OPEN EUROPEAN JOURNAL ON VARIABLE STARS
November 2018 http://var.astro.cz/oejv ISSN 1801–5964

NEW TOOL WITH GUI FOR FITTING O-C DIAGRAMS

Gajdoš, Pavol[1] and Parimucha, Štefan[1]

1) Institute of Physics, Faculty of Science, Pavol Jozef Šafárik University, Košice, Slovakia,
pavol.gajdos@student.upjs.sk

Abstract: There are many different methods for fitting and analysing O-C diagrams. We present a new fitting tool for an analysis of O-C diagrams. Our package use Genetic Algorithms (GA) and Markov Chain Monte Carlo (MCMC) methods. Unlike many others fitting routines, our method does not need any initial values of fitted parameters. Fitting using presented software is quite simple thanks to a very intuitive graphic user interface. Currently, nine most common models of periodic O-C changes are included in this software.

1 Introduction

The precise timing of minima of eclipsing binaries (EBs) or extrema times of any other periodic events, which is a quite simple observational task and is also achievable with amateur's technique, could help us to discover interesting processes in the studied system. The most common phenomena are connected with a mass transfer between the components of EB, the presence of another body in the system, apsidal motion or angular momentum lost from the system.

Unfortunately, the exact physical model of many of these effects is complicated and strongly non-linear. Using classical fitting routines is problematic. The good initial values of parameters are needed by many of them. One can use the approximation of the model by Taylor or Fourier series but the physical properties of the model are loosening. In this paper, we present a new fitting tool for analysing the changes of times of minima of EBs or exoplanetary transits. We use Genetic Algorithms to remove the necessity of any input values of the model's parameters. Final values of them together with their statistically significant uncertainties are obtained using Markov chain Monte Carlo fitting. The combination of these two algorithms allows us to analyse the exact physical model of the observed variations.

2 Period changes models of eclipsing binaries and exoplanets

Minima times T_C of EBs or transits of extrasolar planets can be simply calculated by the linear ephemeris

$$T_C = T_0 + P \times E \tag{1}$$

which predicts minima times T_C of EB with orbital period P. Here E is an epoch of the observation and it counts how many eclipses elapsed since the zero epoch (i.e. from time T_0).

Minima times determined from observations for the same epoch (T_O) are in general different to times T_C. The behaviour of this difference (O-C), often shown in O-C diagrams, is caused by perturbation δT

$$T_O - T_C \equiv O - C = \delta T. \tag{2}$$

This perturbation is generally a sum of different effects and it indicates period changes of the binary system. Formally, we can write

$$\delta T = (\Delta T_0 + \Delta P \times E) + Q \times E^2 + \delta T_i \tag{3}$$

where the part in bracket generates linear trend in (O-C)s and it is caused by wrong linear ephemeris (Equation 1), a quadratic term ($Q \times E^2$) describes changes due to mass transfer and δT_i means more complex periodic variations in (O-C)s (described in the next sections).

2.1 The light-time effect

The apparent changes of binary stars period caused by the distance variation of the system from the observer are often called the light-time effect (LiTE). The EB moves around the common barycentre of a wider triple system. It produces periodic variations in the observed minima times with respect to linear ephemeris of the EB with a period corresponding to the period of the third body. The secondary minima have identical behaviour in (O-C)s with primary ones.

Irwin (1952) determined analytical formula for O-C changes caused by LiTE, which can be written in the form

$$\delta T_{LiTE} = \frac{a_{12} \sin i_3}{c} \left[\frac{1 - e_3^2}{1 + e_3 \cos \nu_3} \sin(\nu_3 + \omega_3) + e_3 \sin \omega_3 \right], \tag{4}$$

where $a_{12} \sin i_3$ is the projected semi-major axis of the binary star around the barycentre of a triple system, i_3 is the inclination of the third body's orbit, e_3 is the eccentricity, ν_3 is the true anomaly of the binary orbit around the system's barycentre and ω_3 is the longitude of pericenter. Two another orbital parameters, the period of the third body P_3 and the time of pericenter passage t_{03} are hidden in ν_3 calculation, which have to be solved using Kepler equation. Equation 4 can be simply extended for another body(ies) in the system.

