with the HST including [1, 2, 5, 9, 10, 17, 25].

While the HST has superb internal astrometric precision, obtaining accurate absolute astrometry has been a challenge in the previous years. The positional information provided by the observatory's attitude control system is limited by the position accuracy of individual Guide Star Catalog ([14]) stars. The *Hubble* Source Catalog (HSC, [27]) solves this problem by cross-matching sources detected in individual HST visits ([3]) and matching the brightest detected sources to deep reference catalogs: PanSTARRS, SDSS and 2MASS. The Gaia catalog will be used as reference in future HSC versions.

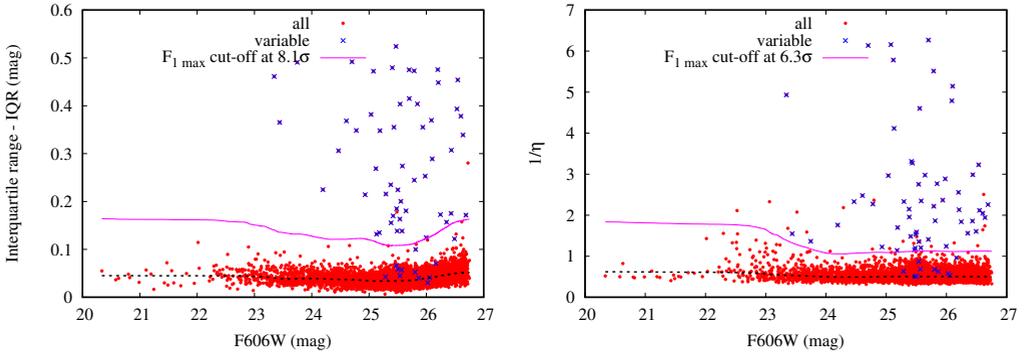
The HSC provides access to a uniform reduction of the majority of publicly available images obtained with the WFPC2, ACS/WFC, WFC3/UVIS, and WFC3/IR instruments. The HSC is very inhomogeneous: it includes 112 instrument-filter combinations; some filters are more popular than others. Most fields were observed only few times and the time between visits varies. The current HSC version 2 has 89 fields containing 1.7 million sources that were visited at least 5 times while only 7 fields containing 290000 sources visited  $\geq 25$  times. The HSC is based on visit-combined images that are deeper than individual exposures and mostly clean of cosmic rays. However, that comes at a price: information about changes on timescales shorter than one visit (that may last a few hours) is averaged-out and the resulting number of independent measurements is by a factor of a few smaller than the number of HST exposures of the field.

We aim to define a set of algorithms that will detect and validate candidate variable sources within the HSC, producing the *Hubble* Catalog of Variables (HCV). This is a work in progress that will lead to the release of the first version of the HCV in 2018. The initial project overview is given by [8].

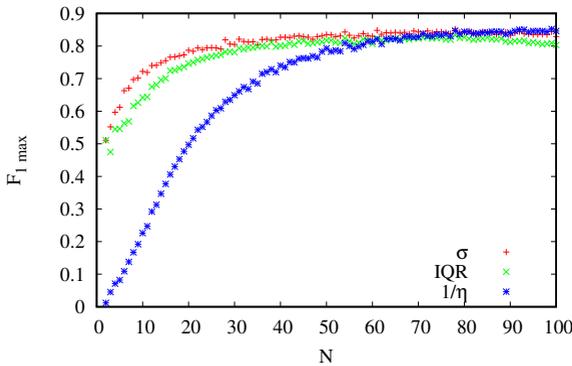
## 2 Variability search problem

It is common in optical photometry that measurement errors are not accurately known. The contribution of random background variations and photon noise to the brightness measurements uncertainty can be estimated easily for CCD observations. However, the hard-to-quantify residual systematic effects (pixel-to-pixel sensitivity variations, charge transfer inefficiency, blending with nearby stars) limit the photometric accuracy for brighter objects. Some measurements get corrupted by cosmic ray hits, CCD cosmetic defects, incorrect background estimation near frame edges and other undetected data processing anomalies. Since the majority of field stars are not variable at a few per cent level, we utilize them to estimate typical photometric accuracy in a given dataset as a function of magnitude.

Ground-based photometry was used by [22] to compare a number of variability indices that characterize “how variable” a lightcurve appears by quantifying its scatter and smoothness. We extend this work by comparing the variability indices listed in Table 1 to simulations based on HSC data. We inject variability with random amplitude into HSC lightcurves of non-variable objects using the technique described by [22]. The simulations confirm that the interquartile range (IQR) of the measured magnitudes,  $m_i$  together with the inverse von Neumann ratio that characterizes the lightcurve smoothness  $1/\eta = \frac{\sum_{i=1}^N (m_i - \bar{m})^2}{\sum_{i=1}^{N-1} (m_{i+1} - m_i)^2}$  (where  $\bar{m}$  is the mean magnitude and  $N$  is the number of measurements) can recover a broad range of variability patterns and are robust against individual outlier measurements. These indices do not depend on the estimated errors (Table 1) that may be unreliable. Figure 1 presents IQR and  $1/\eta$  indices as a function of magnitude in one of the simulations. Figure 2 illustrates how the variability detection efficiency of the indices changes as a function of the number of points in a lightcurve.



