Period, epoch and prediction errors of ephemeris from continuous sets of timing measurements

H. J. Deeg

Instituto de Astrofísica de Canarias, C. Via Lactea S/N, E-38200 La Laguna, Tenerife, Spain
Universidad de La Laguna, Dept. de Astrofísica, E-38200 La Laguna, Tenerife, Spain
E-mail: hdeeg@iac.es

Received 20 November 2014; accepted 14 March 2015

ABSTRACT

Space missions such as Kepler and CoRoT have led to large numbers of eclipse or transit measurements in nearly continuous time series. This paper shows how to obtain the period error in such measurements from a basic linear least-squares fit, and how to correctly derive the timing error in the prediction of future transit or eclipse events. Assuming strict periodicity, a formula for the period error of such time series is derived: $\sigma_P = \sigma_T (12/1)(N^3 - N)^{1/2}$, where $\sigma_P$ is the period error; $\sigma_T$ the timing error of a single measurement and $N$ the number of measurements. Relative to the iterative method for period error estimation by Mighell & Plavchan (2013), this much simpler formula leads to smaller period errors, whose correctness has been verified through simulations. For the prediction of times of future periodic events, the usual linear ephemeris where eclipse errors are quoted for the first time measurement, are prone to overestimation of the error of that prediction. This may be avoided by a correction for the duration of the time series. An alternative is the derivation of ephemerides whose reference epoch and epoch error are given for the centre of the time series. For long continuous or near-continuous time series whose acquisition is completed, such central epochs should be the preferred way for the quotation of linear ephemerides. While this work was motivated from the analysis of eclipse timing measures in space-based light curves, it should be applicable to any other problem with an uninterrupted sequence of discrete timings for which the determination of a zero point, of a constant period and of the associated errors is needed.

Key words. <Ephemerides – Time – Occultations – Techniques: photometric – Methods: data analysis – binaries: eclipsing>

1. Motivation and objectives

The space missions MOST, Kepler and CoRoT have been dedicated to the acquisition of near-continuous photometry over longer time scales. From them, timing measurements of eclipse or transit events have become available of a different nature from those from ground-based campaigns. Their main difference is the completeness of coverage between the first and the last measurement, with duty cycles of about 90% (Michel 2013, for completeness of coverage between the first and the last measurement) of such time series. From them, timing measurements of eclipse or transit events (or of ‘O-C’ residuals derived against a preliminary ephemeris) is the correct way for a determination of a period and its error, leading to a very simple equation to estimate the period error. The period error in a linear ephemeris is elaborated in the second part (Sect. 3). This error, if quoted as usual for the first timing measurement in a dataset, is shown to be a non-optimum description of the zero-point error in a linear ephemeris, which may lead to overestimated prediction errors. Correct ways to estimate the prediction error of future events are then given.

2. Derivation of the period error

Our objective is the estimation of the error of the period $P$ given a continuous sequence of $N$ timing measurements $T_E$ at integer Epochs $E$, with $E = 0, \ldots, N - 1$, and assuming that the period to be measured is intrinsically constant (e.g. $P$ does not vary with $E$). It is also assumed that all timing measurements have an identical time error $\sigma_T$. A linear ephemeris given by

$$T_{c,E} = P \cdot E + T_{c,0},$$

$T_{c,0}$ being the time of zero-epoch, can then be derived from minimizing the residuals $T_{c,E} - T_{c,E}$. We note that $T_{c,E}$ corresponds to the commonly used (O-C) or ‘observed - calculated’ residuals.

The $\chi^2$ minimization to determine best-fit parameters for $P$ and $T_{c,0}$ is then given by the linear regression:

$$\chi^2 = \sum_{E=0}^{N-1} \frac{(T_E - (P \cdot E + T_{c,0}))^2}{\sigma_T^2}. \quad \quad (2)$$

The least squares estimate of the slope $b$ of a linear fit $y = a + bx$ to data-tuples $(x_i, y_i)$ can be found in many basic works on
statistics (e.g. Kenney & Keeping 1962; Press et al. 1992) and is given by:

\[ b = \frac{N \sum x_i y_j - \sum x_i \sum y_j}{N \sum x_i^2 - (\sum x_i)^2} \]  

(3)

where we use summations over \( i = 0, \ldots, N-1 \) and changing to the nomenclature of \( a \to T_i, b \to P, y_j \to T_j \) and for the convenience of writing, replacing \( E \) by \( i \), we obtain:

\[ P = \frac{N \sum T_i - \sum i \sum T_i}{N \sum i^2 - (\sum i)^2} \]  

(4)

It is of note that the \( T_i \) in Eq. 4 may be either the measured times themselves (quoted for example in BJD) or O-C' residuals against some other (preliminary) ephemeris. In that case, Eq. 4 delivers the difference to the period of that ephemeris.