Because we are not able to find inclination of the orbit i_3 only from O-C analysis, we can determine only so-called mass function

$$f(M_3) \equiv \frac{(M_3 \sin i_3)^3}{M^2} = \frac{(a_{12} \sin i_3)^3}{P_3^2} \tag{5}$$

of the third body. Here, $M = M_1 + M_2 + M_3$ is a total mass of the system (M_i - masses of components).

Semi-amplitude of changes on O-C diagram generated by LiTE is given by equation:

$$K_3 = \frac{a_{12} \sin i_3 \sqrt{1 - e_3^2 \cos \omega_3}}{c} \tag{6}$$

2.2 Agol's models

Agol et al. (2005) presented two models of O-C changes concerning on transits of exoplanets. O-C variations are caused by gravitational interaction with another body (planet) in the system. They assumed that the orbits of transiting exoplanet and the third body are coplanar with inclination $i = 90°$.

The first one is the model of perturbations due to an interior planet on small orbit:

$$\delta T_{AgolIn} = -P\frac{\mu_3 r_3 \cos(\nu + \omega_3)\sqrt{1-e^2}}{2\pi a(1 - e\sin\omega)} \tag{7}$$

where P is a period of transiting planet, a is its semi-major axis, e eccentricity of its orbit, ν its true anomaly and ω its longitude of pericenter; r_3 is the radius of the orbit of the 3^{rd} body, ω_3 its longitude of pericenter and $\mu_3 \equiv M_3/M$ ($M = \sum_i M_i$ is a mass of the whole system) is its reduced mass. Semi-amplitude of changes on O-C diagram generated by Agol's interior perturber is given by the equation:

$$K_3 = P\frac{\mu_3 r_3 \sqrt{1-e^2}}{2\pi a\,(1 - e\sin\omega)}. \tag{8}$$

For perturbations due to an exterior planet on a large eccentric orbit, Agol et al. (2005) determined the formula:

$$\delta T_{AgolEx} = \frac{\mu_3}{2\pi(1-\mu_3)}\frac{P^2}{P_3}\left(1-e_3^2\right)^{-3/2}(\nu_3 - l_3 + e_3\sin\nu_3) \tag{9}$$

where P_3 is the period of the 3^{rd} body, e_3 eccentricity of its orbit, ν_3 its true anomaly and l_3 its mean anomaly. Other parameters are the same as for a case of interior perturber. Semi-amplitude of changes on O-C diagram generated by this model is given by the equation:

$$K_3 \approx \frac{\mu_3}{2\pi(1-\mu_3)}\frac{P^2}{P_3}\left(1-e_3^2\right)^{-3/2} 2\left[\arctan\frac{e_3}{1+\sqrt{1-e_3^2}} + e_3\right]. \tag{10}$$

2.3 Apsidal motion

Apsidal motion is the precession of the orbit due to a gravitational quadruple moment induced by tidal distortion in a binary star. The secondary minima are in antiphase with primary ones in the O-C diagram. The pericenter position ω at epoch E is defined by the linear equation $\omega = \omega_0 + \dot{\omega} \times E$ where ω_0 is initial position of pericenter and $\dot{\omega}$ is angular velocity of line of the apsides.

Giménez & Bastero (1995) determined the formula for O-C changes in the form

$$\delta T_{APS} = \frac{P}{\pi}\sum_{n=1}^{\infty}(-\beta)^n\left(\frac{1}{n} + \sqrt{1-e^2}\right)\sin n\nu \tag{11}$$

where P is anomalistic period of EB (i.e. time between two primary or secondary eclipses), $\nu = \theta - \omega + \pi/2$ is true anomaly ($\theta = 0$ for primary minima and $\theta = \pi$ for secondary minima) and $\beta = e/\left(1+\sqrt{1-e^2}\right)$.

