Automated data analysis – algorithms for Gaia Cepheids

Zoltán Tamás Kiss, László Szabados
Konkoly Observatory of the Hungarian Academy of Sciences,
Budapest 1525, P.O. Box 67, Hungary
E-mail: kissz@konkoly.hu

Abstract. Gaia space astrometric mission, to be launched in 2012, is expected to provide a bulk of data observing a thousand million objects within its five year time-span. Due to its scanning law the all sky will be covered, and all the available objects will be observed repeatedly with a moderate time sampling, therefore Gaia will obtain a large amount of photometric data that carry valuable information on a wide range of variability phenomena.

The Gaia group of the Konkoly Observatory works on the preparation of automated characterisation algorithms of pre-classified Cepheid type sources within the Variability Processing group of Gaia.

Gaia is expected to provide observational data on all available objects with a number of data points gradually increasing during the mission. Based on the scanning law a moderate number of data points per source are expected. That, though makes the processing necessarily to be fine tuned, is sufficient for the characterisation including determination of period(s), Fourier parameters, possible period changes, and hints on binarity. Therefore we use for testing public and unpublished data sources that are comparable to those expected from Gaia in term of sampling combined with simulated Gaia measurements. Depending on the phase coverage ground based supplementary observations may be needed to improve the accuracy of the results.

The final product of this work will be a contribution to the Gaia database, that will be compiled primarily based on the Gaia on-board measurements, and probably include a number of newly discovered Cepheids. This work may lead to an improved knowledge of the nature of these objects, that are fundamental to approach a more precise cosmic distance scale.

1. Introduction

1.1. Gaia overview
Gaia is designed to continue and extend the efforts made by the Hipparcos, the former astrometric space mission of the ESA, to draw the three dimensional picture of our astronomical environment. Gaia measures all objects detectable down to 20th mag with unprecedented accuracy, therefore the expected number of objects to be detected is around $10^9$. The expected photometric and astrometric uncertainty is 0.05 mag at 20th mag and 24 μas at 15th mag. However this is not the only substantial improvement with respect to the Hipparcos. To determine all the kinematical properties of the observed objects, not just the accurate spatial location but also the complete velocity vector, Gaia measures radial velocities of the target objects performing spectroscopic measurements. Measurements are repeated several times for all objects during the mission in order to reveal possible temporal changes in the observable properties of astronomical objects. Scheduled for launch in 2012, Gaia will perform its...
measurements in the L2 Lagrange point of the Sun-Earth system rotating with a period of 6 hours around its spin axis that points 50 degrees from the direction of the Sun with a precession period of 70 days. The optical system has two distinct lines of sights with 106 degrees separation, both are perpendicular to the rotation axis. Due to this geometry and rotation each object crosses its fields of view 70 times on average during the 5 year expected lifetime of the Gaia. This number of observations per object allows us to study the variability of the objects detected. Within the frame of the Coordination Unit 7 (CU7) of the Gaia Data Processing and Analysis Consortium (DPAC) the Gaia group of the Konkoly Observatory is responsible for the development and testing of software related to data on Cepheid variables to be observed by Gaia.

1.2. Gaia Cepheid observations

Due to the fact that all detected objects will be measured several dozen times, Gaia allows us to investigate the brightness variations of the objects observed. Though the expected number of data points per objects is relatively small, 70 on average, in combination with the unprecedented photometric accuracy and number of detected objects it is still enough for accurate characterisation of light variations. The applied scanning law provides appropriate time sampling, leading to a good phase coverage, with the exception of those objects that have periods nearly in resonance with the Gaia’s rotational or precession period.

The procedures being developed by the Konkoly Gaia group are meant to be used within the Gaia data processing pipeline. The input data are the pre-processed observations organised into photometric time series. During the pre-processing a period search and a classification are to be done. Our procedures work only with data of objects that were found to be variable and classified as Cepheids based on the data available in the earlier stages of the pipeline by procedures developed by the DPAC Coordination Units 1-6. Though the number of objects to be processed is not so huge as the total number of objects detected by Gaia, because of the possible misclassification, still much larger than the expected final number of Cepheids to be found by Gaia. The main issue, however, is the diversity of the possible input data. The resulting accuracy may be sensitive to the time sampling and periods of light variations as well as the possible variations in both period and light curve shape that are often unpredictable even for the known variables. The results to be obtained using algorithms that have to be applied as part of a pipeline are quite sensitive to the way how these effects are taken into account.

We perform the coding of the developed algorithms to Java language to make the source code fit to the Gaia framework, and make it applicable for automated operation in processing the future Gaia data, and implement the results in the Gaia catalogue.

2. Methods and input data

2.1. Methods

The procedures being developed are meant to determine the period or periods, calculate the coefficients of the Fourier decomposition of the light curves for the significant harmonics in all available colours. At first a Fourier decomposition is performed to find out the rough values of periods. In case of a single period it is refined applying a least squares fit to the light curve using a function in the following form:

\[ m(t) = C + A_0 \sin(2\pi(f t + \Phi_0)), \]  

where \( t \) is the time in days with respect to an arbitrary initial epoch; \( m \) is the magnitude for the given \( t \); \( C \) is the mean magnitude; \( A_0 \) is the amplitude for the main frequency \( f \); and \( \Phi_0 \) is the initial phase at \( t = 0 \). Phase values are ranging from 0 to 1 according to the formula. This procedure is to be applied for light curves of the given object measured in all available colours. The final refinement of the period value is done calculating the mean of the values obtained in different colours inversely weighted with their uncertainties.
Using the refined value of period obtained, the parameters of the harmonics are searched for in the following form:

\[ m(\phi) = C + A_0 \sin(2\pi(\phi + \Phi_0)) + A_1 \sin(2\pi(2\phi + \Phi_1)) + \ldots + A_n \sin(2\pi(n\phi + \Phi_n)), \]