Making use of the identities \( \sum_{i=0}^{N-1} i = N(N-1)/2 \) and \( \sum_{i=0}^{N-1} i^2 = N(N-1)(2N-1)/6 \), we find, after some basic algebra:

\[ P = \frac{12 \sum T_i (i - N)}{N^3 - N} = c_N \sum T_i w_i : c_N = 12/(N^3 - N) \]  

(5)

The terms \( w_i = i - N/2 \) acts as weighting coefficients for the timings \( T_i \), the highest weight being given for the timings at the beginning and end of a dataset, and with little or no weight for those near the centre.

From above equation for \( P = P(T_0, \ldots, T_{N-1}) \), we can then derive the period error \( \sigma_P \) using error propagation; e.g. \( \sigma_P^2 = \sum (\sigma_T^2 w_i^2) \), which leads immediately to:

\[ \sigma_P^2 = c_N^2 \sum w_i^2 \]  

(6)

Using again the identities for \( \sum_{i=0}^{N-1} i \) and \( \sum_{i=0}^{N-1} i^2 \), we find that \( \sum w_i^2 = 1/c_N \) and arrive at the final result:

\[ \sigma_P^2 = c_N \sigma_T^2 \frac{12}{N^3 - N} ; N \geq 2 \]  

(7)

2.1. Comparison with the period errors of Mighell & Plavchan

In the following, period error estimates from Eq. 7 are compared to similar ones given by M&P. They use an iterative algorithm, denominated ‘Period Error Calculator (PEC)’ that obtains the period error through multiple combinations of the errors of the manifold 2-point measurements that are present within a series of timing measurements. For \( N = 2 \) to 8 timing measurements, M&P quote explicit values \( c_N \) that correspond to the \( c_N \) coefficients of Eq. 5 or 7 given above. A comparison of these values is shown in Table 1. Their results agree with Eq. 7 only for the case of \( N = 2 \) or \( N = 3 \). Up to \( N = 8 \), differences remain small within \( \approx 30\% \). Without implementing their PEC algorithm, we can also compare with their example of a strictly periodic variable with \( N = 171 \) timing measurements, each with an uncertainty of \( \sigma_T = 0.0104 \) days, for which they derive a period error of 23 microdays. From our Eq. 7 with \( c_{171} = 2.40 \times 10^{-6} \), we derive:

\[ \chi^2_{red} = \frac{\chi^2}{N - 2} \]

1 Usually, summations over indices going from 1 to \( N \) are assumed in Eq. 4 and also in Eq. 5. The change to indices going from 0 to \( N - 1 \) has no consequences as long as the summations go over a total of \( N \) terms. We prefer here indices starting with 0 in order to start with \( E = 0 \) in the linear ephemeris, as in Equation 4.

2 See the ‘Reduced’ values in the table 1 of M&P. They give them for \( M = 1, \ldots, 7 \) period cycles, which correspond to \( N = 2, \ldots, 8 \) timing measurements.

3 See M&P’s Figure 1 and accompanying text, which is for 170 cycles, corresponding to 171 measurements.

Table 1. Comparison between \( c_N \) values of this work and of Mighell & Plavchan

| \( N \)  | \( c_N \) this work | \( c_N \) M&P |
|--------|-------------------|-------------|
| 2      | 2                 | 2           |
| 3      | 0.5               | 0.5         |
| 4      | 0.2               | 0.22222     |
| 5      | 0.1               | 0.11806     |
| 6      | 0.05714           | 0.07125     |
| 7      | 0.03571           | 0.04574     |
| 8      | 0.02381           | 0.03163     |
| \dots  | \dots             | \dots       |
| 171    | 2.40×10^{-6}      | 4.89×10^{-6}|

Notes. (a) \( N \) corresponds to \( M + 1 \) in M&P. (b) Corresponding to the ‘Reduced’ square of the period error estimates from M&P’s table 1.

However, a period error of 16.1 microdays, which implies that their ‘reduced value’ for \( N = 171 \) is \( (23/161)^2 = 2.04 \) times larger than \( c_{171} \).

