1983MNRAS.203..917D

Mon. Not. R. astr. Soc. (1983) **203**, 917–924

A period-finding method for sparse randomly spaced observations or "How long is a piece of string?"

M. M. Dworetsky *Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT*

Received 1982 August 17; in original form 1982 March 26

Summary. A string-length method for establishing the period of a variable star from a relatively small number of randomly spaced observations over a long span of time is investigated. Criteria for establishing the validity of indicated periods are presented. The method is particularly suited for determination of periods in the limiting case of relatively few observations of reasonably high accuracy. A revised period and orbital elements for a spectroscopic binary observed by Abt & Levy (17 Lyr) are given. The period given by Abt & Levy for 18 Com is not confirmed; the data are insufficient to determine the correct period.

1 Introduction

As Petrie (1962b) pointed out, there was at the time he wrote no systematic method by which one could unambiguously determine the correct period of a spectroscopic binary from randomly spaced observations taken many periods apart. The surest technique involves 'saturation' observing but this is undesirable mainly because it is expensive in terms of telescope time. Another problem arises because astronomers preferentially observed objects near meridian passage, so that alias periods can be introduced and one may choose incorrectly

$$P^{-1}_{\text{false}} = 1.0027 \text{ day} \pm P^{-1}_{\text{true}} \tag{1}$$

(Tanner 1948); the problem is especially acute for systems with $P < 30$ days. Equation (1) is the most common form of a more general expression.

Several schemes for searching for periods in data have been proposed. In addition to Fourier transform techniques (e.g., Gray & Desikachary 1973; Fahlman & Ulrych 1982) other techniques can be grouped under the heading of least-squares methods. Most of these are based on the concept of minimizing the sum of the squares of the differences in ordinate between one data point and the next (or the mean of several adjacent data points) after the points have been ordered in phase for a given test period. The trial period yielding the smallest sum is customarily taken to be the correct period (with suitable precautions which are not always observed). The earliest published application of this method in astronomical work was that of Lafler & Kinman (1965). Variants and elaborations of this method have also been described by Morbey (1978), Stellingwerf (1978) and Renson (1978). The latter two methods take observational errors into account, and Renson in particular attempts to

© Royal Astronomical Society • Provided by the NASA Astrophysics Data System

1983MNRAS.203..917D

deal with problems which might be caused by accidental large phase differences between adjacent points. In general all such methods give good results as the data base grows larger and larger. Stellingwerf's method is particularly useful if one is also searching for secondary frequencies of low amplitude in variable stars.

Burke, Rolland & Boy (1970) used a method which is a true 'string-length' technique. In this method the quantity to be minimized is simply the sum of the lengths of line segments joining successive points (m_i, ϕ_i) in a phase diagram. The period chosen is that for which

$$\sum_{i=1}^{n-1} [(m_i - m_{i-1})^2 + (\phi_i - \phi_{i-1})^2]^{1/2} + [(m_1 - m_n)^2 + (\phi_1 - \phi_n + 1)^2]^{1/2} \tag{2}$$

is a minimum, with n the number of observations. Burke, Rolland & Boy (1970) did not provide any discussion of the properties of this method, a situation remedied in this paper. The straightforward conceptual basis of the method and its ability to determine periods on the basis of a minimal amount of data are found to be powerful recommendations in favour of its use in many situations.

2 Some mathematical properties of the string-length method

Renson (1978) has criticised the fact that this method introduces an inhomogeneous expression in equation (2) because the measured ordinate m_i has different units (e.g., km s^{-1}) from the phases (unit = trial period). This objection can be dealt with if we first scale the observations into a suitable range. Otherwise, in the limits of large m_i the string-length method approaches the Lafler–Kinman technique, while for very small m_i the string-length approaches unity

$$\left(= \sum_{i=1}^{n} \Delta\phi_i\right).$$

What is needed is a scale factor for the m_i which places equal emphasis on the two types of terms in equation (2); in this way one finds a minimum string-length only when the observations are arranged in phase order in a way which produces a smooth-looking variation.

The Δm's and $\Delta\phi$'s will have equal importance in equation (2) when $|dm/d\phi| = 1$. This cannot be true over the entire range of phases except for a variation consisting of linear rises and falls of equal duration, in which case we should scale the measured ordinates m_i so that $m'_{\max} - m'_{\min} = 0.5$. For a Keplerian variation with $m' = \frac{1}{4}\,[e\cos\omega + \cos(\omega + v)]$ it is easy to show that the mean absolute slope $\langle |dm/d\phi| \rangle = 1$. This is also true for any variation consisting of linear rises and falls. Therefore it seems wisest to scale the observations so that

$$m'_i = (m_i - m_{\min})/2(m_{\max} - m_{\min}) - 0.25 \tag{3}$$

in order to give equal weight to measures and phases in equation (2).

