

ON A NEW TECHNIQUE FOR DISCOVERING VARIABLE STARS

A. V. Mironov,¹ A. I. Zakharov¹ and F. N. Nikolaev²

¹ *Sternberg State Astronomical Institute, Universitetskij prospekt 13,
Moscow 119992, Russia*

² *Moscow Lomonosov University, Physical Department, Vorob'evy Gory
1, Moscow 119992, Russia*

Received November 5, 2003

Abstract. A technique for discovering variable stars based on the calculation of the correlation coefficients is proposed. Applications of the technique are shown on the results of numerical experiments and on the *Hipparcos* photometric data.

Key words: techniques: photometric – stars: variables: general – orbiting observatories: Hipparcos

1. INTRODUCTION

This paper is devoted to the discovery of variable stars using the results from the space project *Hipparcos* (Hipparcos 1997) and similar data. Until recent times, variable stars were discovered, as a rule, using a comparison in between of a large number of pairs of photo plates. Some variables were discovered just by chance after accumulating several accurate photometric measurements for a certain star.

The General Catalogue of Variable Stars (GCVS) in its current form contains information on nearly 37 500 Galactic variable stars. Among them only about 8 000 stars have magnitudes in maximum light brighter than 10.0. About 40% of these bright variables have been included in the GCVS after analyzing the results of the *Hipparcos* project. Future space projects are planned for coming years including such an immense project as *Gaia* (Gaia 2000). The project *Lyra* (Lyra 1999) is scheduled for carrying out in the Russian segment

of the International Space Station. It is intended to do ten-color photometry of stars down to 15th magnitude to 1% accuracy. All these projects will provide astrometric and photometric data on hundreds of millions of objects which are sure to comprise millions (or even tens of millions) variable stars. Hence, we seem to be standing on the threshold of a new epoch in the research of variable stars and the application of the results of their investigation in studies of the structure and evolution of stellar systems.

Traditional methods to detect variability and to check the suspected variability are sure to fail when dealing with the huge volume of information. There is no physical possibility for a team of astronomers to scan all light curves with their own eyes (as they do now). Thus it is right time to develop another approach for the automatic analysis of the observational data.

There exists another aspect of the problem. As a rule, we do not have a sufficient database to investigate the distribution of variables in the Galaxy by statistical techniques. In any region of the sky, one cannot be sure that all variables in it have been already discovered. Moreover, one cannot even determine the fraction of variables already known.

An analysis of digital experiments and of the *Hipparcos* and *Tycho* photometric data must make us ready for reductions of new results of space astrometric and photometric surveys.

2. Z-STATISTICS USED IN THE HIPPARCOS DATA REDUCTION.

In the *Hipparcos* “Epoch Photometry Annex”, for every transit (having number j) of every star (having number i), its individual magnitude Hp_{ij} as well as the standard error σ_{ij} estimated during the transit are listed.

To discover variable stars, the authors of the *Hipparcos* catalogue used the following statistics:

$$\chi_i^2 = \sum_{j=1}^{N_i} \frac{(Hp_{ij} - \overline{Hp_i})^2}{\sigma_{ij}^2}. \quad (1)$$

The parameter χ_i^2 has a χ^2 -distribution with $\nu_i = N_i - 1$ degrees of freedom. Here i is a star number; j is a transit number; N_i is the total number of the star transits; Hp_{ij} is an individual estimate of the star magnitude in the j th transit; $\overline{Hp_i} = \frac{1}{N_i} \sum_{j=1}^{N_i} Hp_{ij}$ is the

mean magnitude of the i th star. Then a normalized χ^2 (see Kendall & Stuart 1977) was calculated as

$$\mathcal{Z}(\chi_i^2) = \left(\frac{9\nu}{2}\right)^{\frac{1}{2}} \left[\sqrt[3]{\frac{\chi_i^2}{\nu}} + \frac{2}{9\nu} - 1 \right]. \quad (2)$$

The \mathcal{Z} -statistics has an asymptotically normal distribution with zero mean and unit variance. $\mathcal{Z}(\chi_i^2)$ was the basic variability indicator in the *Hipparcos* analysis.

The \mathcal{Z} -statistics has at least one noticeable defect. To calculate it, one has to know the variance σ_{ij}^2 *a priori*. This variance should be derived from data obtained for the stars which are apparently non-variable. But in the *Gaia* mission (as well as any other measurements with a CCD matrix), the variance σ^2 will not be determined. Therefore we need a variability indicator which does not require any preliminary information.