2.2. Simulations of the period error and application notes

Given that period error estimates from Eq. 7 deviate significantly from those of the algorithm by M&P, the results of Eq. 7 were verified by a set of simulations as follows. Assuming that an intrinsic (correct) ephemeris is given by \( T_i(E) = T_0 + E P_i \), a set of \( N \) timing measurements for \( E = 0, \ldots, N-1 \) is generated, with errors \( T_E \) that are randomly drawn from a normal distribution with a standard-deviation of \( \sigma_T \) and centred on the calculated value \( T_i(E) \). The error of each individual timing measurement is also set to be \( \sigma_T \). An example of such a simulated set of measurements in the form of an O-C diagram against the intrinsic ephemeris is shown in Figure 1. A linear ephemeris is then fitted by minimizing \( \chi^2 \) (e.g. as given by Eq. 4 or Eq. 5), which gives us the ‘observed period’ \( P_{obs} \), and the deviation against the intrinsic period, \( \Delta P = P_{obs} - P_i \). This procedure can easily be repeated many times using the same intrinsic ephemeris and a histogram of the deviations \( \Delta P \) be produced, as shown in Figure 2.

The simulations show that, for large numbers of repetitions, the mean of \( \Delta P \) becomes close to zero, and the standard deviation of \( \Delta P \) very closely approaches the value predicted by Eq. 7. Such simulations with 100 000 repetitions were performed for values of \( N = 3, 10, 100, 1000 \), all of them showing the validity of the period error given by Equation 7.

Equation 7 is a fit which is the measured times are normally distributed around the intrinsic (and unknown) linear ephemeris, and that the errors of the individual timing measurements (which are usually assigned by the observer) are of a size similar to the (O-C) residuals of the timings against this ephemeris. When this condition is not given, the reduced chi-square (using Eq. 2 and \( \chi^2_{red} = \chi^2/(N - 2) \) against the intrinsic ephemeris deviates significantly from 1. In practice, we know only the fitted ephemeris. The reduced chi-square against it should be calculated as \( \chi^2_{red} = \chi^2/(N - 2) \), accounting for the two unknown fit parameters. If \( \chi^2_{red} \) deviates substantially from 1, there is in principle no way to know if this is due to unusually large or small random errors in the measurements or if it has other origins. For the case of \( \chi^2_{red} \ll 1 \), we don’t know if measurements have been fortuitously well aligned, or if measurement errors have been overstated. For \( \chi^2_{red} \gg 1 \), measurement errors might have been un-
A simulation with such measurements being fortuitously aligned might be a valid hypothesis. Even if the fitted ephemeris in the previous section, we can derive the intercept of the linear fit, which gives the ephemeris zero epoch $T_{c,0}$ and its error.

3. Epoch and prediction errors

In a similar way to that discussed for the period and period error in the previous section, we can derive the intercept of the linear fit, which gives the ephemeris zero epoch $T_{c,0}$ and its error.

3.1. Epoch errors at the beginning of a measurement sequence

First, we continue to use the usual epoch indices ranging from 0 to $N - 1$ and start from a common equation\(^1\) for the intercept:

$$a = \frac{\sum x_i \sum y_i - \sum x_i \sum y_i}{N \sum x_i^2 - (\sum x_i)^2}. \quad (8)$$

Changing, as before, to the nomenclature of Eq. 2 and with very similar algebra, we obtain:

$$T_{c,0} = \frac{6 \sum T_i (\frac{2N-1}{3} - i)}{N^2 + N} = d_N \sum T_i v_i; \quad d_N = 6/(N^2 + N). \quad (9)$$

Again, terms $v_i = \frac{2N-1}{3} - i$ act as weighting coefficients for the timings $T_i$, with a weight that goes from about $\frac{1}{2}N$ through zero to $-\frac{1}{2}N$. For the error of $T_{c,0}$ we obtain similarly to Eq. 6

$$\sigma^2_{T_{c,0}} = \sigma^2_p = d^2_N \sigma^2_l \sum w_i^2. \quad (10)$$

Evaluating the sum of the weighting coefficients as $\sum w_i^2 = N(N+1)(2N-1)/18$, we then obtain:

$$\sigma^2_{T_{c,0}} = \frac{(4N - 2) \sigma^2_l}{N^2 + N}. \quad (11)$$

3.2. Ephemeris with zero epoch at the centre of the measurement sequence

In the following, a zero epoch at the centre of the measurement sequence is considered. For simplicity, only the case with an odd measurement sequence is considered. For simplicity, only the case with an odd

Article number, page 3 of 5
number of measurements is elaborated. The indices (epochs) of the measurements are now labelled \( j \) and will be \( j = -k, \ldots, k \) with \( k = \frac{1}{2}(N - 1) \). For the period, we start again from the basic Eqs. 5 or rather 4. For the summations, going now from \(-k\) to \(+k\), we employ the identities \( \sum_{j=-k}^{k} j = 0 \) and \( \sum_{j=-k}^{k} j^2 = N(N + 1)(N - 1)/12 \). The value for the period then turns out as:

\[
P = \frac{12}{N^2 - N} \sum T_j j.
\]