It is useful to calculate the expected string-lengths for several types of variation. The length L of a continuous sinusoidal string ($m = \frac{1}{4}\sin 2\pi\phi$) is given by

$$L = \int_0^L dL = \int_0^1 \left(1 + \frac{\pi^2}{4}\cos^2 2\pi\phi\right)^{1/2} d\phi. \tag{4a}$$

Let $\theta = 2\pi\phi$. Then

$$L = \sqrt{\frac{4+\pi^2}{\pi^2}} \int_0^{\pi/2} \left(1 - \frac{\pi^2}{4+\pi^2}\sin^2\theta\right)^{1/2} d\theta = 1.18548\, E\left(\frac{\pi^2}{4+\pi^2}\right) \tag{4b}$$

© Royal Astronomical Society • Provided by the NASA Astrophysics Data System

1983MNRAS.203..917D

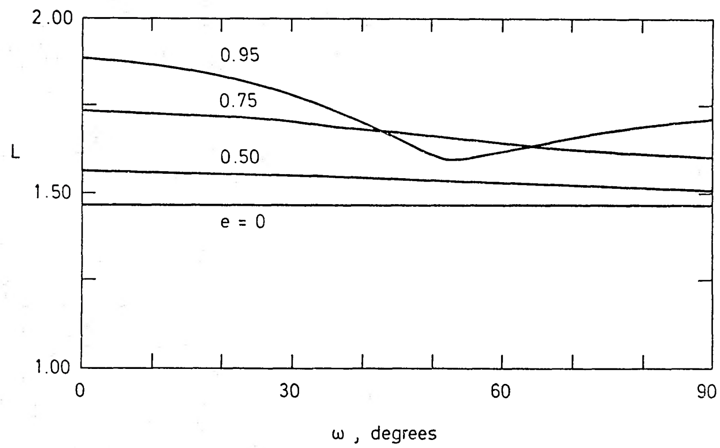

Figure 1. The string-length is a function of e and ω for a spectroscopic binary. However, even for very eccentric orbits ($e = 0.95$) the expected string-length is less than 1.9. The curves are symmetric around the $\omega = 0°$ and $\omega = 90°$ axes, so only the first quadrant is shown.

where E is a complete elliptic integral. Substitution of numerical values (Abramowitz & Stegun 1964, p. 609) gives $L = 1.4637$. Similarly, for a perfect step function $L = 2.0$, for a sawtooth $L = 1.618$, and for a variation consisting of linear rises and falls of equal duration $L = 1.414$. For velocity variations involving elliptical orbits, the string-length is a complicated function of e and ω but even for rather extreme eccentricities (up to $e = 0.95$) the maximum value of L is 1.88 (see Fig. 1). It should be noted that in Batten, Fletcher & Mann's (1978) catalogue of spectroscopic binary orbits there are no systems with $e \geqslant 0.95$. Therefore adoption of the correct period (in the absence of observational errors) should produce values of $L \simeq 1.4$ to 1.8, depending on a number of variables which we cannot predict in advance, although e and ω might be estimated crudely from the distribution of measured velocities (Schlesinger 1915).

2.1 EFFECT OF RANDOM ERRORS

The effect of normally distributed random errors on the string-length L for a sine curve was investigated by means of a Monte Carlo calculation using a random number generator in a computer program. An error parameter ϵ was introduced, scaled so that the standard deviation of the parent population could be expressed as a fraction of the total range of the ordinates. These errors were added to 'observations' made at random times on the interval of phases (0, 1). The 'observations' were then rescaled according to equation (3), and L calculated. This procedure was carried out 20 times for each combination of N_{obs} (20, 25, ... , 100) and ϵ (0.01, 0.02, ... , 0.20) and an average value of L obtained. The increase δL due to random errors was found to be given approximately by

$$\delta L \simeq 0.34\,(\epsilon - \epsilon^2/2)\,(N_{\mathrm{obs}} - \sqrt{10/\epsilon}). \tag{5}$$

For example, in a typical well-observed case we might have $\epsilon = 0.05, N_{\mathrm{obs}} = 40$, so we would anticipate that the best-fit trial period would yield a string-length of about 1.89.

For the case of a spectroscopic binary orbit, as defined above $\epsilon = \epsilon_1/2K$ rather than $\sigma_1/2(V_{\mathrm{max}}^{\mathrm{obs}} - V_{\mathrm{min}}^{\mathrm{obs}})$, where ϵ_1 is the external standard deviation of a single observation from the final least-squares solution, σ_1 is the internal standard error of an observation, and K is the semi-amplitude of the true velocity variation. Thus an exact estimate of ϵ is initially subject to some conceptual uncertainty, but this difference will usually not be a hindrance in judging the acceptability of minima in L. One might assume $\epsilon_1 \simeq 2\sigma_1$ (Petrie 1962a).

