Planetary and Space Science 123 (2016) 117–125

Contents lists available at ScienceDirect

Planetary and Space Science

journal homepage: www.elsevier.com/locate/pss

H, G_1, G_2 photometric phase function extended to low-accuracy data

A. Penttilä [a,*], V.G. Shevchenko [b], O. Wilkman [a], K. Muinonen [a,c]

[a] Department of Physics, P.O.Box 64, 00014 University of Helsinki, Finland
[b] Institute of Astronomy, Kharkiv Karazin National University, Kharkiv 61022, Sumska str. 35, Ukraine
[c] Finnish Geospatial Research Institute, P.O.Box 15, 02431 Masala, Finland

ARTICLE INFO

Article history:
Received 28 April 2015
Received in revised form
12 August 2015
Accepted 13 August 2015
Available online 28 August 2015

Keywords:
Photometry
Magnitude
Asteroids

ABSTRACT

We introduce a constrained nonlinear least-squares algorithm to be used in estimating the parameters in the H, G_1, G_2 phase function. As the algorithm works directly in the magnitude space, it will surpass the possible bias problem that may be present in the existing H, G_1, G_2 fit procedure when applied to low-accuracy observations with large magnitude variations. With constraints on the photometric phase-curve shape parameters G_1 and G_2, it guarantees a physically reasonable phase-curve estimate. With a new data set of 93 asteroids, we re-assess the two-parameter version of the H, G_1, G_2 function. Finally, we introduce a one-parameter version of the phase function that can give a suggestion of the asteroids taxonomic group based only on its phase curve. A statistical model selection procedure is presented that can automatically select between the different versions of the photometric phase functions. An online tool that implements these algorithms is introduced.

1. Introduction

The reliable estimation of the absolute magnitude H for an asteroid from the photometric observations is extremely important. The absolute magnitude (the apparent magnitude at 1 au from the Sun and the Earth, observed in the backscattering direction) relates the brightness of the asteroid to its size, if the albedo of the target is known. Furthermore, the shape of the photometric phase curve (i.e., magnitude as a function of the phase angle) can serve as a proxy for the taxonomic type of the asteroid in cases when spectral information is not available (see, e.g., Oszkiewicz et al., 2012; Shevchenko et al., 2016, and Section 3.5).

The International Astronomical Union (IAU) adopted the so-called H, G photometric phase function in 1989 to be used when estimating the shape of the photometric phase curve and the absolute magnitude H (Bowell et al., 1989). In 2012, the IAU adopted the so-called H, G_1, G_2 function (Muinonen et al., 2010). The H, G_1, G_2 function improved especially the backscattering behavior of the curve with high- and low-albedo asteroids.

Both the H,G and the H, G_1, G_2 phase functions can be applied to targets with multiple high-quality observations. If the number of observations is small, or their accuracy is low, problems may arise. The most apparent problem is that, especially, the parameter G or the parameters G_1, G_2 might be poorly estimated. The solution with the H, G has been to fix value of G to a constant value and estimate only the H. The H, G_1, G_2 offers an improved solution with a two-parameter H, G_{12} function. Nevertheless, even the H, G_{12} function fit can be nonreliable with very small data sets, and procedures of using a fixed G_{12} values have been applied.

We offer a solution that we believe to improve the current situation with the photometric fits with a small number of low-accuracy observations. After a short definition of the H, G_1, G_2 phase function in Section 2, we present a constrained nonlinear least-squares method for fitting the H, G_1, G_2 function that can improve the possible bias with low-accuracy data (Section 3). Then, we revisit the two-parameter phase function with new data in Section 3.4 and offer a new version of that, the H, G_{12}^* phase function. In Section 3.5 we asses the problem with fixed G or G_1, G_2 parameters by introducing one-parameter models that relate to five taxonomic asteroid groups. We tie all the models with three, two or one parameter together in Section 3.6 by introducing a statistical model selection procedure to select the best version to be used with a particular data set. We have developed an online tool that implements the algorithms (Section 4), and show one application example with the near-Earth asteroid (144411) 2004 EW9 in Section 4.1.

2. The H, G_1, G_2 phase function

The H, G_1, G_2 photometric phase function is already extensively presented in Muinonen et al. (2010), thus we recall here only a short summary of the key concepts. The data to be modeled consists of triplets $(\alpha_i, V(\alpha_i), \sigma_i)$, where α is the phase angle between the Sun and the observer, as seen from the target. $V(\alpha)$ is the reduced

* Corresponding author. Tel.: +358 294151227
E-mail address: Antti.I.Penttila@helsinki.fi (A. Penttilä).

http://dx.doi.org/10.1016/j.pss.2015.08.010
0032-0633/

Table 1
H, G_1, G_2 basis functions $\Phi_{1,2,3}(\alpha)$. For details about the splines $\xi_{1,2,3}(\alpha)$, see Appendix A, and for the tabulated basis function values, Appendix B.

Range (deg)	$\Phi_1(\alpha)$	Range (deg)	$\Phi_2(\alpha)$	Range (deg)	$\Phi_3(\alpha)$
0–7.5	$1-\frac{6}{\pi}\alpha$	0–7.5	$1-\frac{9}{5\pi}\alpha$	0–30	$\xi_3(\alpha)$
7.5–150	$\xi_1(\alpha)$	7.5–150	$\xi_2(\alpha)$	30–150	0

observed magnitude at phase angle α, and σ is the error (standard deviation) of the observation. Note that σ is always included in the H, G_1, G_2. If it is not given in the data, an implicit value of 0.03 mag is used. There is also an implicit assumption in the model that the error distribution in the reduced magnitude value is symmetric.

In the (reduced) magnitude value space, the H, G_1, G_2 model is of form

$$V(\alpha) = H - 2.5\log_{10}[G_1\Phi_1(\alpha) + G_2\Phi_2(\alpha) + (1 - G_1 - G_2)\Phi_3(\alpha)], \quad (1)$$

where H, G_1, and G_2 are the parameters of the model, and Φ_i are the basis functions. The basis functions are composite functions consisting of linear parts (in $\Phi_{1,2}$), constant part (in Φ_3), and parts defined by cubic splines $\xi(\alpha)$ (see Appendix A for details of the spline implementation). The model is valid from $\alpha = 0^\circ$ to 150°, and the basis functions are given in Table 1. For convenience, we also give the tabulated values of the basis functions in Appendix B.

In the original H, G_1, G_2 the fit between the data and the model is not done in the magnitude values, but in flux values. The magnitudes V are converted to flux F with a nonlinear relation $F = 10^{-0.4V}$. After the conversion, the model in Eq. (1) can be written as

$$F(\alpha) = a_1\Phi_1(\alpha) + a_2\Phi_2(\alpha) + (1 - a_1 - a_2)\Phi_3(\alpha) \quad (2)$$

with relations

$$H = -2.5\log_{10}(a_1 + a_2 + a_3), \quad G_1 = \frac{a_1}{a_1 + a_2 + a_3}, \quad G_2 = \frac{a_2}{a_1 + a_2 + a_3}. \quad (3)$$

The error in the magnitude space, σ, converts into the flux space, σ_F, depending also on the corresponding magnitude value V, by

$$\sigma_F = 10^{-0.4V}(10^{0.4\sigma} - 1). \quad (4)$$

Finally, the H, G_1, G_2 function in the flux space (Eq. (2)) is fitted to data (α_i, F_i) using the linear least-squares method with the weights $1/\sigma_{F,i}^2$, and the parameter values in the magnitude space are received by applying Eq. (3) to the estimates $(\widehat{a_1}, \widehat{a_2}, \widehat{a_3})$.