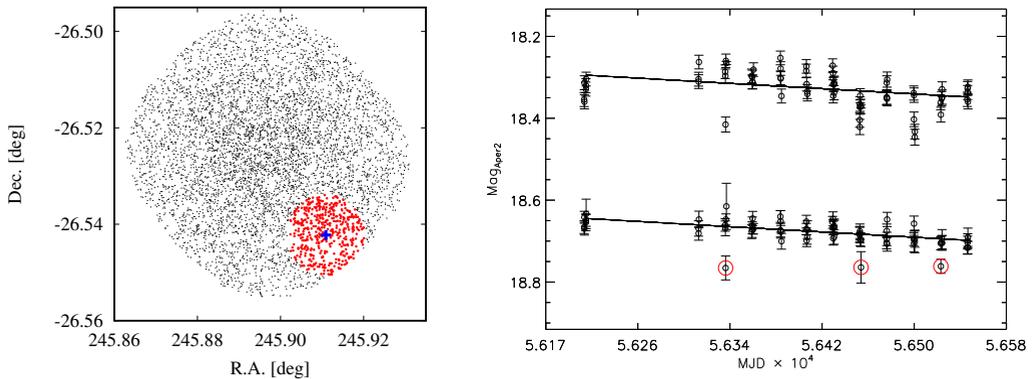
**Figure 1.** Variability indices characterizing scatter (left panel) and lightcurve smoothness (right panel) plotted as a function of magnitude. The indices are computed for HSC lightcurves of non-variable objects in the M31 halo field ([2]). Crosses mark the objects in which artificial non-periodic variability was injected as described in [22]. Selecting objects above the solid line as candidate variables results in the maximum  $F_1$ -score. Dotted line shows the median value of a variability index as a function of magnitude.



**Figure 2.** The efficiency of variable objects selection characterized by the maximum (over all possible cut-off thresholds)  $F_1$ -score as a function of the number of lightcurve points,  $N$ , for three variability indices:  $\sigma$ , IQR and  $1/\eta$  (see the references in Table 1). Based on simulated non-periodic variability injected in HSC lightcurves (M4 field, [17]). For the definition of  $F_1$ -score, see [https://en.wikipedia.org/wiki/F1\\_score](https://en.wikipedia.org/wiki/F1_score).

**Table 1.** Variability indices computed by the HCV pipeline.

Index	Errors	Ref.	Index	Errors	Ref.
<i>Indices quantifying lightcurve scatter</i>			time-weighted Stetson's $J_{\text{time}}$	✓	[7]
reduced $\chi^2$ statistic – $\chi_{\text{red}}^2$	✓	[4]	clipped Stetson's $J_{\text{clip}}$	✓	[22]
weighted std. deviation – $\sigma$	✓	[12]	Stetson's $L$ index	✓	[24]
median abs. deviation – MAD		[29]	time-weighted Stetson's $L_{\text{time}}$	✓	[7]
interquartile range – IQR		[22]	clipped Stetson's $L_{\text{clip}}$	✓	[22]
robust median stat. – RoMS	✓	[19]	consec. same-sign dev. – <i>Con.</i>		[28]
norm. excess variance – $\sigma_{\text{NXS}}^2$	✓	[16]	excursions – $E_x$	✓	[18]
norm. peak-to-peak amp. – $v$	✓	[21]	autocorrelation – $l_1$		[11]
<i>Indices quantifying lightcurve smoothness</i>			inv. von Neumann ratio – $1/\eta$		[20]
Welch-Stetson index – $I$	✓	[26]	excess Abbe value – $\mathcal{E}_{\mathcal{A}}$		[15]
Stetson's $J$ index	✓	[24]	$S_B$ statistic	✓	[6]



**Figure 3.** Lightcurve pre-processing. Left: spatial distribution of HSC objects in the M4 field ([17]). Highlighted are the objects used to determine local magnitude zero-point correction for the object marked with the cross. Right: lightcurve of that object before (top) and after (bottom) applying the local correction. The circles mark identified outlier points. The two lightcurves are shifted along the magnitude axis for clarity.

### 3 HCV data pre-processing and variability detection pipeline

The HSC data are imported and grouped according to the observed field. The measurements of objects near frame edges or obtained with an old version of the image processing pipeline are excluded as unreliable. For each object all its measurements in a given filter that pass the above selection are collected to construct a lightcurve. Each lightcurve is fitted with a straight line using robust regression and outliers from the fit are identified. If a high percentage of measurements obtained during some visit are identified as outliers in the corresponding lightcurves, all measurements associated with this visit are discarded (bad image). For each of the remaining visits, for each object a local zero-point correction is computed as the mean difference between the measured magnitudes and the ones predicted by the robust line fit for all objects within a specified radius around the object being corrected (Fig. 3). This should compensate for the residual large-scale sensitivity variations across the image.