© Royal Astronomical Society • Provided by the NASA Astrophysics Data System

1983MNRAS.203..917D

These results provide some information necessary to define approximate acceptance criteria for $L_{\min}$ *before* embarking on a calculation of string-lengths for trial periods. In order to do this one also requires information about average string-lengths and their statistical distribution for the vastly greater number of cases in which completely incorrect periods have been guessed.

2.2 STRING-LENGTH CRITERIA FOR REJECTION OF FALSE PERIODS

A second Monte Carlo calculation was carried out in order to determine the expected value $\langle L \rangle$ and its standard deviation $\sigma_{\langle L \rangle}$ for circular orbits and randomly spaced observations. The 'observations' were generated at random times on an interval of 1000 days for randomly selected periods between 10 and 20 days and calculations identical to those used in making period searches were then performed, with $N_{\mathrm{obs}} = 20$ to 56 (more extensive tests would have used inordinate amounts of computer time). The frequency interval searched was 0.0 to 0.1 day^{-1}, so the true period was always included. Uniform steps in frequency $\Delta f = 10^{-4}$ were taken so that the maximum phase error for any pair of observations would always be less than 0.1 (for most pairs the error is of course much less). The results indicated that

$$\langle L \rangle \simeq 0.212\, N_{\mathrm{obs}} \tag{6a}$$

and

$$\sigma_{\langle L \rangle} \simeq N_{\mathrm{obs}}/37.5. \tag{6b}$$

Equations (5) and (6) can be combined to deduce the probability p of accidentally finding a wrong period for given values of N_{obs} and ϵ. (Since the true period was included when equation 6 was derived, p will be conservatively estimated.) The results shown in Table 1 are for sinusoidal variations, where n, the expected number of values of L small enough to be mistaken for true periods, is given as a function of N_{obs} and ϵ for a calculation with 10^4 independent trial periods. Some important points are immediately obvious. If the observations are very accurate compared with their total range, one *might* be able to deduce the correct period uniquely with as few as 15 observations. Many more observations are needed if the data are of low relative precision. It must be remembered that the expected number n of false periods is *in addition to* the correct period, which will also appear as a minimum in the string-length calculation. It appears that an attempt to find a period from as few as 20 good observations is almost always justified, if they have been taken at random times.

The major problem here is that observations are never, in practice, randomly distributed. For example, astronomers are usually assigned blocks of several nights with long intervals between blocks. Also a star is most conveniently placed when it is near the observer's meridian. All of these non-random effects will introduce other periodicities into the data: one sidereal day, one synodic lunar month, one calendar year and others. For these reasons the minimum required N_{obs} usually must be increased by an amount which is difficult to define before the true period has been determined.

Table 1. Expected number of false periods for randomly spaced observations and 10^4 trial periods.*

N_{obs} \ ϵ	0.00	0.05	0.10	0.15	0.20
15	0.1	0.1	0.5	2.5	10
20	1.0 (−3)	2.4 (−3)	2.1 (−2)	0.16	1.0
25	4.0 (−5)	1.8 (−4)	2.4 (−3)	2.6 (−2)	0.2
30	4.5 (−6)	3.1 (−5)	4.9 (−4)	0.7 (−2)	6.6 (−2)

*Sinusoidal case only. 1.0 (−3) = 1.0×10^{-3}.

© Royal Astronomical Society • Provided by the NASA Astrophysics Data System

1983MNRAS.203..917D

2.3 COMPUTATIONAL PROCEDURE FOR STRING-LENGTH CALCULATIONS

The observations are first scaled according to equation (3) and the average date T_0 of the observations calculated. Several thousand trial periods are then tested with equal frequency steps $\Delta f = \Delta\phi/(T_{\max} - T_{\min})$. This procedure assures that phase errors are always less than $\Delta\phi$, which is usually chosen as 0.1. Values of L are calculated for each trial period according to equation (2) after the data have been ordered in phase for that trial period. Numerical errors are minimized by referring all phases to T_0. An estimate of ϵ is supplied and values of L less than $1.6 + 1.2\,\delta L$ are flagged, as are values less than $\langle L \rangle - 4\sigma_{\langle L \rangle}$ (equation 6) provided that this is greater than the limit for L. The string-lengths may then either be tabulated (with printed flags) or plotted. The smallest minimum in L is taken as the most likely period, provided its depth is reasonable when compared with the criteria already mentioned (equation 5). Some judgement is required if more than one statistically significant period is found; the results may require the interpretation that further observations are needed. If no period is found, the search should be extended to other frequency intervals. The computing time is approximately linearly proportional to N_{obs}.