The errors for the parameter estimates $(\widehat{H}, \widehat{G_1}, \widehat{G_2})$, as well as other quantities derived from the fit, such as the photometric phase coefficient $k = -(1/(5\pi))(30G_1 + 9G_2)/(G_1 + G_2)$, are estimated using Monte Carlo simulation with the H, G_1, G_2 function. It is assumed that the linear estimates $(\widehat{a_1}, \widehat{a_2}, \widehat{a_3})$ follow the three-dimensional normal distribution. The covariance matrix of the distribution is $\mathbf{\Sigma} = (\mathbf{X}'\mathbf{X})^{-1}$, where $\mathbf{X}$ is the model matrix with the weighted values of the base functions at observed phase angles, $[\mathbf{X}]_{ij} = \Phi_j(\alpha_i)/\sigma_{F,i}^2$.

As m values $(a_1^{(k)}, a_2^{(k)}, a_3^{(k)})$, $k = 1, \ldots, m$ are simulated, Eq. (3) can be used to receive simulated sample $(H^{(k)}, G_1^{(k)}, G_2^{(k)})$. The error for $(\widehat{H}, \widehat{G_1}, \widehat{G_2})$ can then be assessed by computing standard deviation or quantiles from that simulated sample. Using quantiles for error estimation is preferred since the distribution for the nonlinear parameters can be non-symmetric.

3. H, G_1, G_2 fits with low-accuracy data

First of all, we like to stress that the original H, G_1, G_2 photometric phase function described in Section 2 works perfectly well in most cases. Usually, if we have 'enough' data (say, more than 5–10 observations) with modest variation ($\sigma \lesssim 0.05$ mag) and with enough spread in the phase angles, there are no problems. However, there are issues related to sparse data with large variations where an alternative approach in the way the H, G_1, G_2 function is fitted to the data can be fruitful. In what follows, we improve the usability of the H, G_1, G_2 model in these problematic cases, revisit the simplified versions of the model with fewer parameters, and suggest an automated method of selecting between the full model and models with fewer parameters.

3.1. Constraining the fit

In its original form the parameter values, either the flux space parameters (a_1, a_2, a_3) or the magnitude space parameters (H, G_1, G_2), are not constrained in any way. However, there are physical arguments to do so. Firstly, it is generally clear that the flux from the object should decrease when the phase angle increases, since the projected illuminated area of the object, as seen by the observer, is decreasing. Secondly, all the scattering processes that we can identify (i.e., coherent backscattering, shadow-hiding) are amplifying the decrease of flux when moving from exact backscattering. So, we can safely state that the flux space model $F(\alpha)$ in Eq. (2) should have a negative first derivative with all α. Thus, the magnitude space model $V(\alpha)$ in Eq. (1) should always have a positive first derivative.

The requirement of $F'(\alpha) \le 0$ is possible to fulfill by searching (numerically in the case of splines) conditions for (a_1, a_2, a_3). Unfortunately, since there are three composite basis functions, the conditions become quite complicated. To simplify the situation, we can take into account a mathematical or semi-physical requirement that the basis functions $\Phi_{1,2,3}$ in the flux space should have positive coefficients $a_{1,2,3}$. The functions Φ_1 and Φ_2 should bracket the photometric slope behavior of an asteroid (Muinonen et al., 2010, Section 3.2), thus it is natural that $0 \le a_1, a_2$. The basis function Φ_3 is used to introduce backscattering enhancement to the phase function, and therefore is natural to assume that $0 \le a_3$. With the abovementioned conditions, the first derivative of the flux model is always negative.

If we think about the magnitude space parameters (H, G_1, G_2), we can derive conditions for G_1 and G_2 based on the ones for the flux space parameters. From relations in Eq. (3), we can directly see that we can require

$$0 \le G_1, G_2, \quad 1 - G_1 - G_2 \le 1. \quad (5)$$

Introducing these conditions will require changing the linear least-squares method for fitting the model with the data into a general (nonlinear) constrained least-squares method.

3.2. Bias in the magnitude space with the linear fit in the flux space

The parameter of the greatest importance with the H, G_1, G_2 model is the predicted absolute magnitude H of the target. Observations are generally given in magnitudes and not in flux, and data is visualized in phase-magnitude plots. As mentioned earlier, the original H, G_1, G_2 model implicitly assumes symmetric distribution of errors in magnitude values. All these support the general idea that the fit should be unbiased in the magnitude space. Therefore, there is some controversy, at least in theory, in the fact that the function is defined to be fitted in the flux space. Since the transformation between the magnitude and the flux, $F = 10^{-0.4V}$, is nonlinear, the fit cannot be (strictly) unbiased in both spaces, and the error distribution cannot be (strictly) symmetric in both spaces. As the least-squares fit is done in the flux space, the model cannot be unbiased in the magnitude space. As the error distribution is

assumed symmetric in the magnitude space, it cannot be symmetric in the flux space. However, the least-squares fit assumes the errors to be symmetric.

The predecessor of the H, G_1, G_2 phase function, the H, G from Bowell et al. (1989), was also defined so that the fit is done linearly in flux space. It is probably quite safe to assume that the linear fit requiring no iterations was favorable against the iterative nonlinear fit, especially given the limited computing power at the time. While the H, G_1, G_2 function revised the basis functions and added the backscattering enhancement, the linear fit in the flux space was kept. Again, in cases with good data the difference between the linear and the nonlinear fit is negligible. In what follows we try to assess the amount of the bias.

We limit ourselves to the simple case with two observations V_1, V_2 that are done with the same phase angle, and have the same error σ. The estimate for V is simply the (weighted) mean of the magnitudes, $V = (V_1/\sigma^2 + V_2/\sigma^2)/(2/\sigma^2) = (V_1 + V_2)/2$. When we move to the flux space, and convert the magnitudes and errors to flux, the weighted mean in flux is

$$F = \frac{F_1/\sigma_{F,1}^2 + F_2/\sigma_{F,2}^2}{1/\sigma_{F,1}^2 + 1/\sigma_{F,2}^2}. \tag{6}$$

Plugging in $F_i = 10^{-0.4V_i}$ and $\sigma_{F,i} = 10^{-0.4V_i}(10^{0.4\sigma} - 1)$, and moving back to magnitude space $V_F = -2.5\log_{10}(F)$, we get

$$V_F = -5\log_{100}\left(\frac{10^{0.4V_1} + 10^{0.4V_2}}{10^{0.8V_1} + 10^{0.8V_2}}\right). \tag{7}$$