A set of corrected HSC lightcurves of objects observed in a given field with the same instrument-filter combination is the basic unit of variability search. For each lightcurve in a set we compute variability indices and select as candidate variables the objects that have variability index values significantly higher than the typical value for their brightness. Magnitude-dependent cuts in robust indices IQR and  $1/\eta$  are currently used to select candidate variables (Fig. 1). We are investigating ways to efficiently combine information captured by all indices listed in Table 1 by means of the principal component analysis and machine learning.

The HCV pipeline is implemented in JAVA and parallelized using the APACHE SPARK framework. The variability indices implementation in the pipeline is consistent with their implementation in VAST ([23]) which is also used for lightcurve visualization and testing while a dedicated HCV visualization software is being developed.

### 4 Summary

- The HCV is a catalog of variable objects derived from the HSC. It will be released in 2018.
- The catalog is very heterogeneous due to the nature of the HSC dataset. It covers selected fields in the Galaxy, the Local Group and beyond.

- The HCV is very deep, it ventures into poorly explored region of variability parameter space.
- Data pre-processing and variability detection techniques used for HCV are applicable to other multi-epoch surveys.

*Acknowledgments:* This work is supported by ESA under contract No. 4000112940.

## References

- [1] Bernard, E. J., et al., *MNRAS*, **432**, 3047 (2013)
- [2] Brown, T. M., Ferguson, H. C., Smith, E., Kimble, R. A., Sweigart, A. V., Renzini A., & Rich, R. M., *AJ*, **127**, 2738 (2004)
- [3] Budavári, T., & Lubow, S. H., *ApJ*, **761**, 188 (2012)
- [4] de Diego, J. A., *AJ*, **139**, 1269 (2010)
- [5] Dolphin, A. E., et al., *ApJ*, **550**, 554 (2001)
- [6] Figuera Jaimes, R., Arellano Ferro, A., Bramich, D. M., Giridhar, S., & Kuppaswamy, K., *A&A*, **556**, A20 (2013)
- [7] Fruth, T., et al., *AJ*, **143**, 140 (2012)
- [8] Gavras, P., et al., preprint (arXiv:1703.00258) (2017)
- [9] Hoffmann, S. L., et al., *ApJ*, **830**, 10 (2016)
- [10] Jeffery, E. J., et al., *AJ*, **141**, 171 (2011)
- [11] Kim, D.-W., Protopapas, P., Alcock, C., Byun, Y.-I., & Khardon, R., *ASPC*, **442**, 447 (2011)
- [12] Kolesnikova, D. M., Sat, L. A., Sokolovsky, K. V., Antipin, S. V., & Samus, N. N., *AcA*, **58**, 279 (2008)
- [13] Lanzerotti, L. J., *Assessment of options for extending the life of the Hubble Space Telescope: final report* (National Academies Press, Washington D.C, 2005)
- [14] Lasker, B. M., et al., *AJ*, **136**, 735 (2008)
- [15] Mowlavi, N., *A&A*, **568**, A78 (2014)
- [16] Nandra, K., George, I. M., Mushotzky, R. F., Turner, T. J., & Yaqoob, T., *ApJ*, **476**, 70 (1997)
- [17] Nascimbeni, V., et al., *MNRAS*, **442**, 2381 (2014)
- [18] Parks, J. R., Plavchan, P., White, R. J., & Gee, A. H., *ApJS*, **211**, 3 (2014)
- [19] Rose, M. B., & Hintz, E. G., *AJ*, **134**, 2067 (2007)
- [20] Shin, M.-S., Sekora, M., & Byun Y.-I., *MNRAS*, **400**, 1897 (2009)
- [21] Sokolovsky, K. V., Kovalev, Y. Y., Kovalev, Y. A., Nizhelskiy, N. A., & Zhekanis G. V., *AN*, **330**, 199 (2009)
- [22] Sokolovsky, K. V., et al., *MNRAS*, **464**, 274 (2017)
- [23] Sokolovsky, K. V., & Lebedev, A. A., preprint (arXiv:1702.07715) (2017)
- [24] Stetson, P. B., *PASP*, **108**, 851 (1996)
- [25] Villforth, C., Koekemoer, A. M., & Grogin N. A., *ApJ*, **723**, 737 (2010)
- [26] Welch, D. L., & Stetson, P. B., *AJ*, **105**, 1813 (1993)
- [27] Whitmore, B. C., et al., *AJ*, **151**, 134 (2016)
- [28] Wozniak, P. R., *AcA*, **50**, 421 (2000)
- [29] Zhang, M., Bakos, G. Á., Penev, K., Csubry, Z., Hartman, J. D., Bhatti, W., & de Val-Borro, M., *PASP*, **128**, 035001 (2016)