Article

A modified Kwee van Woerden algorithm with reliable error-estimates

Hans J. Deeg [1,2]

[1] Instituto de Astrofísica de Canarias, C. Vía Láctea S/N, E-38205 La Laguna, Tenerife, Spain; hdeeg@iac.es
[2] Universidad de La Laguna, Dept. de Astrofísica, E-38206 La Laguna, Tenerife, Spain
Correspondence: hdeeg@iac.es

Abstract: The Kwee van Woerden (KvW) method for the determination of eclipse minimum times has been a staple in eclipsing binary research for decades, due its simplicity and independence of external input parameters. However, its estimates of the timing error have been known to be of low reliability. During the analysis of very precise photometry of CM Draconis eclipses from TESS space mission data, KvW's original equation for the timing error estimate produced numerical errors, which evidenced a fundamental problem in this equation. This contribution introduces an improved way to calculate the timing error with the KvW method. A code that implements this improved method, together with several further updates over the original method is presented as well. An example application on the CM Draconis light curves from TESS is given, where we show that its timing error estimates of about 1 second are in excellent agreement with error estimates obtained by other means.

Keywords: Eclipsing binary minima timing method; Transit timing variation method; Eclipsing binary stars; CM Draconis; TESS space mission; Computational Methods

1. Introduction

The Kwee van Woerden method, henceforth KvW, has been very popular for eclipse minimum time determination since its publication in 1956 [1]. This is due to its computational simplicity as well as due to its independence from assumptions about the data that are being analyzed, beyond the assumption of data-points spaced equally in time and a symmetric eclipse shape. However, practicioners have long been aware of the unreliability in the algorithm's error estimates, being typically considered as too optimistic [2,3,4]. During the analysis of highly precise eclipse time-series from the TESS space mission, KvW's equation for the timing error estimate went however from unreliable to becoming unsolvable, which motivated the modification to the error estimate described here.

The KvW method, in brief, assumes a light curve of an eclipse of N equidistant points separated by Δt, in which a given data point at time T_1 represents a preliminary time of minimum. Using T_1 as reflection axis, the differences in magnitudes or fluxes between the paired points Δm_k (k=1...n) are taken, and their squared sum is calculated: $S(T_1) \equiv \sum (\Delta m_k)^2$. The symmetry axis is then shifted to $(T_1 - \frac{1}{2}\Delta t)$ and $(T_1 + \frac{1}{2}\Delta t)$, and the corresponding sums $S(T_1 - \frac{1}{2}\Delta t)$ and $S(T_1 + \frac{1}{2}\Delta t)$ are calculated while keeping the number of pairings, n, the same in all reflections. The values of S against time are then fit by a parabola of the form

$$S_{fit}(T) = a\,T^2 + bT + c \,. \tag{1}$$

This parabola has a minimum value of $S_{fit}(T_0)$ given by

$$S_{fit}(T_0) = c - b^2/4a \tag{2}$$

at the time

$$T_0 = -b/2a \, , \tag{3}$$

which is the sought-after time of minimum. For the error of the minimum time, σ_{T0}, KvW give the following equation:

$$\sigma^2_{T0} = 4ac - b^2 / (4\,a^2\,(Z-1)) \, , \tag{4}$$

where Z is the maximum number of independent flux pairings, with Z=N/2 in the case of equidistant points. We note that the original KvW uses just three reflections for the calculation of *S*; later implementations may also use five or seven reflections spaced by further $\frac{1}{2}\Delta t$-steps away from T_1; we call them 3, 5 and 7-fold implementations of the KvW.
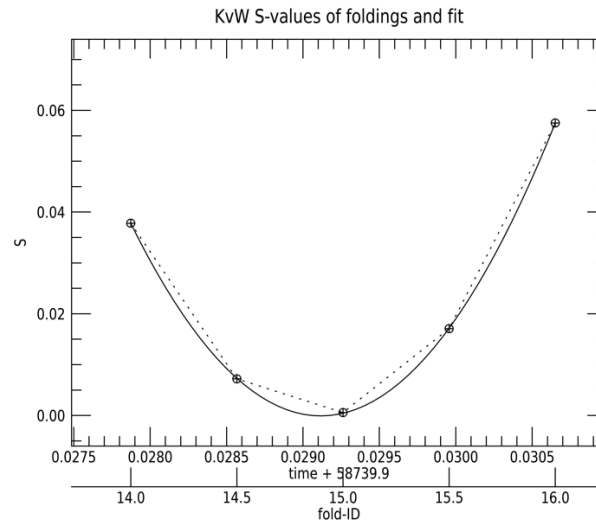