Let us consider the bias $b = V - V_F$ when the two observations are δ magnitude apart from each other, i.e., $V_2 = V_1 + \delta$. In that case,

$$b = \frac{\delta}{2} + 5\log_{100}(1 + 10^{0.4\delta}) - 5\log_{100}(1 + 10^{0.8\delta}). \tag{8}$$

The bias, as a function of the difference δ between the two observations, is plotted in Fig. 1. We can see, for example, that the bias for two observations with the same phase angle and the same error when operating in the flux space, is 0.03 mag when the difference between the observations is about 0.3 mag, and 0.3 mag when the difference is about 1 mag. Since 0.03 mag is the default error in the H, G_1, G_2 model, it could be reasonable to require that the bias should not exceed this. Even though Eq. (8) holds only for two observations, the limit of about 0.3 mag variation in the data could act as the limit when we should start to operate in the magnitude space with nonlinear least-squares (see Section 3.3) instead of the linear least-squares in the flux space. The limit of '0.3 mag in variation' means variation between observations with similar phase angles (say, $\sim 2^\circ$ apart), and not the complete variation over all the phase angles. With high-quality observations of large/bright targets, the bias should not be a problem. With small near-Earth objects, for example, one can easily find targets where the local variation in the observations exceeds 0.3 mag.

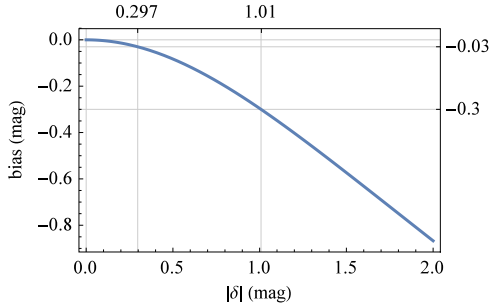

Fig. 1. The bias between magnitude- and flux-space weighted mean for two observations with the same phase angle and magnitude error, as a function of the separation δ between the observations.

3.3. Implementation of the nonlinear fit

In Sections 3.1 and 3.2, we gave two reasons for using the constrained nonlinear least-squares method to fit the observations and the model, namely the need to constrain the possible values of the model parameters, and to reduce the bias. The nonlinear least-squares method is quite straightforward to implement in a computing environment, where a constrained nonlinear optimization algorithm is available. The target function g to be minimized is

$$g(H, G_1, G_2) = \sum_{i=1}^{n} \frac{(V(\alpha_i; H, G_1, G_2) - V_i)^2}{\sigma_i^2}, \tag{9}$$

where $(\alpha_i, V_i, \sigma_i)$ are the ith phase angle, observed magnitude, and error in the data, and $V(\alpha)$ is from Eq. (1). The constraints, as given in Section 3.1, are implemented so that there is an additional penalty factor if the parameter values do not meet the constraints. In our example implementation (see Section 4), we are using the derivative-free simplex optimization code COBYLA for finding the best constrained estimate for the parameters (H, G_1, G_2) (Powell et al., 1994). The estimates (H^L, G_1^L, G_2^L) from the linear fit in the flux space can be used as the starting values for the optimization.

The method to assess the error estimate for the nonlinear fit needs to differ from the one for the linear fit (see Section 2). We cannot assume anymore that the linear parameters would be normally distributed. If there would not be any parameter constraints, we could approximate the error in nonlinear parameters with the normal distribution. However, this approximation is poor if the estimated parameter values are close to the limits set by the parameter constraints $0 \le G_1, G_2, 1 - G_1 - G_2 \le 1$.

To estimate the parameter errors correctly, we can use the so-called bootstrapping method. With the regression problem, the choice is to bootstrap the residuals (Efron and Tibshirani, 1993). In practice, this means that we will form new bootstrap samples (with replacement) from the fitted model and its residuals. Using the estimated parameter values $(\widehat{H}, \widehat{G_1}, \widehat{G_2})$ we can compute the estimated model value at the original phase angles, $\widehat{V_i} = V(\alpha_i; \widehat{H}, \widehat{G_1}, \widehat{G_2})$, and the residuals of the fit, $e_i = V_i - \widehat{V_i}$, where V_i's $(i = 1, \ldots, n)$ are the original observed values.

The k bootstrap samples have the same size n as the original data. In the bootstrap sample j, the ith phase-magnitude pair $(\alpha_i^{(j)}, V_i^{(j)})$ is formed by randomly selecting a number $\eta \in [1, n]$, and then adding the η th residual to the ith phase angle–magnitude estimate pair:

$$(\alpha_i^{(j)}, V_i^{(j)}) = (\alpha_i, \widehat{V_i} + e_\eta). \tag{10}$$

Actually, note that the errors σ_i in the observations need to be taken into account if all the observations do not share the same error value. We do this by using standardized residuals in Eq. (10). So, in fact, we will use $s_\eta = (V_\eta - \widehat{V_\eta})/\sigma_\eta$ instead of the e_η's. As the weights are already included in the bootstrapped magnitudes, a non-weighted nonlinear least-squares fit is used in computing the bootstrapped best-fit parameters.

All the k bootstrap samples $\{(\alpha_i^{(j)}, V_i^{(j)}) \mid i = 1, \ldots, n\}$ should have the same statistical properties as the original one. By fitting the nonlinear model to the bootstrap samples, we receive a sample of k parameter vectors $\{(H^{(j)}, G_1^{(j)}, G_2^{(j)}) \mid j = 1, \ldots, k\}$, and can assess the error in the parameters by computing standard errors or quantiles from these.

One can easily notice that the abovementioned method of estimating the errors in the parameters is far more computing-intensive than the method with the linear fit. Both methods will require simulation, but with the linear fit one needs only to draw random numbers from a three-dimensional normal distribution. With the bootstrap method one needs to iterate the nonlinear optimization for every bootstrap sample, i.e., k times. The number of all the different bootstrap samples for a target with n observations is $\binom{2n-1}{n}$.

For $n = 8$ this number could still be feasible, 6435 combinations. The next case of $n = 9$ gives already 24,310 combinations, so in practice one has to accept a reasonably large k for the error limits to converge, preferably several hundreds or thousands.

3.4. Revisiting the H, G_{12} model

Muinonen et al. (2010) noticed that there is a strong correlation between the G_1 and G_2 parameters with the collection of 22 objects that were analyzed in the article. They decided to benefit from this by introducing a two-parameter model, the so-called H, G_{12} model. The 'metaparameter' G_{12} was related to the G_1 and G_2 by a piecewise linear function that modeled the correlation between G_1 and G_2. The relation was piecewise because the E-type asteroids were behaving different from the others, and needed a distinct treatment. The H, G_{12} function can be estimated with linear least squares in flux space separately for the two parts of the piecewise relation, but a nonlinear method is needed for the error analysis.

Recently, Shevchenko et al. (2016) gathered data of phase-magnitude observations at small phase angles for 93 asteroids. The data consist of selected main-belt asteroids of various taxonomic types, and the reduced magnitude values are averaged over the rotational lightcurves of the targets. We were able to collaborate so that the H, G_1, G_2 function was fitted to their data using the constrained nonlinear method described in Sections 3.1–3.3. The complete list of the targets and the estimated H, G_1, G_2 parameter values are found in Table 3 in Shevchenko et al. (2016). This set of data gave us an excellent opportunity to revisit the G_{12} relation with over 4 times more objects.