Semi-amplitude of changes on O-C diagram is

$$K_3 \approx \frac{Pe}{\pi}. \tag{12}$$

The sidereal period of EB is given by the equation

$$P_S = P\left(1 - \frac{\dot{\omega}}{2\pi}\right) \tag{13}$$

and the period of apsidal motion (and also period of the variation in O-C diagram) is

$$U = P_S\frac{2\pi}{\dot{\omega}} = P\left(\frac{2\pi}{\dot{\omega}} - 1\right). \tag{14}$$

3 Fitting methods

Many classical numerical methods based on the iterative minimization of the sum of squares (e.g. Levenberg-Marquardt algorithm or Simplex method) are used to obtain the optimal set of model's parameters. These algorithms can be simply implemented in many programming languages and data analysis packages and solution can be found relatively fast. However, the quality of the results (i.e. their convergence to the global optimum) is strongly dependent on an initial guess of fitted parameters which have to be close enough to the final one.

Here we introduce our approach to improve it. Our method does not require any starting values. It needs only searching interval where the values of parameters should be. The Genetic algorithms (GA) are used to obtain the first guess of parameters. They are improved by Markov Chain Monte Carlo (MCMC) fitting. MCMC is also used to calculating the realistic statistically significant estimation of uncertainties of the fitted parameters. These errors are underestimated in many cases if some classical fitting method, which calculates errors using the covariance matrix, is used.

3.1 Genetic Algorithms

Genetic Algorithms are commonly used for solving optimization problems which could not be solved using classical optimization methods (Weise, 2011). The basic scheme of GA is inspired by biologic evolution. The general principles of evolution could be summarized in several points:

1. There is a population of individuals.

2. There is a variability of properties of individuals.

3. Individuals with different properties have a different reproductive capacity.

4. Properties of descendant are correlated with properties of parents.

Similarly to biologic evolution, generating a new generation of individuals in GA include the impact of crossover and mutations which change individuals in the next generation. When fitting data using GA, individuals are made up of values of fitted parameters of the model. Crossover is realized by changing the values of some parameters between two individuals in parent generation. Mutations cause changing of the parameter by adding random value from a Gaussian distribution.

The main advantage of fitting data with GA is that start values of fitted parameters are unnecessary. We only need to know search limits for these parameters. In our case, we could set really wide limits which could be simply set from physics of the EBs or directly from O-C diagram. Fitting is not very computationally intensive and takes a relatively short time comparing to Monte Carlo method. On the other hand, GA fitting does not give any information about the errors of fitting parameters. This problem could be solved using MCMC fitting.

3.2 Monte Carlo method

Monte Carlo is one of the most used stochastic optimization methods (Brooks et al., 2011). The basic idea is to obtain states of the studied system which are from the required distribution, by generating random numbers from the uniform distribution. We use Metropolis algorithm which generates Markov chains. The MCMC method converges to the normal distribution of states. Uncertainties of all parameters can be easily estimated from this final distribution.

MCMC method gives us, unlike the GA, the estimation of errors of fitted parameters. However, MCMC fitting needs to know the initial values of parameters. For this reason, we have to use another fitting method before using MCMC itself. Fitting with MCMC is very computationally intensive and takes approx. five times longer than fitting with GA (with the same number of calculations of the model).

4 Package description

We developed a stand-alone package written in python for fitting O-C diagrams of EBs and exoplanets. Its name is OCFit. The package contents four main classes for fitting linear (FitLinear) and quadratic trends (FitQuad) and complex changes on O-C diagram (OCFit) and for loading the saving state of O-C analysis (OCFitLoad). There are also three auxiliary classes for working with GA and for analysis of GA and MCMC sampling.

The parameters of class OCFit could be saved to file using its function Save. After that, the state of class OCFit could be fully restored using function Load or using class OCFitLoad.