Figure 1. S values (sum of squared differences between fluxes in pairings) for the first complete primary eclipse of CM Dra observed by TESS. The upper X axis shows the time in units of BJD-2400000 and the lower one gives the enumeration ('fold-ID') of the flux-values on which (or between which) the folding was performed. The round points are S values from five foldings at the given fold-ID, while the solid line is the second-order polynomial that is fitted through these five points. We note that the minimum of the fitted curve is slightly below zero.

2. Identification of the problem

The failure of Eq. (4) became apparent when we employed the original 3 or 5-fold KvW to determine T_0 on individual eclipses of the well characterized M4-M4 binary CM Dra in very precise light curves from the TESS space mission [6]. Individual TESS light curves have lengths of about 28 days, and therefore contain 17-19 primary and secondary eclipses of CM Dra, which has a period of 1.268 days, with primary and secondary eclipses of rather similar depths of 47.5% and 44.5%, respectively.

In several individual eclipses, a computational error arose when attempting to solve Eq. 4, caused by a negative value of the term *4ac - b²*. This condition of *4ac - b² < 0* is equivalent to the minimum value of S_{fit} (Eq. 2) becoming negative (see also Fig. 1).

While the derivation of σ_{T0} from Eq. 4 was plainly failing in some cases, we also note that σ_{T0} may approach zero - and potentially be underestimated by orders of magnitudes - whenever a positive term *4ac -b²* approaches zero.

Tests were then performed with TESS light curves with artificially added noise. In these cases, the minimum values $S_{fit}(T_0)$ increase and the numerical problems in the determination of σ_{T0} vanish. Hence, problems in the determination of σ_{T0}, as well as serious underestimations of σ_{T0} from very small values of *4ac - b* have been a consequence of the significant increase in photometric precision in the more than 60 years since the algorithm's publication.

3. A revised determination of the timing error

Considering the average noise of the flux-measurements to be $\mu = <\mu_i>$, with μ_i being the observational error of an individual measurement, the average noise in the difference of two fluxes $\Delta m_k = m_{+k} - m_{-k}$ is given by $\mu\sqrt{2}$.

Instead of using the sum of the squared residuals $S(T_j)$, with T_j being the epoch of the reflection axis, we may as well calculate the usual $\mathcal{X}^2$ statistical values for these pairings, which are then

$$\mathcal{X}^2(T_j) = S(T_j) / 2\mu^2 . \tag{5}$$

If we assume that the flux from the observed binary is perfectly symmetric around the minimum time T_0, then the observed value of $\mathcal{X}^2(T)$ at the time of the minimum $T=T_0$ is only due to the average noise μ. Hence, $\mathcal{X}^2(T_0)$ would be determined by the number of degrees of freedom Z, and be given by:

$$\mathcal{X}^2(T_0) = Z - 1 . \tag{6}$$

Following Eq. 5, the equivalent expression for the value $S(T_0)$ is then:

$$S(T_0) = (Z - 1)\, 2\mu^2 . \tag{7}$$

We note that KvW's Eq. 13 also indicates that $S(T_0)$ *should* have this value, except for a factor Z/(Z-1) that is close to unity.

In our modified KvW algorithm, after an initial determination of the coefficients *a, b* and *c* as usual, we therefore offset $S(T_0)$ so that $S(T_0)$ is *defined* by Eq. (7). Using Eq. (2), the coefficient *c* is then recalculated so that $S(T_0)$ complies with Eq. (7), with *c* now given by:

$$c = (Z-1)\, 2\mu^2 + b^2/4a . \tag{8}$$

The average flux error μ should preferentially be supplied as an external parameter, from a measurement of a time-series' noise outside of the eclipses. If this is not possible, as an alternative we may assume that the lowest $S(T)$ obtained from foldings on or between data points is dominated by the noise μ, while contributions to that $S(T)$ from remaining imperfections in the folding's symmetry are relatively small. μ can then be derived from an inversion of eq. (7). In practical cases, if at least two data points are in an eclipse's central flat part, this method leads to values of μ that are within 50% of a μ measured from the off-eclipse noise.