The larger data set continues to support the previous result that the E-types behave differently in the (G_1, G_2)-space, see Fig. 2. Furthermore, also the D-types seem to be off from the general trend. In Muinonen et al. (2010), the E-types were separated from the rest of the asteroids both in G_1 and in G_2. With the recent data, however, we start to see S- and M-type asteroids that have similar G_1 values than the E-types. It seems that the piecewise linear model in Muinonen et al. (2010) is not the best approach. We propose to supersede the previous H, G_{12} model with a new H, G_{12}^* model where one linear function of first degree is used to model the G_1, G_2 correlation for the asteroids excluding the E- and D-types, and the E- and D-types will be handled separately. This separate modeling is described in Section 3.5.

Our suggestion for the two-parameter H, G_{12}^* phase function is to model the G_1, G_2 correlation for all the taxonomic types except the E- and D-types with a first degree linear model $y = \beta_0 + \beta_1 x$. Since we can estimate the error for both the G_1 and G_2 parameters, we will fit the model to the data using the (weighted) total least squares (TLS, i.e., orthogonal regression) method (Van Huffel and Vandewalle, 1991). The TLS fit gives us a line from $(G_1, G_2)_1 = (0, 0.53513350)$ to $(G_1, G_2)_2 = (0.84293649, 0)$. As we constrain the H, G_1, G_2 model to $0 \le G_1, G_2 \le 1$, we will construct a parametric relation u as

$$(G_1, G_2) := u(G_{12}^*) = (0, 0.53513350) + G_{12}^*(0.84293649, -0.53513350), \quad (11)$$

where $0 \le G_{12}^* \le 1$. This H, G_{12}^* function can be fitted similar to the constrained nonlinear H, G_1, G_2 function by rewriting Eq. (9) as

$$g_{12}^*(H, G_{12}^*) := g(H, 0.84293649\, G_{12}^*, 0.53513350 - 0.53513350\, G_{12}^*). \quad (12)$$

Note that the values fitted for G_{12} as described in Muinonen et al. (2010) and the values for G_{12}^* as described here are not commensurable.

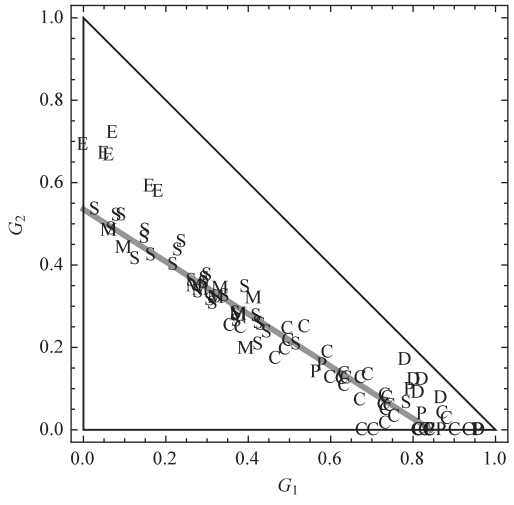

Fig. 2. The fitted G_1 and G_2 values for the 93 asteroids in Shevchenko et al. (2016). The letter indicates the asteroid type, and more complex types are simplified into general types (B, CP, CB → C; V, Q → S; T → D; X, Xc → P). The grey solid line shows the $(G_1, G_2) = u(G_{12}^*)$ relation.

3.5. One-parameter H phase function for targets with small numbers of observations

We have introduced constraints for the parameter values in the H, G_1, G_2 function, and revised the two-parameter simplification, the H, G_{12}^* function. However, there are still cases where the number of observations is small and we have difficulties in fitting both the three- and the two-parameter models. Still, we could benefit from an estimate of the absolute magnitude H of the object, even though the uncertainty of this estimate would be large. For this purpose, we introduce the one-parameter version of the H, G_1, G_2 function, the H function that can be applied even to targets having only one observation.

Shevchenko et al. (2016) constructed the average values for the G_1, G_2 parameters that can be applied to the one-parameter model. By fitting only the H parameter we are only changing the level of the magnitude model, and not the shape. Therefore it suits well to have a small set of possible model shapes, and then select the best based on the goodness of the fit.

The average values for the G_1, G_2 parameters for asteroid classes E, S, M, C, P, and D were given in Shevchenko et al. (2016). The values were computed by minimizing the combined model $g(H_i, G_1, G_2)$ (see Eq. (9)) for all the targets in a specific asteroid class at the same time. The absolute magnitude H_i was left to vary from one asteroid to another, but all the asteroids shared the G_1, G_2 values, i.e., the shape of the photometric phase curve. It can be seen from Shevchenko et al. (2016, Table 4) that the S- and M-type asteroids have very similar G_1, G_2 parameters. That is why we choose to build our one-parameter H model as a set of 5 one-parameter models that correspond to E-, S/M-, C-, P-, and D-type asteroids. Fig. 3 (a)–(e) shows the model, observations, and the values of the fitted G_1, G_2 for these five types. The additional value in this kind of an approach is that we can directly suggest a best-fit asteroid class for an object. The H model is based in minimizing Eq. (9) but with constant values of G_1, G_2, so $g_H(H) := g(H, G_1^t, G_2^t)$, where the pair (G_1^t, G_2^t) is one of the five value pairs shown in Fig. 3(a)–(e).

It can easily be seen from Eq. (1) that, once the parameters G_1, G_2 are taken as constants, the estimate for H is just the weighted mean as