(12)

Given that \( j = i - \frac{N-1}{2} \), this is identical to Equation 5. The equation for the period error is then of course also identical to Equation 7. For the intercept, or zero-epoch, and starting from Eq. 8, however, we obtain a different and much simpler expression:

\[ T_{c,0} = \frac{1}{N} \sum T_i. \]

(13)

This ‘centre’ or ‘middle epoch’ has been labelled \( T_{c,0} \) in order to distinguish it from the usual zero epoch, \( T_{c,0} \), at the first measurement. The corresponding error is given by:

\[ \sigma^2_{T,c,0} = \frac{\sigma^2_T}{N}. \]

(14)

The different outcome for the epoch error, depending on its location at the beginning or in the centre of a measurement sequence, can be explained in the following way. From the epoch error at the centre of a measurement sequence \( \sigma_{T,c,0} \) and from the period error \( \sigma_p \), we can calculate the expected timing error at the beginning (or at the end) of the sequence, \( \sigma_{T_{begin}} \), using the square-sum of errors:

\[ \sigma^2_{T_{begin}} = \sigma^2_{T,c,0} + \left( \frac{N-1}{2} \sigma_p \right)^2, \]

(15)

where \( \frac{N-1}{2} \) is the number of periods between the sequence’s beginning and its centre. Inserting the expressions for \( \sigma_{T,c,0} \) and \( \sigma_p \) into Eq. 15 we find that \( \sigma_{T_{begin}} = \sigma_{T,c,0} \) with \( \sigma_{T,c,0} \) being given by Equation 6. This means that the error of the zero epoch of an ephemeris quoted in the conventional way, with \( E = 0 \) corresponding to the first timed event, is really the error-sum of the ‘true’ epoch error \( \sigma_{T,c,0} \) in the middle of the sequence plus a contribution from the period error.

3.3. Consequences for the prediction of events beyond the measurement sequence

A fundamental function of an ephemeris is the prediction of transit or eclipse events beyond the end of a measurement sequence, giving both the time and the time uncertainty of such future events. The usually quoted \( \sigma_{T,c,0} \) and \( \sigma_p \) may, however, easily lead to an overestimation of this timing error if a naive error-sum given by \( \sigma^2_{T,c,E} = \sigma^2_{T,c,0} + (E \sigma_p)^2 \) is used, as is illustrated by the dashed slopes in Figure 4. There are two solutions to circumvent such overestimation of the ‘prediction error’. The first solution, \( \sigma_{T,c,0} \), can be retrieved from the conventionally quoted epoch errors by reversing Eq. 15

\[ \sigma^2_{T,c,0} = \sigma^2_{T,c,E} - \frac{(N-1)^2}{2} \sigma_p^2. \]

(16)

As a side-note to Eq. 14 the calculation of \( \sigma_{T,c,0} \), for existing ephemerides may also serve as a diagnostics on the correct sizing of these ephemeris errors since \( \sigma_{T,c,0} \) can easily be related to the size of the errors of individual timing measurements through Eq. 14 and, of course, \( \sigma_{T,c,0} \) needs to come out as a positive number – or else \( \sigma_{T,c,0} \) is underestimated and/or \( \sigma_p \) is overestimated.

4. Conclusions

In the first part of this communication, a simple formula for the derivation of period errors in continuous sequences of timing measurements with identical timing errors has been derived and verified through a set of simulations. For the extraction of a linear ephemeris from a set of timing measurements, which implies that an intrinsic constant period is assumed, there is no apparent reason to use other methods than a linear fit based on an error minimization. There is no reason to use another method for the estimation of the fit-parameters errors beyond an error propagation from the equations that determine the fit parameters (in this case, from Eqs 5 and 8). In the case of identical measurement errors in a continuous series of data, equations for these
errors simplify to those given in Sect. 2 for the period (Eq. [7]) and Sect. 3 for the epoch (Eqs [11] and [14]). The only point open to variation is the type of error minimization used in the linear fit, where other methods beyond chi-square minimization might be considered, such as minimizing absolute errors; and/or robust fits that reject outliers. These may lead to slightly different error-estimates, all of them, however, are based on the residuals against the best linear best fit, independently on how that fit was obtained.

It is not the aim of this communication to revise or analyse the ‘PEC’ algorithm of M&P, which derives the period error from a combination of timing errors between any pairing of two time measurements within a sequence. Certainly, the algorithm by M&P is a vastly more complicated way to derive the period error. Differences between the error estimates from Eq. [7] of this work and M&P’s ‘PEC’ – which increase with the number of timing measures – are probably caused by an incorrect weighting of the individual 2-point timing measurements from which PEC constructs its final result. Verifying this would, however, need a detailed analysis of PEC, which is beyond of the scope of this study.