Inserting the revised value of *c* from Eq. (8) into KvW's original error determination, Eq. (4) simplifies to:

$$\sigma_{T0}^2 = 2\mu^2 / a . \tag{9}$$

We note that a conversion to $\mathcal{X}^2$ statistics by dividing Eq. (9) with $2\mu^2$ leads to $\sigma_{T0} = 1/\sqrt{a}$. This corresponds to the usual definition of the 1-σ region of confidence for one free parameter, where a quadratic function describing $\mathcal{X}^2(T)$ increases at $T = T_0 \pm \sigma_{T0}$ by 1 over the minimum value $\mathcal{X}^2(T_0)$.

4. Code implementation of the Kwee van Woerden method with improved error estimate

A code kvw.pro has been written in the IDL language that implements the KvW method with the timing error estimate as described above (see the *supplementary materials* for a link to the code). It is not overly complex and is heavily commented, which should facilitate its implementation in other languages. The code provides also several further improvements over a basic implementation of the KvW method, which are itemized here as follows:

- Test for equidistance of input flux points: Similar to the original KvW method, the code requires data points that are equally spaced in time. The code tests if variations in temporal spacing that larger then 1% of the median spacing occur, and if so, does halt further processing. If this is considered too stringent, the rejection value can be modified. The currently code does not facilize the processing of unequally spaced data. If needed, input data to kvw.pro should be converted into equidistant flux-

points through prior linear (as proposed in the original paper by KvW) or higher-order interpolations or fits.

- Selection of data points: While the user has to take care that an input time-series contains only data taken during an eclipse (usually requiring some minimum flux-drop against the off-eclipse flux, see Fig. 2), asymmetric data-coverage around the center of an eclipse is recognized by the code. The code always selects the maximum available number of data points that are available for pairings on either branch of an eclipse, and hence balances the coverage of in- and of egress (Fig. 3).

- Employment of more than 3 foldings around the initial minimum time estimate: The number of foldings needs to be odd and the use of 5 or 7 foldings is recommended.

- For the initial minimum time estimate, the algorithm uses by default the central point of the supplied light curve, but the user may also select to use the point with the lowest flux. The central value is the better choice unless there are considerable asymmetries in the eclipse light curve. The point of least flux should only be used in low-noise data, when this point is well defined against the noise of the curve.

- Symmetrizing the fit to $S(T)$: If the lowest value of $S(T)$ does not correspond to a folding that is close to the initial estimate of the minimum time, an asymmetric fitted curve S_{fit} will be obtained, with the longer branch either to the left or right of the lowest $S(T)$ having a larger weight onto the coefficients a, b and c (see also Fig. 4). To avoid this, the outer values of $S(T)$ in the longer branch are cropped, so that this branch is at most one point larger than the shorter one. The fit for S_{fit} is then performed on the reduced set of values $S(T)$. This cropping can only be performed if $S(T)$ has been obtained at more than 3 foldings.

- The code permits also a determination of the timing error using the original procedure of KvW 1956, which does not require an explicit determination of the noise of the flux values.

- Optionally, graphical output similar to Figs. 1, 3 and 4 can be produced, which may provide useful diagnostics in the revision of the timing measurements.
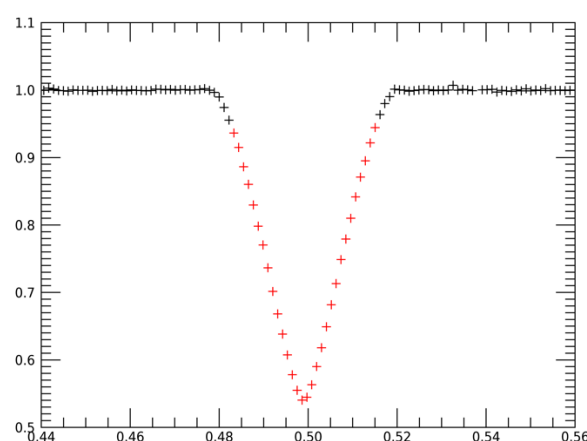