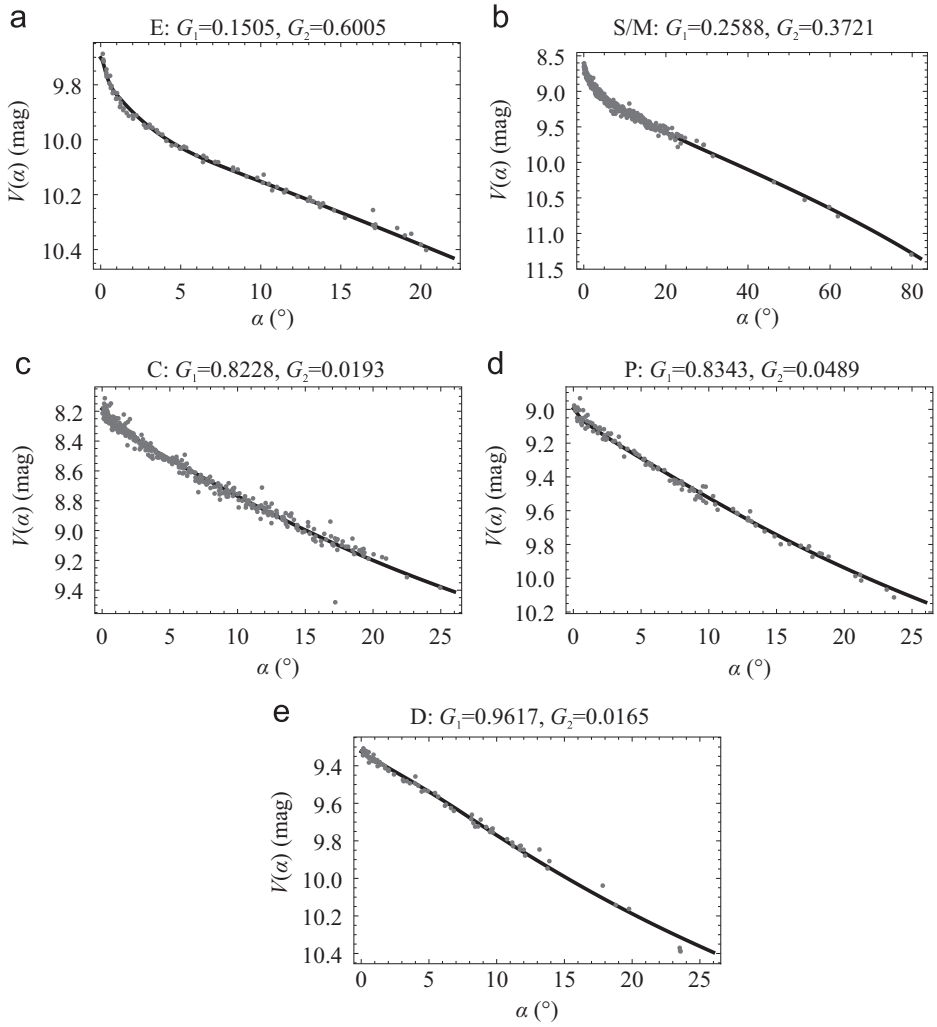

Fig. 3. Observations and average fits for the asteroid classes E, S/M, C, P, and D. The numbers of asteroids and the numbers of observations in the classes are, respectively, E-class, 6 and 82; S- and M-classes together, 40 and 514; C-class, 34 and 371; P-class, 7 and 109; D-class, 6 and 78. The magnitudes of the observations within an asteroid class are shifted so that the joint absolute magnitude for the class is the mean of the absolute magnitudes over the asteroids in that class.

$$\widehat{H} = \frac{\sum_{i=1}^{n}\left(V_i + L_i\right)/\sigma_i^2}{\sum_{i=1}^{n} 1/\sigma_i^2} \quad \text{where}$$

$$L_i = 2.5\log_{10}(G_1^{\mathrm{t}}\Phi_1(\alpha_i) + G_2^{\mathrm{t}}\Phi_2(\alpha_i) + (1 - G_1^{\mathrm{t}} - G_2^{\mathrm{t}})\Phi_3(\alpha_i)). \tag{13}$$

3.6. Automatic selection between three-, two- and one-parameter models

To summarize Sections 3.1–3.5, we have now altogether seven alternative models for the asteroid magnitude-phase relation, one three- and one two-parameter model, and five one-parameter models. There is an evident question of which model to use? In fact, this problem has been around already with the previous H, G function, where constant values for G were often used in cases where data did not allow fitting G reliably. With the H, G_1, G_2, this problem continued with the full and the two-parameter models. To this date, we are not aware of any systematic and quantitatively justified approach on selecting between the models. We are now offering such an automatic selection procedure, based on a statistical model selection.

The (weighted) sum-of-squared-errors (SSE) and the (weighted) root-mean-square (RMS) are generally measures of the goodness of the fit. On the other hand, adding more free parameters to the model improves the fit automatically. To compare models with different numbers of parameters, we need to penalize the SSE with the number of parameters p. In statistics this is done using the adjusted R^2, Mallows's C_p, Akaike information criterion, or the Bayesian information criterion (*BIC*). We prefer to use the last one. The *BIC* for our models is

$$BIC = n\log\left(\frac{\sum_{i=1}^{n}\left(V_i - \widehat{V}_i\right)^2/\sigma_i^2}{\sum_{i=1}^{n} 1/\sigma_i^2}\right) + p\log(n), \tag{14}$$

where p is the number of parameters in the model. The smaller the value of *BIC*, we can fit all the seven models to data, and sort the models from the best to the worst fit. With this automated procedure, the user should not need to decide herself if, e.g., the H, G_{12}^* function should be used or if there is enough data for applying the full H, G_1, G_2 function. In addition, the five one-parameter models corresponding to main asteroid taxonomic types are also ordered, and we receive a suggestion of the most probable taxonomic class for our object.

4. Online implementation

The new functionality in the H, G_1, G_2 phase function that we propose in Section 3 is fully implemented in a web-based online tool