Our python package allows performing the complex study of O-C changes. Its detail description is above the scope of this paper. The manual for using our class in own python script could be found on our GitHub page[1].

[1] https://github.com/pavolgaj/OCFit

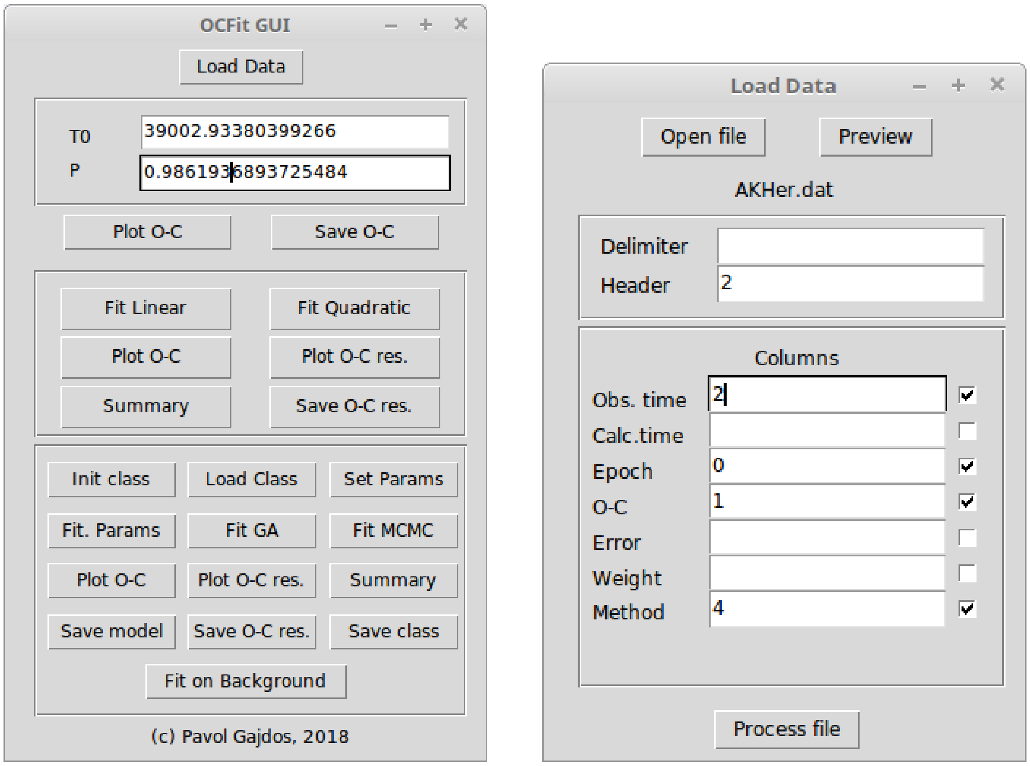

Figure 1: Main window of GUI (*left*). Window for loading data (*right*).

5 Graphic user interface

Our graphic user interface (GUI) can be used to basic control of class OCFit (see Sec. 4). Not all functions and control parameters are available here. However, the combination of this GUI with your own python script to work with OCFit class is still possible.

Design of GUI is quite simple (see Figure 1). It is created using standard python package Tkinter[2]. Many buttons are available only after successful executing previous necessary actions (e.g. showing the results is not available before fitting).

5.1 Loading data and linear ephemeris

Input O-C data can be loaded from any text file in which are data stored in columns. The columns could be separated by any character which is not used in data columns. Using header with an optional number of rows is also possible. Rows marked as python comments (i.e. started with "#") are skipped automatically.

The columns in input file could be:

- observed times of minima in Julian dates
- calculated times of minima in Julian dates

[2] https://wiki.python.org/moin/TkInter

- epochs
- O-Cs in days
- errors of O-C in days
- weights
- used methods of O-C determination

Not all of them are necessary. The minimal required data are observed times of minima or calculated times with O-Cs. The setting of errors is also recommended. If the column with the used methods is presented, the error or weight of each individual method will be set in the next step. It is better to use errors of O-C measurements instead of their weights. Figure 1 shows the window for loading data. The parameters in this figure are set for loading data from a text file with O-C data from on-line database O-C gateway operated by the Czech Astronomical Society[3] (Paschke & Brat, 2006).