Figure 2. Example of an eclipse of CM Dra that had been extracted from TESS data towards a minimum time analysis. Shown is the first eclipse that is completely covered in TESS mission data, which is a secondary eclipse at BJD 2458739.2936. The x-axis is the fractional orbital phase and the y-axis is in flux values that are normalized to the off-eclipse flux (from fitting to the section of the light

curve that is shown in this figure). Points shown in red have a flux of < 0.95 and were used as input into the KvW algorithm.
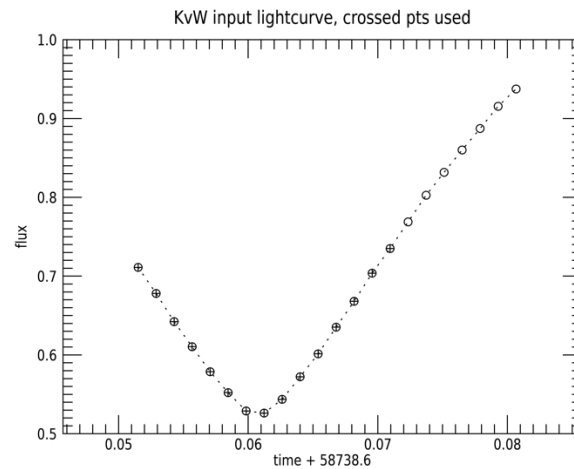

Figure 3. Input light curve to the KvW algorithm of the first – but incomplete – CM Dra eclipse in TESS data, which is a primary eclipse at BJD 2458738.6607. The y axis is in normalized flux units and the x axis in BJD-2400000. The code selects the maximum amount of data-points that can paired for foldings (filled circles) and ignores the others (open circles). Numbering the points by counting from zero, the eclipse minimum is between points 6 and 7.
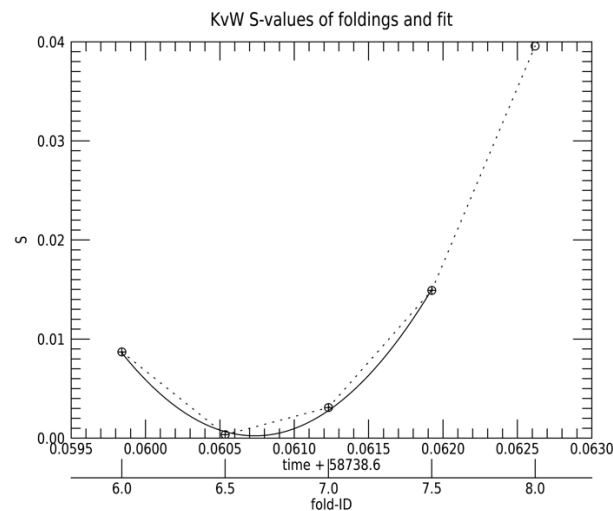

Figure 4. Similar to Figure 2, but showing the *S*-values from the incomplete eclipse of Figure 3. In this case, the distribution of *S*-values is asymmetric around the lowest *S*-value at fold-ID 6.5. The code then rejects the right-most folding (open circle) and performs a fit only on the four remaining S values.

5. Example application to TESS data and verification of the error estimates

In the following we show the application of the modified KvW to the aforementioned TESS data of CM Dra. The analyzed data are the first ones obtained by TESS on CM Dra, in its Sector 16, which was acquired from 2019 Nov. 11 to 2019 Oct. 7. Besides Sector 16, CM Dra was also observed in TESS sectors 19 and 22 to 26; their analysis is however beyond the scope of this publication. The light curve was downloaded from NASA's MAST archive, and had been processed with its pipeline version spoc-4.0.28. (We note that light-curves available on MAST until Spring 2020, which had been processed by versions prior to spoc-4.0.26, have time-stamps that are too large by 2 seconds [7]). From this dataset we used the 'PDCSAP_FLUX' values, which are fluxes that underwent a 'Pre-search Data

Conditioning' procedure to remove common instrumental effects [8]. Around each individual eclipse, a time-series covering about 3 times the eclipse-duration of 0.050d was extracted; see Fig. 2 for an example of an extracted section. For each eclipse snippet, a second order polynomial was then fitted to the off-eclipse sections before and after the eclipse. The fluxes were then divided by the polynomial fit, resulting in an eclipse light curve whose off-eclipse flux is normalized to 1, and which is free of gradients and other signals on time-scales larger than a few hours, be they from CM Dra or from instrumental effects (see again Fig. 2).