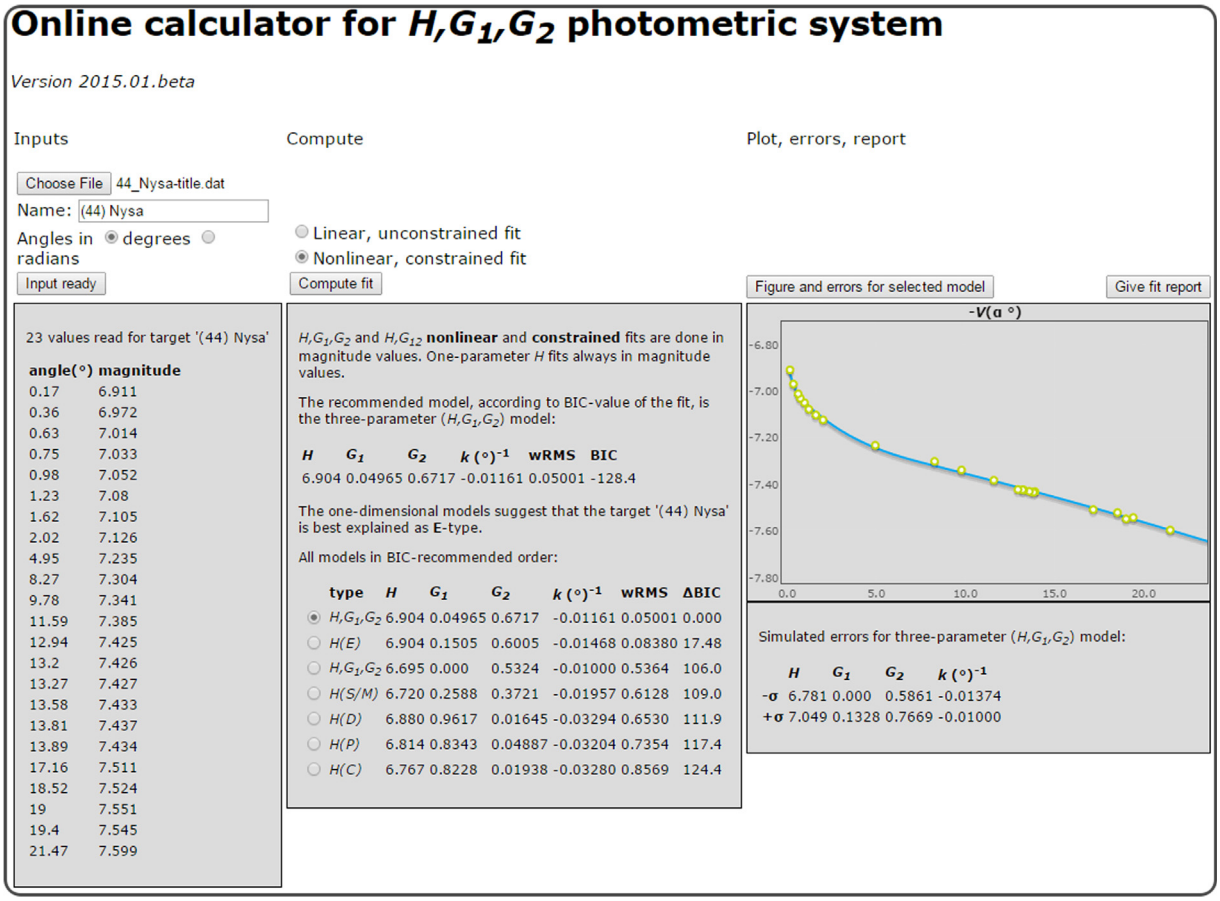

Fig. 4. Screenshot of the OCP user interface. In the example, 23 observations of the asteroid (44) Nysa are used.

(Penttilä, 2015). Tools for photometric phase curve fitting with the H, G_1, G_2 function as it was accepted by the IAU are already available.[1] The existing implementation is available in Fortran 77, Fortran 2003, Mathematica, Java, and Python programming languages. For this latest development, we wanted to offer a platform-independent implementation that does not require any specific computing environment. Therefore we decided to implement the functionality as a web-based application that will run in a web-browser.

Our online H, G_1, G_2 tool (Online Calculator for Photometric phase-curves, OCP) is implemented using the standard languages in the WWW pages, HTML5 and CSS for describing the layout of the page, and JavaScript for implementing the algorithms. The JavaScript runs completely on the client side, i.e., using the resources on the user's computer. The OCP uses two third-party JavaScript libraries, JSCobyla2 (Gustafsson, 2012) and Flotr2 (Humble Software Development, 2014) which are both included in the OCP, and their use requires no action from the user. The JSCobyla2 implements the Cobyla-algorithm (Powell et al., 1994) for the constrained nonlinear derivative-free optimization, and the Flotr2 is used to create the figures of the fitted H, G_1, G_2 function.

The view of the OCP user interface is shown in Fig. 4. First, the user can upload the input file that contains the phase angles and the observed reduced magnitudes of the target. The loaded value pairs are shown on the page. Second, the user can choose to compute either the linear or the nonlinear H, G_1, G_2 fit to the data. The OCP will report the best model, based on the *BIC* statistics (see Section 3.6). Also the taxonomic class that fits the best with the observations among the five single-parameter H models is reported. Finally, a complete list of all the seven fitted models is reported.

Once the fits are ready, the user can choose the model to which the error estimates are simulated and the figure with the fit and the observations is plotted. The errors are reported as $-\sigma$ and σ-limits for the parameters, but to be exact, the limits are the 68.3% confidence interval limits. For normally distributed data, the one-σ interval holds the 68.3% of the data around the best-fit estimate. For non-symmetric distributions (e.g., error distribution for G_1 or G_2 close to zero or one (the interval is not symmetric around the best-fit estimate and cannot be described with a single σ-value. The simulation of the error bounds is usually very fast for linear fits, typically it takes a few seconds. For nonlinear fits the error simulation will take longer, typically some tens of seconds.

Finally, an ASCII-text report of the results can be generated for exporting to other programs. The observations and the best fit are reported, and tabulated values of the fitted model are given.

4.1. Application to one near-Earth asteroid

We will show, as an example, the performance of the linear and nonlinear H, G_1, G_2 models for the near-Earth asteroid (144411) 2004 EW_9. In particular, we choose to fit the H, G_1, G_2 function using the observations by the Pan-STARRS 1 telescope in Haleakala (observatory code F51), and then see how this fit agrees with the 2004 EW_9 observations by all available observatories.

We have 30 observations by Pan-STARRS, ranging from 1.45° to 42.8°. We decided to drop one outlier, which had ~2 mag larger magnitude than the other observations around that phase angle.

[1] HG1G2 tools, http://wiki.helsinki.fi/display/PSR/HG1G2+tools and an online Java tool, see Oszkiewicz et al. (2011)

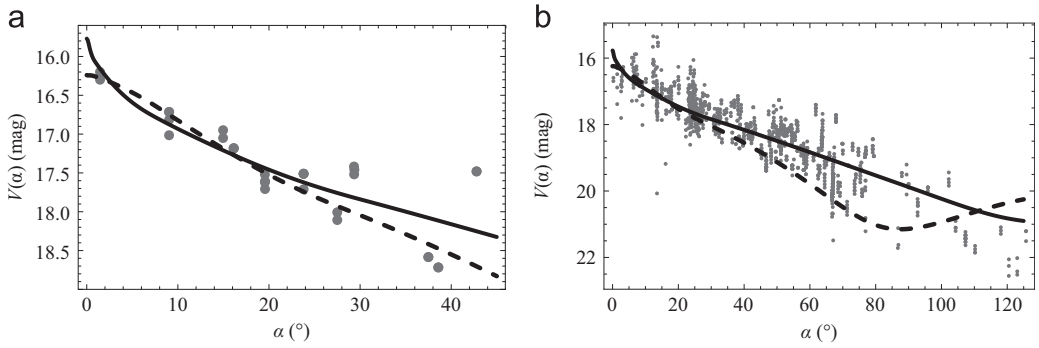

Fig. 5. The linear (dashed line) and nonlinear (solid line) fits to observations of asteroid (144411) 2004 EW_9: (a) only the observations by the Pan-STARRS 1 telescope are shown; (b) all the available data are shown. In both cases, the H, G_1, G_2 fits have been carried out using only the Pan-STARRS data.

Including all the observatories, the data has 1379 observations from 0.25° to 125.7°. The data is publicly available from the IAU Minor Planet Center, but the magnitudes have been corrected to correspond to the V-filter.[2]

Both the linear and nonlinear fits are shown with the Pan-STARRS data in Fig. 5(a). These observations present a typical case where the nonlinear fit is beneficial. The variation in the data is very large, for example there is a difference of 0.69 mag for observations only 1.84° apart. It should be noted that, in Muinonen et al. (2010), the recommendation is to use the H, G_{12} function and not the full three-parameter version with this kind of low-accuracy data. The linear fit gives $H = 16.24$, $G_1 = 1.6478$, $G_2 = -0.5759$, so the conditions mentioned in Section 3.1 are clearly violated with the values estimated for G_1 and G_2. The nonlinear fit gives $H = 15.77$, $G_1 = 0.4438$, $G_2 = 0$. The model selection suggests that the C-type phase curve is the most suitable model among the one-parameter models. Interestingly, Birlan et al. (2013) recently presented spectral observations of this target, and classified it as a Cb-, Xk-, or Cg-type asteroid.