3. STATISTICS BASED ON CORRELATION COEFFICIENTS

There is a natural way to solve the problem. The modern astrophotometry usually uses multicolor photometric systems. Quite often, multichannel detectors are applied for this purpose and an individual registration channel is used for every passband. In such a case, radiation is recorded in different channels simultaneously or quasi-simultaneously. It is obvious that if we measure a variable flux, then a correlation between signals in different channels must exist.

Let us have a multicolor photometric system consisting of n_b passbands, so that radiation from a star numbered i is measured simultaneously in all these passbands. Let us suppose that we have made N_i measurements in every band and obtained n_b sets of magnitudes m_{ijk} . ($j = 1, 2, \dots, N_i$; $k = 1, 2, \dots, N_b$). For any pair of these sets (numbered k and l , $l = k + 1, \dots, N_b$) one can calculate a linear correlation coefficient

$$\rho_{ikl} = \frac{\text{COV}_{ikl}(m_{ik}, m_{il})}{\sigma_{ik}\sigma_{il}}, \quad (3)$$

where

$$\begin{aligned} \text{COV}_{ikl}(m_{ik}, m_{il}) &= \sum_{j=1}^{N_i} \frac{m_{ijk}m_{ijl}}{N_i} - \bar{m}_{ik}\bar{m}_{il}; \\ \sigma_{ik}^2 &= \frac{1}{N_i} \sum_{j=1}^{N_i} (m_{ijk} - \bar{m}_{ik})^2; \quad \bar{m}_{ik}^2 = \frac{1}{N_i} \sum_{j=1}^{N_i} m_{ijk}^2. \end{aligned} \quad (4)$$

As magnitudes of different stars are measured unequal number of times, the ratio of the correlation coefficient to its error should be used. An estimate of the correlation coefficient error is:

$$\sigma(\rho_{ikl}) = \frac{1 - \rho_{ikl}^2}{\sqrt{N_i - 2}}. \quad (5)$$

To search for the variability of a star numbered i , we suggest that the statistics \mathcal{MZ}_{ikl}

$$\mathcal{MZ}_{ikl} = \rho_{ikl} / \sigma_{\rho_{ikl}} \quad (6)$$

should be used. We have designated it \mathcal{MZ} (i.e., \mathcal{M} ulti- \mathcal{Z}) in contrast to the single-channel \mathcal{Z} -statistics applied by the authors of *Hipparcos*.

Since there exist n_b channels, one can construct a vector \mathcal{MZ}_{ikl} with $n_p = \mathcal{C}_{n_b}^2 = n_b(n_b - 1)/2$ components, where $\mathcal{C}_{n_b}^2$ is the number of combinations of size 2 from a set of size n_b . From components of the vector, one can calculate an integrated statistics \mathcal{MZ} for every star

$$\mathcal{MZ}_i = \frac{1}{\sqrt{n_p}} \sum_{k=1}^{n_b} \sum_{l=k+1}^{n_b} \mathcal{MZ}_{kl} \quad (7)$$

reflecting all the channels being generally correlated. It is easy to see that the integrated \mathcal{MZ} is a projection of n_p -dimensional vector \mathcal{MZ}_{ikl} to n_p -dimensional cube diagonal passing through the point of origin. We shall call this diagonal the main diagonal.

4. EFFICIENCY OF THE TECHNIQUE

If we examine non-variable stars only and there are no correlations between signals in different channels then the points corresponding to the ends of the \mathcal{MZ} -vectors (\mathcal{MZ} -points) will have approximately normal distribution with zero mean and unit variance without any dependencies on the signal level, the signal variation and the total number of channels. For variable stars, the \mathcal{MZ} -points and the center of their distribution will move along the main diagonal in the positive direction. So, if \mathcal{MZ}_i is higher than +1.96 the star is not constant with a probability of 0.975; if $\mathcal{MZ}_i > 2.58$ – with a probability of 0.995; and if $\mathcal{MZ}_i > 3.29$ – with a probability of 0.9995.