The next step of the basic workflow is to enter the linear ephemeris of the studied system. The linear ephemeris should be very accurate. After that, the O-C diagram calculated from this ephemeris could be shown. If the linear ephemeris is not correct or also the quadratic trend is presented on the O-C diagram, it could be fitted separately before analysing more complicated changes. Now, the OCFit class could be initialized.

These steps could be skipped by loading already saved OCFit class from the file.

5.2 Parameters of model

Currently, there is nine model available for fitting. Adding other models is, of course, possible and it is planned. Available models are

- LiTE3 - LiTE caused by the 3^{rd} body (see Sec. 2.1)
- LiTE34 - LiTE caused by the 3^{rd} and 4^{th} body (see Sec. 2.1)
- LiTE3Quad - combination of LiTE3 model and quadratic trend
- LiTE34Quad - combination of LiTE34 model and quadratic trend
- AgolInPlanet - model of inner perturber (see Sec. 2.2)
- AgolInPlanetLin - combination of AgolInPlanet model and linear trend
- AgolExPlanet - model of exterior perturber (see Sec. 2.2)
- AgolExPlanetLin - combination of AgolExPlanet model and linear trend
- Apsidal - model of apsidal motion (see Sec.2.3)

It is necessary to set the values of all the fixed parameters of the model. If only the MCMC method (see Sec. 3.2) will be used, the values of fitted parameters are also

[3] http://var2.astro.cz/ocgate/

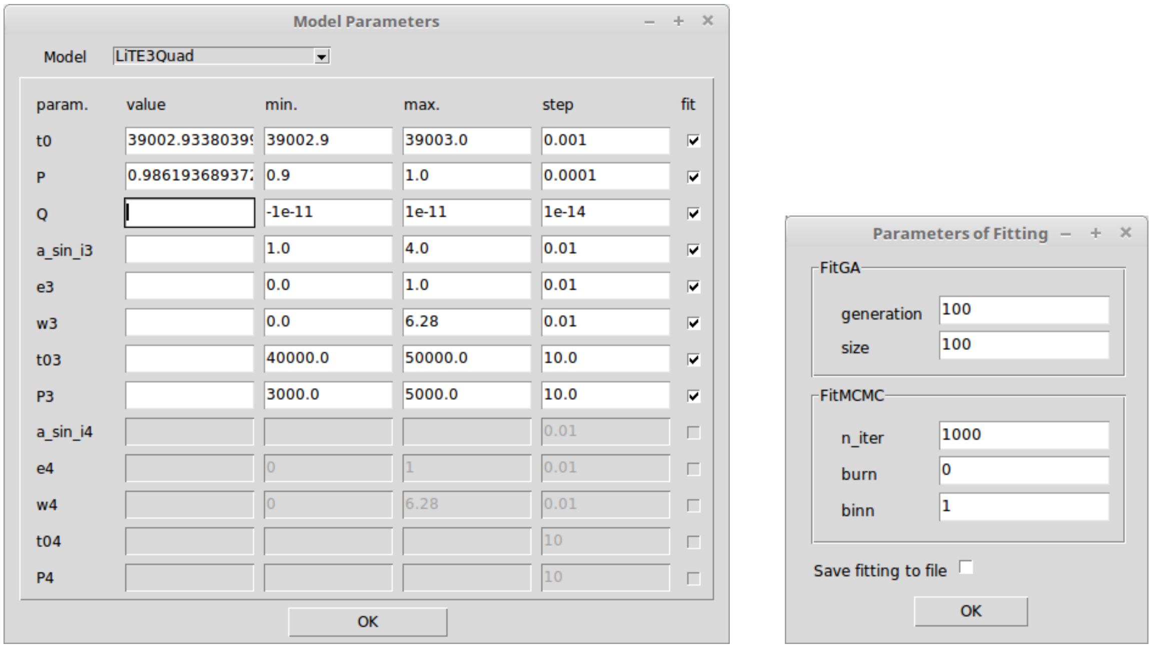

Figure 2: Windows for setting parameters of the model (*left*) and that of the fitting methods (*right*).

required. The limits (minimum and maximum) and steps of all fitted parameters are needed in a case of using any of fitting method.