The estimates for the H vary quite a lot between the linear and nonlinear fits. It is hard to say which one is more 'correct' — the backscattering enhancements within the last two degrees are quite different between the fits, but there is no data. In any case, the linear fit is clearly biased towards larger magnitudes with larger phase angles. This is due to the nonlinear transformation when moving to the flux space where the linear fit is done and then back to the magnitude space (see Section 3.2).

If we plot the fits for the Pan-STARRS observations with all the data points (Fig. 5 (b)), we can see how the linear fit will turn and start to have decreasing magnitude as the phase angle increases after about 85°. Also, the fit is clearly too biased to large magnitudes before that. The nonlinear fit is quite reasonable over the whole data range, even though only the Pan-STARRS observations were used in the fit. This example shows the danger of applying linear fit in the flux space with near-Earth asteroid data, or with any other data that has large variations in magnitudes.

5. Conclusions

We offer updated versions of the algorithms that are used to fit the photometric H, G_1, G_2 phase function. These algorithms are based on the constrained nonlinear least-squares fit and the automatic model selection using the Bayesian information criterion. With these updates, the H, G_1, G_2 function can be better applied to observations with large magnitude variations, and to targets with only a few observations.

The proposed two-parameter function H, G_{12}^* is estimated using a much larger set of targets that was possible with the original H, G_{12} model. We are able to simplify the G_1, G_2 relation from the composite linear relation with two functions into a single linear relation. In addition, we recognize five distinct shapes for the photometric phase curve that are related to the asteroid taxonomic types E, S/M, C, P, and D. The automatic model selection criterion both gives the best-fit taxonomic type for the target, and also helps us to decide which model to choose among the three-, two-, and one-parameter functions.

We provide these updated methods, together with the original linear fit H, G_1, G_2 model, as an open online tool requiring only a JavaScript-enabled web-browser. This will help the transition from the H, G function into the IAU-adopted H, G_1, G_2 function.

Acknowledgments

A.P. and K.M. acknowledge the ERC Advanced Grant no. 320773 entitled *Scattering and Absorption of Electromagnetic Waves in Particulate Media* (SAEMPL), and K.M. and O.W. the Academy of Finland Grant no. 1257966 entitled *Electromagnetic Wave Scattering in Complex Media*. We thank Petr Pravek and Alan Harris for their comments on the manuscript.

Appendix A. Cubic spline basis

The basis functions $\xi_{1,2,3}$ in the H, G_1, G_2 function are implemented as cubic splines. The spline system is defined by its knots (i.e., (x,y)-value pairs that the spline should pass), and its derivatives d at the knots. We give the complete list of knot and derivative values in Tables A2 and A3.

Let us have the triplets (x^i, y^i, d^i) ordered according to the knot x^i values so that $x^{(1)} < \cdots < x^{(k)}$, and a value x so that $x^{(j)} \le x < x^{(j+1)}$. Then, the cubic spline value $\xi(x)$ is

$$\xi(x) = (1-t)\,y^{(j)} + t\,y^{(j+1)} + t\,(1-t)((1-t)\,a + b\,t) \text{ where}$$
$$t = \frac{x - x^{(j)}}{x^{(j+1)} - x^{(j)}}, \quad a = d^{(j)}\,(x^{(j+1)} - x^{(j)}) - (y^{(j+1)} - y^{(j)}),$$
$$b = -d^{(j+1)}\,(x^{(j+1)} - x^{(j)}) + (y^{(j+1)} - y^{(j)}). \tag{A.1}$$

[2] Magnitude reduction by M. Granvik, see Granvik et al. (2009).

Table A1
Spline knots and derivatives for $\xi_1(y_1, d_1)$ and for $\xi_2(y_2, d_2)$.

x (deg)	y_1	d_1	y_2	d_2
7.5	7.5×10^{-1}	−1.9098593	9.25×10^{-1}	$-5.7295780 \times 10^{-1}$
30	3.3486016×10^{-1}	$-5.5463432 \times 10^{-1}$	6.2884169×10^{-1}	$-7.6705367 \times 10^{-1}$
60	1.3410560×10^{-1}	$-2.4404599 \times 10^{-1}$	3.1755495×10^{-1}	$-4.5665789 \times 10^{-1}$
90	5.1104756×10^{-2}	$-9.4980438 \times 10^{-2}$	1.2716367×10^{-1}	$-2.8071809 \times 10^{-1}$
120	2.1465687×10^{-2}	$-2.1411424 \times 10^{-2}$	2.2373903×10^{-2}	$-1.1173257 \times 10^{-1}$
150	3.6396989×10^{-3}	$-9.1328612 \times 10^{-2}$	1.6505689×10^{-4}	$-8.6573138 \times 10^{-8}$

Table A2
Spline knots and derivatives for ξ_3.

x (deg)	y_3	d_3
0	1	$-1.0630097 \times 10^{-1}$
0.3	8.3381185×10^{-1}	-4.1180439×10^{1}
1	5.7735424×10^{-1}	-1.0366915×10^{1}
2	4.2144772×10^{-1}	−7.5784615
4	2.3174230×10^{-1}	−3.6960950
8	1.0348178×10^{-1}	$-7.8605652 \times 10^{-1}$
12	6.1733473×10^{-2}	$-4.6527012 \times 10^{-1}$
20	1.6107006×10^{-2}	$-2.0459545 \times 10^{-1}$
30	0	0

Appendix B. Tabulated values of the H, G_1, G_2 basis functions

For convenience, we give here in Table B1 the tabulated basis function values for the H, G_1, G_2 function. These can be used to implement the basis functions with a linear approximation.

Table B1
Tabulated values of the basis functions $\Phi_{1,2,3}(\alpha)$.