The flux error μ was determined from the off-eclipse point-to-point *rms* of these normalized light curves. Among individual eclipses, it varied between $0.98 \cdot 10^{-3}$ and $2.45 \cdot 10^{-3}$ in normalized flux units. Since the higher of these *rms* values were dominated by individual flux-peaks in the off-eclipse data, for the further analysis we used a noise that was averaged over all eclipses, which was $\mu = 1.38 \cdot 10^{-3}$ or 1380 *ppm*. The modified KvW was then executed separately on each of the primary and secondary eclipses, using only data-points with relative fluxes of less than 0.95 (red points in Fig. 2).

Resulting minimum times of all complete eclipses with their errors are shown in Table 1, while Fig. 5 shows their O-C values against the ephemeris given by [5] (We note that in [5], the epochs of reference are given in BJD_TAI. These have been converted to BJD_TDB by adding 32.184 seconds).

Based on Table 1, the average of the individual timing errors is $1.25 \cdot 10^{-5}$ d for primary minima, and $1.34 \cdot 10^{-5}$ d for secondaries (or 1.08 and 1.16 sec), with individual errors varying only very slightly around these averages.

An independent estimate of the timing error can be obtained from the standard deviation of the O-C values against the mean O-C value (dashed lines in Fig 5). This value is $1.18 \cdot 10^{-5}$ d for primary and $1.48 \cdot 10^{-5}$ d for secondary eclipses. This is in very good agreement to above mentioned errors, and shows that the timing errors from individual eclipse timings are reliable.

A further independent verification can be obtained from equations for the expected timing precision in eclipse or transit light curves [9]. That work provides several equivalent formulae for the timing error, based on a light curve's noise, and on the eclipse depth and duration. Here we use their Eq. 7, given as:

$$\sigma_{T0} \ = \mu \, T_{\nabla} \, / \, (2 \, \Delta F \, \sqrt{n_{\nabla}}) \tag{10}$$

where ΔF the relative depth of the eclipse, T_{∇} is the combined in- and egress duration, and n_{∇} is the number of data-points within T_{∇}. For the photometric noise we use again $\mu = 1.38 \cdot 10^{-3}$ and for ΔF the above mentioned eclipse depths of 0.475 and 0.445. The duration we determine as $T_{\nabla} = 0.050$d, which corresponds to $n_{\nabla} = 36$ data points, given the 2-minute cadence of TESS. With Eq. (10) we obtain then timing errors of $1.21 \cdot 10^{-5}$ d for primary and $1.29 \cdot 10^{-5}$ d for secondary eclipses – which is again in excellent agreement with the results obtained by the modified KvW.
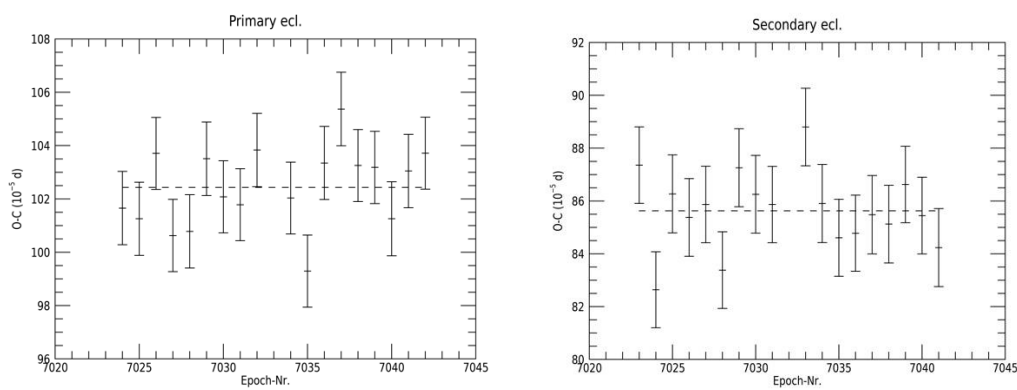
Figure 5. Observed minus Calculated (O-C) eclipse minimum times for CM Dra primary and secondary eclipses in TESS sector 26 data. The observed times correspond to Table 1 and the calculated times and epoch numbers are from the ephemeris of [5]. The dashed line is the average O-C value of these eclipeses.