(2)

where \( \phi \) is the running phase, with respect to the phase at an arbitrary initial epoch, obtained from the moments of individual observations and the fitted period; \( m \) is the magnitude at the given \( \phi \); \( C \) is the mean magnitude; \( A_0, A_1, \) and \( A_n \) are the amplitudes and \( \Phi_0, \Phi_1, \) and \( \Phi_n \) are the initial phases of the main frequency, the 1st, and the \( n \)th harmonics. The values of these parameters are obtained applying a least squares fit to the observed data. The Fourier coefficients are calculated as follows: \( R_{ij} = A_i/A_j \) and \( \Phi_{ij} = \Phi_i - \Phi_j \). These are meant to be used for the determination of the subclass of the Cepheid and its pulsation mode. The binarity status is determined from the relative amplitudes observed in different colours (Klagyivik & Szabados 2009).

In case of sufficient data points and phase coverage, possible changes in the period and light curve shape are also examined. For the expected number of data points for Gaia time series is usually low, about 70 or even smaller, period changes or even changes in the light curve shape can be pointed out by splitting time series into separate cycles. The data points of the different cycles are compared to each other. To describe period changes the least squares fit to the time series is determined using a period changing linearly with time. This fitting is quite sensitive to the initial values of the parameters, especially the rate of period change. These values are estimated determining the periods of the splitted parts of time series separately. A more complex picture of the period changes can be achieved by generating \( O-C \) diagrams of the given light curve data. This method, however, needs longer time series to reach meaningful results.

2.2. Input data for testing

While the algorithms being developed are meant to be used for the processing of future Gaia data, testing of these procedures can only be performed on simulated Gaia data and archive datasets. Appropriate test data must be similar to the expected Gaia data at least in some particular aspects. For this reason the applicability of existing data archives is quite limited. Time series with the same level of photometric uncertainties can only be obtained from simulating Gaia measurements. However, there are archival datasets with time series similar in extent, time sampling, and number of data points. Though the uncertainty of these data are notably higher, this affects the tests only in the sense that errors of the test results will be much higher than expected for Gaia results. Accordingly, while the accuracy of the Gaia results can be estimated only from simulated Gaia data, the detectability of some particular phenomena constrained by mostly the time sampling, length and number of elements of the dataset can be confirmed by testing with archival data. For this purpose Berdnikov (2008) and OGLE III (Soszynski et al. 2008) data are used. The automated procedures were successfully tested on more than 3000 time series in total.

3. Preliminary results

Fourier spectra were generated for all the time series used for testing, and the least squares fit mostly resulted in an accurate period value for the given epoch. In some cases more than one period was found according to the different modes or the different periods of the components of a binary Cepheid. The obtained Fourier spectra showed significant peaks corresponding to the previously published period values. The determination and refinement of the period resulted in an accuracy between \( 10^{-5}, 10^{-8} \) d\(^{-1} \) depending on the number of elements and length of the time series, and the accuracy of the observed data. Fourier parameters were also determined. The number of significant harmonics is between 3 and 12 depending on the phase coverage. The
The typical value of the residual of light curve fitting was around 0.01 mag. In some cases period change could also be recognised splitting the data into separate cycles, though mostly for time series with larger numbers of elements as it was expected from the Gaia observations. In case of changing period there could not be found a constant period for which data points of all the different cycles overlap each other. For some of these, use of period changing linearly with time resulted in a much better fit to the observed data, even when the length of the time series was not longer than the expected lifetime of Gaia. For some others, mostly time series with length of a decade or longer, no reasonably good fit could be found even using linearly changing period. For such cases generating $O-C$ diagrams showed period changes more complicated than linear. However, significant results could only be obtained for time series with a number of elements much larger than expected from Gaia.

Splitting time series into separate parts can reveal changes in the shape of the light curves, however, because of insufficient data points even the presence of this phenomenon is doubtful. A poor fit with constant period may be caused by the presence of two or more periods in the pulsation. Fourier decomposition is still useful in this case to estimate the rough values of the periods excited simultaneously.

### 4. Summary

This work concentrates on the development of algorithms for Gaia data processing, the testing of the algorithms developed, and its probable results. Algorithms that can be used to automatically determine and refine periods as well as period changes of Cepheids were described briefly. Further algorithms being developed to characterise possible changes in light curve shape and binarity status of the measured object were mentioned. It was pointed out that the expected characteristics of Gaia data in terms of time sampling, phase coverage, photometric accuracy and the amount of data per object are suitable to provide us important information of many previously known and several thousands of new Cepheid variables with unprecedented accuracy using these algorithms. Periods, Fourier parameters and in many cases period change rates can be determined more accurately than ever. Though, for some Cepheids with slowly changing period Gaia data alone might not be enough to gain useful information, combining with simultaneous ground based observations and archival measurements, these data can also lead to important results.

These studies also help us to learn more about the relations between pulsating stars and stellar evolution, moreover due to the fact that Cepheids are fundamental distance indicators, the cosmic distance scale can also be established more precisely.

### 5. Acknowledgements

This research is supported by the European Space Agency (ESA) and the Hungarian Space Office via the PECS programme (contract No. 98090).

### References

Berdnikov L N 2008 *VizieR Online Data Catalog* II 285

Klagyivik P and Szabados L 2009 *A&A* **504**, 959

Soszynski I, Poleski R, Udalski A, Szymanski M K, Kubiak M, Pietrzynski G, Wyrzykowski L, Szewczyk O and Ulaczyk K 2008 *Acta Astr.* **58**, 163