α (deg)	$\Phi_1(\alpha)$	$\Phi_2(\alpha)$	$\Phi_3(\alpha)$	α (deg)	$\Phi_1(\alpha)$	$\Phi_2(\alpha)$	$\Phi_3(\alpha)$	α (deg)	$\Phi_1(\alpha)$	$\Phi_2(\alpha)$	$\Phi_3(\alpha)$
0	1.00000000	1.00000000	1.00000000	6	0.80000000	0.94000000	0.14221716	51	0.17795700	0.39463881	0.00000000
0.025	0.99916667	0.99975000	0.99806370	6.5	0.78333333	0.93500000	0.12942389	54	0.16204721	0.36768217	0.00000000
0.05	0.99833333	0.99950000	0.99261656	7	0.76666667	0.93000000	0.11914414	57	0.14746343	0.34202387	0.00000000
0.075	0.99750000	0.99925000	0.98406203	7.5	0.75000000	0.92500000	0.11071705	60	0.13410560	0.31755495	0.00000000
0.1	0.99666667	0.99900000	0.97280360	8	0.73356510	0.91993295	0.10348178	63	0.12187880	0.29417930	0.00000000
0.15	0.99500000	0.99850000	0.94378889	8	0.73356510	0.91993295	0.10348178	66	0.11070881	0.27185223	0.00000000
0.2	0.99333333	0.99800000	0.90880018	9	0.70205874	0.90940957	0.09076411	69	0.10052653	0.25054189	0.00000000
0.25	0.99166667	0.99750000	0.87106525	10	0.67230993	0.89839618	0.07980824	72	0.09126291	0.23021646	0.00000000
0.3	0.99000000	0.99700000	0.83381185	11	0.64424623	0.88692759	0.07025206	75	0.08284886	0.21084408	0.00000000
0.35	0.98833333	0.99650000	0.79968720	12	0.61779520	0.87503862	0.06173347	78	0.07521532	0.19239292	0.00000000
0.4	0.98666667	0.99600000	0.76901633	13	0.59288439	0.86276408	0.05394627	81	0.06829321	0.17483114	0.00000000
0.45	0.98500000	0.99550000	0.74154373	14	0.56944137	0.85013879	0.04680785	84	0.06201346	0.15812689	0.00000000
0.5	0.98333333	0.99500000	0.71701389	15	0.54739370	0.83719757	0.04029153	87	0.05630700	0.14224835	0.00000000
0.55	0.98166667	0.99450000	0.69517129	16	0.52666893	0.82397523	0.03437061	90	0.05110476	0.12716367	0.00000000
0.6	0.98000000	0.99400000	0.67576042	17	0.50719463	0.81050658	0.02901839	93	0.04634750	0.11285041	0.00000000
0.65	0.97833333	0.99350000	0.65852577	18	0.48889834	0.79682645	0.02420818	96	0.04201539	0.09932372	0.00000000
0.7	0.97666667	0.99300000	0.64321182	19	0.47170764	0.78296964	0.01991328	99	0.03809846	0.08660818	0.00000000
0.75	0.97500000	0.99250000	0.62956307	20	0.45555008	0.76897098	0.01610701	102	0.03458670	0.07472832	0.00000000
0.8	0.97333333	0.99200000	0.61732399	21	0.44035322	0.75486528	0.01276361	105	0.03147014	0.06370871	0.00000000
0.85	0.97166667	0.99150000	0.60623909	22	0.42604461	0.74068735	0.00986117	108	0.02873879	0.05357391	0.00000000
0.9	0.97000000	0.99100000	0.59605283	23	0.41255183	0.72647201	0.00737872	111	0.02638265	0.04434847	0.00000000
0.95	0.96833333	0.99050000	0.58650972	24	0.39980241	0.71225408	0.00529530	114	0.02439175	0.03605695	0.00000000
1	0.96666667	0.99000000	0.57735424	25	0.38772394	0.69806837	0.00358992	117	0.02275609	0.02872391	0.00000000
1.25	0.95833333	0.98750000	0.53374972	26	0.37624396	0.68394970	0.00224164	120	0.02146569	0.02237390	0.00000000
1.5	0.95000000	0.98500000	0.49331752	27	0.36529003	0.66993288	0.00122947	123	0.02048884	0.01701331	0.00000000
1.75	0.94166667	0.98250000	0.45592704	28	0.35478972	0.65605272	0.00053245	126	0.01970700	0.01257580	0.00000000
2	0.93333333	0.98000000	0.42144772	29	0.34467057	0.64234406	0.00012962	129	0.01897989	0.00897685	0.00000000
2.5	0.91666667	0.97500000	0.36065328	30	0.33486016	0.62884169	0.00000000	132	0.01816723	0.00613196	0.00000000
3	0.90000000	0.97000000	0.30965499	33	0.30686620	0.58974571	0.00000000	135	0.01712876	0.00395661	0.00000000
3.5	0.88333333	0.96500000	0.26712671	36	0.28089877	0.55271081	0.00000000	138	0.01572421	0.00236629	0.00000000
4	0.86666667	0.96000000	0.23174230	39	0.25685779	0.51762803	0.00000000	141	0.01381329	0.00127648	0.00000000
4.5	0.85000000	0.95500000	0.20228652	42	0.23464320	0.48438840	0.00000000	144	0.01125575	0.00060269	0.00000000
5	0.83333333	0.95000000	0.17798763	45	0.21415492	0.45288296	0.00000000	147	0.00791131	0.00026038	0.00000000
5.5	0.81666667	0.94500000	0.15818479	48	0.19529288	0.42300275	0.00000000	150	0.00363970	0.00016506	0.00000000

References

Birlan, M., Nedelcu, D-A., Popescu, M., 2013. Dynamical evolution and physical properties of six NEAs. Poster Presentation at the Third IAA Planetary Defense Conference, Flagstaff, USA.

Bowell, E., Hapke, B., Domingue, D., Lumme, K., Peltoniemi, J., Harris, A., 1989. Application of photometric models to asteroids. In: Binzel, R., Gehrels, T., Matthews, M. (Eds.), Asteroids II. The University of Arizona Press, Tucson, pp. 524–555.

Efron, B., Tibshirani, R., 1993. An Introduction to the Bootstrap. Chapman & Hall/CRC.

Granvik, M., Virtanen, J., Oszkiewicz, D., Muinonen, K., 2009. OpenOrb: Open-source asteroid-orbit-computation software including Ranging. Meteorit Planet. Sci. 44 (12), 1853–1862.

Gustafsson, A., 2012. Cobyla2 Port to Javascript. Computer Software, retrieved from URL ⟨http://code.google.com/p/jscobyla/⟩.

Humble Software Development, 2014. Flotr2 Canvas graphing Library. Computer Software, retrieved from URL ⟨http://www.humblesoftware.com/flotr2⟩.

Muinonen, K., Belskaya, I., Cellino, A., Delbò, M., Levasseur-Regourd, A.-C., Penttilä, A., Tedesco, E., 2010. A three-parameter magnitude phase function for asteroids. Icarus 209 (2), 542–555.

Oszkiewicz, D., Bowell, E., Wasserman, L., Muinonen, K., Penttilä, A., Pieniluoma, T., Trilling, D., Thomas, C., 2012. Asteroid taxonomic signatures from photometric phase curves. Icarus 219 (1), 283–296.

Oszkiewicz, D., Muinonen, K., Bowell, E., Trilling, D., Penttilä, A., Pieniluoma, T., Wasserman, L., Enga, M.-T., 2011. Online multi-parameter phase-curve fitting and application to a large corpus of asteroid photometric data. J. Quant. Spectrosc. Radiat. Transf. 112 (11), 1919–1929.

Penttilä, A., 2015. Online calculator for H, G_1, G_2 photometric system. Computer algorithm, retrieved from URL ⟨http://www.helsinki.fi/project/psr/HG1G2/⟩.

Powell, M., 1994. A direct search optimization method that models the objective and constraint functions by linear interpolation. In: Gomez, S., Hennart, J.-P. (Eds.), Advances in Optimization and Numerical Analysis. Kluwer Academic Publishers, pp. 51–67.

Shevchenko, V., Belskaya, I., Muinonen, K., Penttilä, A., Krugly, Y., Velichko, F., Chiorny, V., Slyusarev, I., Gaftonyuk, N., Tereschenko, I., 2016. Asteroid observations at small phase angles. IV. Average parameters for the new H, G_1, G_2 magnitude system. Planet. Space Sci. 123, 101–116.

Van Huffel, S., Vandewalle, J., 1991. The Total Least Squares Problems: Computational Aspects and Analysis. SIAM Publications, Philadelphia, PA.