Table 1. CM Dra eclipse minimum times with errors, from TESS Sector 26 observations. Epoch numbers are relative to BJD 49830.757712 for primary and BJD 2549831.390742 for secondary eclipses (converted from [5] to BJD-TDB).

Epoch -Nr.	T_0	σT_0
	(BJD-TBD)	(day)
primary eclipses		
7024	58739.9291169	0.0000125
7025	58741.1975015	0.0000125
7026	58742.4659011	0.0000126
7027	58743.7342864	0.0000126
7028	58745.0026702	0.0000125
7029	58746.2710774	0.0000125
7030	58747.5394452	0.0000126
7031	58748.8078623	0.0000126
7032	58750.0762635	0.0000125
7033	58752.6130101	0.0000126
7034	58753.8813997	0.0000125
7035	58755.1498107	0.0000125
7036	58756.4182122	0.0000125
7037	58757.6865806	0.0000127
7038	58758.9549907	0.0000126
7039	58760.2233508	0.0000126
7040	58761.4917508	0.0000124
7041	58762.7601436	0.0000126
secondary eclipses		
7023	58739.2935944	0.0000134
7024	58740.5619511	0.0000136
7025	58741.8303857	0.0000134
7026	58743.0987637	0.0000133
7027	58744.3671428	0.0000135
7028	58745.6355326	0.0000135
7029	58746.9039510	0.0000133
7030	58748.1723260	0.0000133
7031	58749.4407154	0.0000136
7032	58751.9775280	0.0000133
7033	58753.2458785	0.0000133
7034	58754.5142489	0.0000136
7035	58755.7826651	0.0000134
7036	58757.0510572	0.0000134
7037	58758.3194374	0.0000134
7038	58759.5878263	0.0000135
7039	58760.8562353	0.0000135
7040	58762.1246012	0.0000133

References

1. Kwee K.K.; van Woerden, H. A method for computing accurately the epoch of minimum of an eclipsing variable. Bull. Astron. Inst. Netherlands **1956**, *12*, 327–330
2. Mikulášek, Z.; Wolf, M.; Zejda, M.; Pecharová, P., On Methods for the Light Curves Extrema Determination. Ap&SS **2006**, 304, 363-365

3. Mikulasek, Z. et al. Kwee-van Woerden method: To use or not to use. Contr. Astr. Obs. Skalnaté Pleso **2014**, *43*, 382–387

4. Breinhorst, R. A.; Pfleiderer, J.; Reinhardt, M.; Karimie, M. T. On the determination of minimum times of light curves. A&A **1973**, 22, 239-245

5. Deeg, H. J.; Ocaña B.; Kozhevnikov V. P.; Charbonneau D.; O'Donovan F. T.; Doyle L.R.. Extrasolar planet detection by binary stellar eclipse timing: evidence for a third body around CM Draconis, A&A **2008**, 480, 563–571

6. Ricker, G.R. et al. Transiting Exoplanet Survey Satellite (TESS). J. of Astron. Telescopes, Instruments, and Systems, **2015**, 1, id. 014003

7. Memo to TESS Data Release Note 29. Available online: https://archive.stsci.edu/missions/tess/doc/tess_drn/tess_s21_dr29_data_product_revision_memo_v02.pdf

8. Tenenbaum, P; Jenkins J. TESS Science Data Products Description Document, EXP-TESS-ARC-ICD-0014 Rev D, NASA Technical Report, **2018**

9. Deeg, H.J; Tingley B. TEE, an estimator for the precision of eclipse and transit minimum times. A&A **2017**, 599, A93

Supplementary Materials: The code described in this paper is available at https://github.com/hdeeg/KvW

Conflicts of Interest: The author declares no conflict of interest.