Both this work and the one by M&P assume identical errors for all individual timing measurements. In practice, even in space missions such as CoRoT and Kepler, imperfect duty cycles cause occasional misses or incomplete transits. Furthermore, cosmic-ray hits may degrade light curves of individual transits, leading to larger timing errors. For practical applications of the equations presented here, occasional measurements with strongly deviating timing errors (or missed measurements) can be ignored as long as a predominant timing error can be identified. If such a predominant error cannot be identified (e.g. owing to a change of integration time in Kepler light curves), or if a significant fraction of timing measurements is missing, the simplified equations presented here will not be reliable. The ephemeris and its errors should then be derived from a numerical least-squares minimization of Eq. [2] using individual timing errors (e.g. Press et al., 1992 sect. 14.2).

In the second part of this article, two equations for the epoch error of continuous timing sequences are derived. In the first of these equations, the error is given for a ‘zero’ epoch that corresponds to the first timing measurement (Eq. [11]). This is the conventional way in which ephemerides are indicated. In the second equation, a much simpler expression is obtained for the epoch error at the centre of a timing sequence (Eq. [14]). It is shown that these errors are equivalent, the epoch error at the beginning (or end) of a timing sequence being the error sum of the central epoch error plus the period error. Ephemerides of long sequences of timing measurements would therefore be more logically expressed with epochs and epoch errors for a timing measurement at or near the sequence’s centre. With such a ‘central ephemeris’, the estimation of timing errors beyond the end of the original measurement sequence will then be correctly performed by a simple error sum between epoch error and the period error. The ‘central ephemeris’ serves also to correctly estimate the uncertainties of the true (intrinsic) event times during the measurement sequence. With conventional ephemerides, on the other hand, a correction for the duration of the measurement sequence is needed in order to derive correct timing errors for predictions past the measurement sequence. For the photometric follow-up of CoRoT planets and planet candidates, Eq. [18] has been implemented for several years in an online calculator. There, the numbers N of observed transit events during the CoRoT pointings are estimated from the target periods and the pointing durations, ranging from 28d to 159d. Its predictions of future transit events have been shown to be reliable for CoRoT’s planet candidate verification programme (described initially in Deeg et al. 2009), as well as for an ongoing re-observation of CoRoT planet transits (Klagyivik et al., in prep.). The use of Eq. [18] over naive error-sums of epoch and period errors (counting the epochs since the beginning of the measurements) is still more important when ephemerides of targets from the Kepler mission are considered, since in that case the difference between ‘naive’ and correct prediction errors is much larger, due to the 3.9 yr coverage of its light curves.

A successful re-observation of transits or eclipses of objects discovered by space missions such as CoRoT or Kepler depends critically on the correct prediction of their transit time and timing errors. Such predictions are essential when only a few hours are available to observe a given transit or eclipse. A best possible derivation of the ephemeris errors and the prediction errors is therefore very important for the legacies of the space missions that have brought us these wonderful datasets of long, continuous and highly precise time series. Precise ephemeris measurements may also be expected to have a similar impact on follow-up observations of future planet detection missions, namely TESS (Ricker et al. 2014) and PLATO (Rauer et al. 2014).

Acknowledgements. The author acknowledges support through grant AYA2012-39346-C02-02 of the Spanish Ministerio de Economía y Competitividad (MINECO). He also thanks the anonymous referee for his comments which led to an improvement in the presentation of this paper.

References
Deeg, H. J., Gillon, M., Shporer, A., et al. 2009, A&A, 506, 343
Kenney, J. J. F. & Kepping, E. S. 1962, in Mathematics of Statistics, Pt. 1, 3rd ed (Princeton, NJ: Van Nostrand), 252–285
Michel, E. 2013, in Advances in Solid State Physics, Vol. 31, Stellar Pulsations: Impact of New Instrumentation and New Insights, ed. J. C. Suárez, R. Garrido, L. A. Balona, & J. Christensen-Dalsgaard, 145
Mighell, K. J. & Plavchan, P. 2013, AJ, 145, 148
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical recipes in C, The art of scientific computing (Cambridge: University Press, c1992, 2nd ed.)
Rauer, H., Catala, C., Aerts, C., et al. 2014, Experimental Astronomy, 38, 249
Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9143, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 20

5 http://www.iac.es/proyecto/corot/followup/ (Access will be provided upon request to the author)