As an illustration, consider Figure 1. In it the results of a numerical experiment are shown. A model with eight channels was examined. The standard deviation of signal (standard error: $SE = \sigma_{const}$)

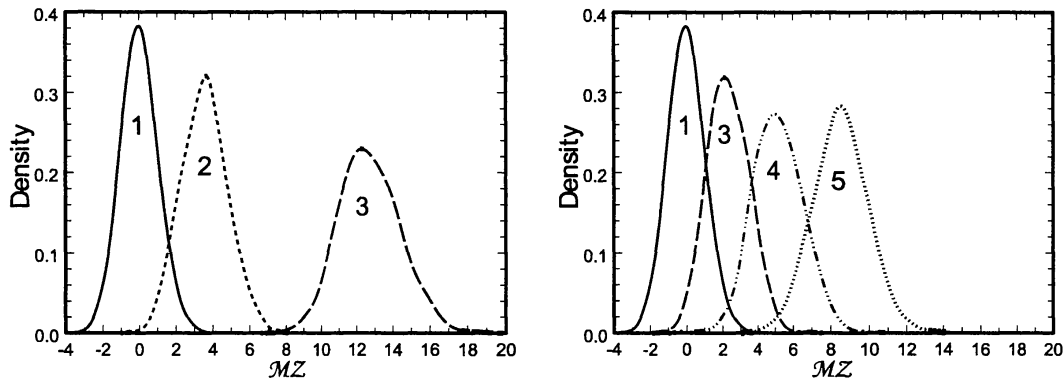


Fig. 1. The distribution of the MZ -statistics model values. Left panel: variables with sinusoidal brightness variations; right panel: variables with brightness variations like classical cepheids; (1) – constant stars; (2) $A_{\text{var}}/SE = 0.3$; (3) $A_{\text{var}}/SE = 0.6$; (4) $A_{\text{var}}/SE = 0.9$; (5) $A_{\text{var}}/SE = 1.2$.

was about 0.1 mag. The curve (1) corresponds to the constant star models. Other curves show variable star models with sinusoidal light curves and with light curves like Delta Cephei stars. Variability amplitude A_{var} varies from 0.03 mag up to 0.15 mag. It is seen that for the sinusoidal light curves one can discover variables if the ratio A_{var}/SE is equal to 0.3 (noise is 3 times stronger than the variability amplitude!) with the probability about 0.90, and if the ratio A_{var}/SE is equal to 0.6, with the probability almost equal to unity. For the cepheid-like light curves one can discover variables if the ratio A_{var}/SE is equal to 0.6 with the probability about 0.60, and if the ratio is 1.2, with the probability exceeding 0.99.

Besides the numerical experiment we have applied the technique to data in the *Hipparcos* and *Tycho* “Epoch Photometry Annex” to try to rediscover the known variables of different types included in the GCVS. The three channels H_p , B_T and V_T were considered. The results are shown in Table 1.

5. CONCLUSIONS

We consider the technique to be effective to discover variables and to estimate fractions of found and missed objects. It is appropriate to mention here that individual measurements of V_T and B_T magnitudes have large errors (≈ 0.1 mag) which make searching for variables more difficult. In spite of that, the results in Table 1 appear satisfactory. We recommend to use the technique in searching for new variables on the base of multichannel photometric observa-

Table 1. The number of the known variables of some types rediscovered with the suggested technique.

GCVS Type of variability	Total known variables	Rediscovered with a probability		
		of 0.975 (%)	of 0.995 (%)	of 0.9995 (%)
ACV	205	66.8	59.0	49.2
BCEP	79	83.6	74.7	69.7
DCEP	189	98.4	98.4	96.3
DSCTC	130	56.9	49.2	42.3
EA	316	86.1	83.5	78.8
EB	105	97.1	97.1	96.2
EW	74	99.9	98.7	95.6
GCAS	92	95.6	93.5	93.5
M	129	97.7	97.7	97.7
RRAB	56	96.4	96.4	94.6
RRC	19	94.7	89.5	89.5
SRA	53	98.1	96.2	96.2
SRB	312	97.4	97.1	96.5
All types	2867	88.3	85.0	80.8

tions.

The technique does not use any information on the observation moments and do not depend on a value of the variation period. Only in the case when all observations are done in the same phase, the variability cannot be detected. In all other cases the star will be recognized as variable if there is a sufficient correlation of its variations measured in different channels.

This study was supported by the Russian Foundation for Basic Research (Grant No. 02-02-16069).

REFERENCES

- Gaia 2000, *ESA: 2000, Gaia: Composition, Formation and Evolution of the Galaxy*, Technical Report, ESA-SCI (2000)4
- Hipparcos 1997, *ESA 1997, The Hipparcos and Tycho Catalogues*, ESA SP-1200
- Kendall M., Stuart A. 1977, in *The Advanced Theory of Statistics*, Vol.1, Chapter 16, Charles Griffin, London
- Lyra 1999, *The Long-term Program of Applied Science Researches and Experiments Planned for Implementation on the Russian Segment of the International Space Station, Technical Task on the Space Experiment "Multicolor Photometric Sky Survey"*