The limits of the parameters could be set extremely large in many cases (see Figure 2) and could be guessed from the O-C diagram. For example, eccentricity (e) and longitude of pericenter (w) could run in their whole possible range, i.e. e from 0 to 1 and w from 0 to 2π. The relative width of the interval of a period of the 3^{rd} (or 4^{th}) body could be about 40% and the same for its time of pericenter passage. The values of the projected semi-major axis of the binary star around the barycentre (a_sin_i) could go from very small values to values a few times larger than the real value. However, the interval for linear ephemeris should be as small as possible. In general, we recommend not to fit linear ephemeris together with more complicated models at all.

5.3 Fitting model

For the successful fitting of the model, setting good parameters of the fitting method is needed (see Figure 2). The parameters (together with their suggested values) of GA (see Sec. 3.1) are

- generation - the number of generations; 100 for testing and at least 1000 for final fitting
- size - size of one generation; 100 for testing and at least 1000 for final fitting

The values of both parameters are the same in many cases. Generally, the extending the size of generation would produce a better solution than increasing the number of

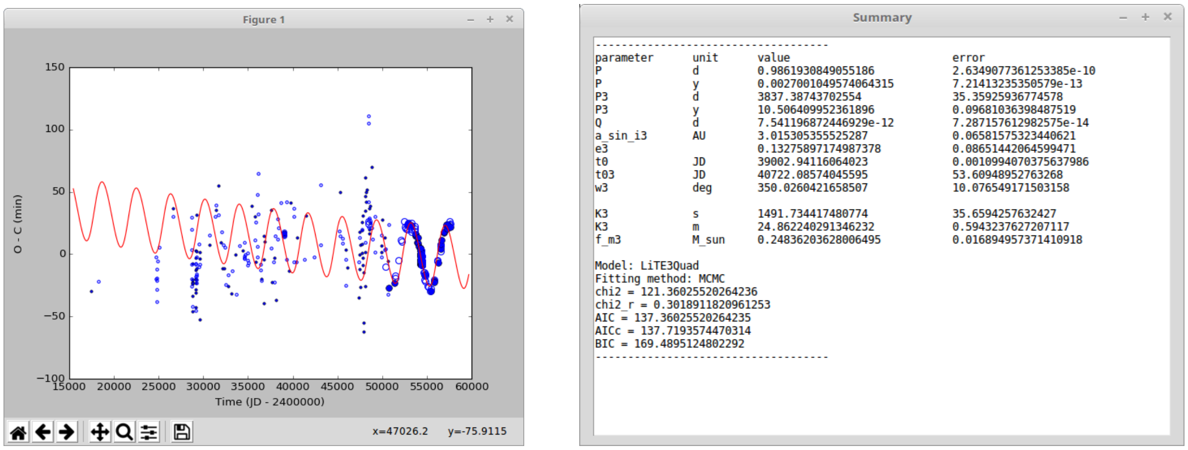

Figure 3: Graphic window with fitted O-C diagram of AD And (*left*) and window with summary of the model's parameters for the same O-C diagram (*right*).

generation by the same value. The number of calculations of the model is *the number of generations* × *the size of one generation*.

