

Searching for Periodic Variables in the EROS-2 Database

P. Dubath^{1,2}, I. Lecoœur^{1,2}, L. Rimoldini^{1,2}, M. Süveges^{1,2},
J. Blomme³, M. López⁴, L. M. Sarro⁵, J. De Ridder³, J. Cuypers⁶,
L. Guy^{1,2}, K. Nienartowicz^{1,2}, A. Jan^{1,2}, M. Beck^{1,2}, N. Mowlavi^{1,2},
D. Ordóñez-Blanco^{1,2}, J. B. Marquette⁷, J. P. Beaulieu⁷,
P. Tisserand^{8,9}, É. Lesquoy^{7,9} and L. Eyser¹

¹Observatoire astronomique de l'Université de Genève, Switzerland
email: Pierre.Dubath@unige.ch

²ISDC Data Centre for Astrophysics, Université de Genève, Switzerland

³Instituut voor Sterrenkunde, K.U. Leuven, Belgium

⁴Centro de Astrobiología (INTA-CSIC), Villanueva de la Canada, Spain

⁵Dpt. de Inteligencia Artificial, UNED, Madrid, Spain

⁶Royal Observatory of Belgium, Brussels, Belgium

⁷Institut d'Astrophysique de Paris, 75014 Paris, France

⁸Research School of Astronomy & Astrophysics, Mount Stromlo Observatory, Weston ACT 2611, Australia

⁹CEA, DSM, DAPNIA, Centre d'Études de Saclay, 91191 Gif-sur-Yvette Cedex, France

Abstract. We started a systematic search for periodic variable-star candidates in the EROS-2 database in the context of preparatory work for the GAIA satellite mission. The goal is to evaluate different classification tools and strategies, and to identify a large sample of variable candidates. In this paper we present the results of an assessment study of a three-step identification and classification process. In the study we took a sample of about 80,000 stars from one of the LMC EROS fields.

Keywords. methods: statistical – techniques: photometric – stars: variables: general.

The identification and classification of the periodic variable candidates are achieved through three steps. First, a number of variability criteria are computed to identify variable candidates. Secondly, a supervised classifier is used to isolate periodic variables. Thirdly, another supervised method is employed to classify the periodic stars.

1. Variability Detection

A number of variability criteria are computed from the light curves, and p -values are derived for each criterion. A star is considered to be a variable candidate if the p -value is smaller than a specified threshold.

Fig. 1 shows the number of selected stars as a function of the p -value threshold for 2 of our 6 different criteria. The triangles give the fraction of selected stars which are also identified as variables in the EROS-2 catalogue (Tisserand *et al.* 2007); their total number is indicated by the horizontal line. The vertical dashed lines give the p -value thresholds used in this study. A sample of 11,888 variable candidates was extracted from the total sample of about 80,000 stars.

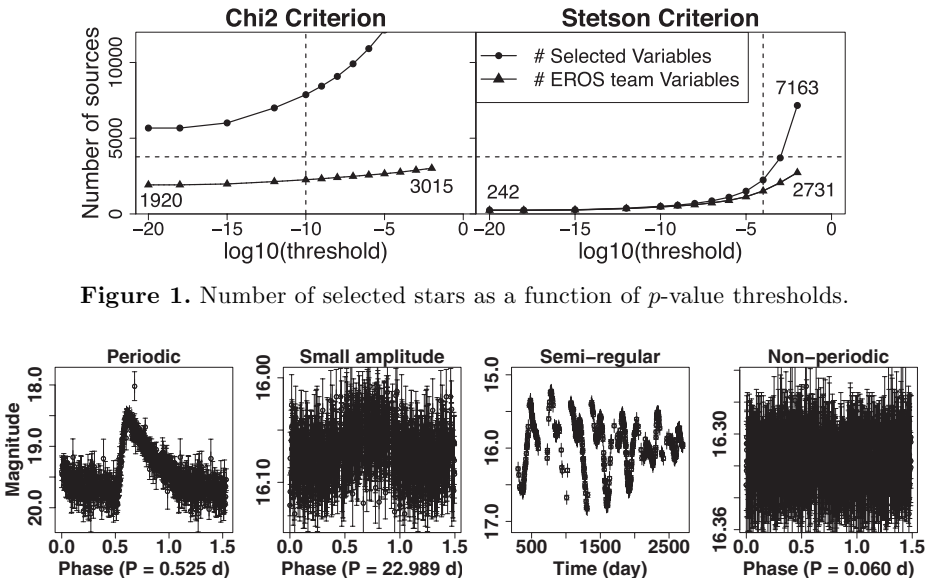


Figure 1. Number of selected stars as a function of p -value thresholds.