The setting of the good values of the MCMC method (Sec. 3.2) could be more complicated. They are

- n_iter - number of iterations; 1000 for testing and at least 10^6 for final fitting
- burn - number of the removed MC steps (before equilibrium); 0 for testing and at least 1000 for final fitting
- bin - size of one block for binning; 1 for testing and at least 10 for final fitting

In both fitting methods, it is necessary to set an appropriate number of iterations. This number is a compromise between computing time and the quality and accuracy of the results. The fitting routine (obtained sampling of the parameters) could be saved to file for later analysis. We use our own set of functions written in python for this purpose for GA sampling. Functions of python modules pymc (Patil et al., 2010) and PyAstronomy could be used to analyse MCMC sampling. These analyses are useful for evaluating the quality of the fitting.

The separate fitting using GA and MCMC is possible. We recommend using GA as a first fitting method. The first guess of values of parameters is obtained in this step. Their final values together with the estimation of their uncertainties are gotten using the MCMC method. It is possible to run fitting also on the background or to a generate fitting script which could be run later and/or on another computer (e.g. on computing server).

5.4 Working with results

We use python module matplotlib (Hunter, 2007) for displaying the final O-C diagram. The original O-C data together with model values could be displayed, e.g. O-C diagram of AD And in Figure 3. To check the quality of the fitting, the plot of residual O-Cs could be used.

All model's parameters are listed in a separated window (see Figure 3). The amplitude of (O-C)s (K) is also determined according to the used model (using the equations in Sec. 2). We also calculate the basic statistical indicators for evaluating the goodness of fit. They are the chi-squared error (χ^2), reduced chi-squared statistic (χ^2_r), Akaike information criterion (AIC), corrected Akaike information criterion (AICc) and Bayesian information criterion (BIC).

The model (O-C)s and residual (O-C)s could be saved to a text file. Eventually, the parameters of whole class OCFit could be saved to file and be used for later analysis. We strongly recommend to do it.

6 Conclusion

The presented python package OCFit combines the advantages of GA and MCMC method to analyse O-C diagrams. The only disadvantage could be quite long computing time but this is not so big problem considering the current state of computer technology. We currently included nine models of O-C changes but it is possible to add other ones. We developed simple GUI to make work with our module easier. Both of them, python package and GUI, are available on our GitHub page[4].

Our method gives very good results which are at least comparable with any other method. We already used it in two our papers (Gajdoš et al., 2017; Parimucha et al., 2018).

Acknowledgements: This paper has been supported by the grant of the Slovak Research and Development Agency with number APVV-15-0458. This article was created under project ITMS No.26220120029, based on the supporting operational Research and development program financed from the European Regional Development Fund. The research of P.G. was supported by internal grant VVGS-PF-2017-724 of the Faculty of Science, P. J. Šafárik University in Košice.

References

Agol, E., Steffen, J., Sari, R., & Clarkson, W., 2005, *MNRAS*, **359**, 567, 2005MNRAS.359..567A

Brooks, S., Gelman, A., Jones, G.L. & Meng, X., 2011. Handbook of Markov Chain Monte Carlo. Chapman & Hall/CRC, New York.

Gajdoš P., Parimucha Š., Hambálek Ľ., Vaňko M., 2017, *MNRAS*, **469**, 2907, 2017MNRAS.469.2907G

Giménez, A., & Bastero, M., 1995, *Ap&SS*, **226**, 99, 1995Ap&SS.226...99G

[4] `https://github.com/pavolgaj/OCFit`

Hunter, J. D., 2007, *Comput Sci Eng*, **9**, 90, 2007CSE.....9...90H

Irwin, J. B. 1952, *ApJ*, **116**, 211, 1952ApJ...116..211I

Parimucha Š., Gajdoš P., Kudak V., Fedurco M., Vaňko M., 2018, *RAA*, **18**, 47, 2018RAA....18...47P

Paschke, A., & Brat, L., 2006, *OEJV*, **23**, 13, 2006OEJV...23...13P

Patil, A., Huard, D. & Fonnesbeck, C. J., 2010, *J Stat Softw.*, **35**, 1, doi:10.18637/jss.v035.i04

Weise, T., 2011. Global optimization algorithm - theory and application. 3rd ed. http://www.it-weise.de/projects/bookNew.pdf