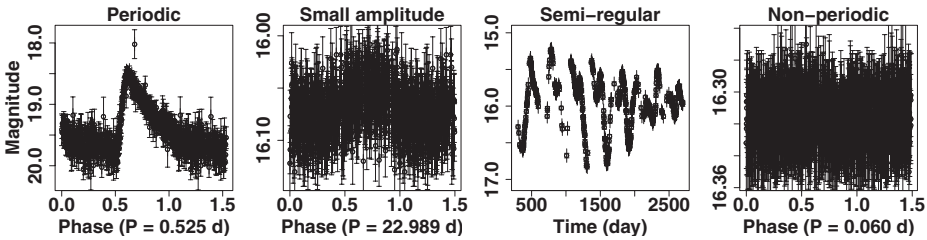


Figure 2. Examples from the four ad-hoc categories used in the visual classification.

| $\frac{P}{S}$ | $\frac{S}{P}$ | $\frac{P}{S}$ | $\frac{S}{P}$ | |
|---------------|---------------|---------------|---------------|--------------------------------|
| 1867 | 13 | 1 | | NP (Non-Periodic stars) |
| 117 | 54 | 9 | | SA (Small Amplitude variables) |
| | 12 | 87 | 1 | PV (Periodic Variables) |
| | 3 | 2 | 15 | SR (Semi-regular variables) |

Figure 3. Confusion matrix from random forest classification.

The harmonic least-squares analysis method (Zechmeister & Kürster 2009) was used to search for frequencies in the sample of 11,888 variable candidates. Stars with frequencies close to known spurious frequencies (near zero and 1, 2, 3, 4, and 5 cycles per day) were removed, and a sub-sample of 2181 light curves folded using the main period was inspected and visually classified into the four ad-hoc categories illustrated in Fig. 2.

2. Periodicity Detection Training

We followed the classification procedures used in a previous paper (Dubath *et al.* 2011) to identify periodic stars. A supervised random forest classifier (Breiman 2001) was trained on the sub-sample of 2181 visually classified stars by using a large number of attributes.

Fig. 3 shows that the relatively small sample of periodic variables in our data can be identified with good efficiency. The main attributes used in the classification process, ranked according to the importance provided by the random forest methodology, include (1) measures of the model amplitude-to-noise ratio in the B and R bands, (2) the period search false-alarm-probability, (3) the B and R magnitude ranges, (4) the weighted skewness of the magnitude distribution, (5) the point-to-point scatter, (6) the Stetson (1996) variability criterion and (7) the relative difference between the frequencies derived from the B and R bands.

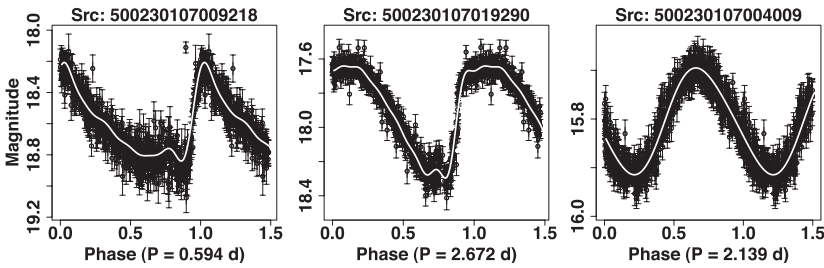


Figure 4. Light curves of three periodic variables.

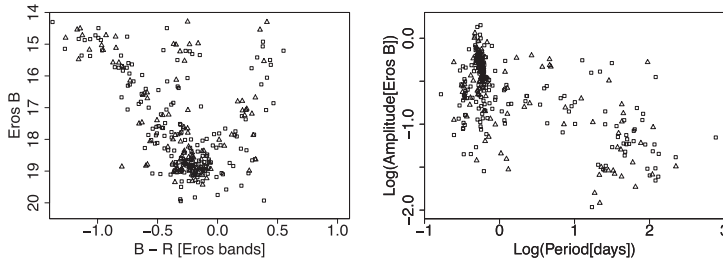


Figure 5. Colour-Magnitude and Period-Luminosity diagrams. Triangles and squares denote the training and test sets, respectively.

The classification model trained with 2181 stars was used to predict types for the sample of 3750 variable candidates with frequencies outside the spurious ranges. That yielded a total sample of 292 periodic variables (100 and 192, from the training and the test sets, respectively). Fig. 4 shows some examples of folded light curves obtained for these sources.

When a similar efficiency for the other EROS fields is assumed, the total number of periodic variables that could be extracted is of the order of 200,000.

3. Classification Plans

The next step in our study is to identify good representatives of the main types of periodic variables included in our sample; they will then be used to train a periodic variable classifier. The structures seen in the colour-magnitude and period-amplitude diagrams displayed in Fig. 5 give us good confidence that it will be possible to classify our set of periodic stars into the main variability types.

References

- Breiman, L. 2001, *Machine Learning*, 45(1), 5
- Dubath, P., *et al.* 2011, *MNRAS*, 414, 2602
- Stetson, P. B. 1996, *PASP*, 108, 851
- Tisserand P., *et al.* 2007, *A&A*, 469, 387
- Zechmeister, M., Kürster M. 2009, *A&A*, 496, 577