3 Application to cases in the literature

In order to test the usefulness of the string-length method two published examples of spectroscopic binaries were analysed which initially had incorrect periods assigned.

3.1 HD 1826

Tidy (1940) published an orbit for this star based on 44 radial velocities observed over a span of 1566 days. His solution gave $P_1 = 3.28325$ days, and a nearly circular orbit. Tanner (1949) revised the period and added four more velocities: $P_2 = 1.43233$ days; again the orbit was nearly circular. The two periods are related by equation (1):

$$P_1^{-1} = 1.0027 - P_2^{-1}$$

with P_2 undoubtedly the correct period. From equation (5) and the published errors, $L \simeq 1.96$ is expected. For the original 44 observations the string-length calculation found a deep minimum ($L = 1.92$) at the correct period P_2, and a second weaker minimum ($L = 2.16$) near the alias period P_1. Several other values somewhat less than $\langle L \rangle - 4\sigma_{\langle L \rangle}$ were found but none of them approached $L \simeq 2$. These minima were caused by non-random spacings in the data due to spurts of observing followed by long periods of inactivity.

This method was successful in identifying the true period in this classic case and also correctly identified the alias as being less significant. No other significant periods were found.

3.2 HD 192276

Hube (1976) published an orbit for this star with $P_1 = 5.7797$ days based on only 17 velocities, of which four were obtained on nights immediately following other observations. Hube & Lowe (1980) later showed that the original period was wrong and used 51 velocities to derive $P_2 = 7.18584$ days. The two periods are related by

$$P_1^{-1} = (29.53)^{-1} + P_2^{-1}$$

where 29.53 days is the lunar synodic period.

© Royal Astronomical Society • Provided by the NASA Astrophysics Data System

1983MNRAS.203..917D

For the original 17 observations, the string-length calculation yields a single convincing minimum ($L = 1.71$) at $P_1 = 5.7798$ days, but also produced a series of larger marginally significant minima ($L \simeq 1.9 - 2.2$) at and around P_2. These secondary minima correspond to the correct period, and its aliases due to annual spacings and gaps of about four years between sets of observations.

The interpretation of these results, with the benefit of hindsight, is that P_1 is the most likely period (based on the original 17 observations only), but that further observations were required to check whether P_2 might be more correct. In this case, the string-length method yielded an ambiguous result, due to the very small number of observations. The suspicious lunar frequency difference between the two strongest minima would properly have given rise to doubts about which was the correct period. Hube used Tanner's method to check for one-day aliases; however, this method is incapable of detecting lunar alias frequencies.

4 Revised orbit for 17 Lyr

In their extensive study of binary statistics among solar-type stars, Abt & Levy (1976) published radial velocities and orbital elements for several apparently variable objects. Inspection of the published velocity curves and doubts expressed by the authors led me to test the periods they found for some of their stars (those with large velocity ranges only) with the following results:

HR 3991: The period published (28.098 day) is probably correct but further confirmation is needed as there are some plausible alias periods.

HR 4753 (18 Com): The period published (17.954 day) is not correct; additional observations are required as the period cannot be determined uniquely.

HR 7261 (17 Lyr): The period published (49.09 day) is not correct and a revision of the orbit is given in Table 2.

HR 8034 (1 Equ): The period published is confirmed.

HR 8566 (37 Peg): The period published is confirmed but further observations are needed.

The new velocity curve for 17 Lyr is shown in Fig. 2. The data are consistent with an assumed circular orbit. Further observations and more precise cross-correlation velocity measurements for 17 Lyr and 18 Com are needed (Pike, Lloyd & Stickland 1978).

It is worth remarking that the 19 velocities attributed to 18 Com by Abt & Biggs (1972) are actually observations of 17 Com.

Table 2. Revised orbital elements of 17 Lyr.

$\gamma = -36.62 \pm 1.18$ km s^{-1} (se)
$K = 13.06 \pm 1.58$ km s^{-1}
$T_0 =$ JD 2 441 195.33 $\pm$ 1.21 days
$P = 42.857 \pm 0.065$ days
$e \equiv 0$
ω is undefined
$a \sin i = (7.70 \pm 0.93) \times 10^6$ km
$f(m) = (9.91 \pm 3.59) \times 10^{-3} M_\odot$
Internal standard error $\epsilon_1 = \pm 4.78$ km s^{-1}
External standard error $\sigma_1 = \pm 5.09$ km s^{-1}.
Abt & Levy period = 49.09 days

© Royal Astronomical Society • Provided by the NASA Astrophysics Data System

1983MNRAS.203..917D

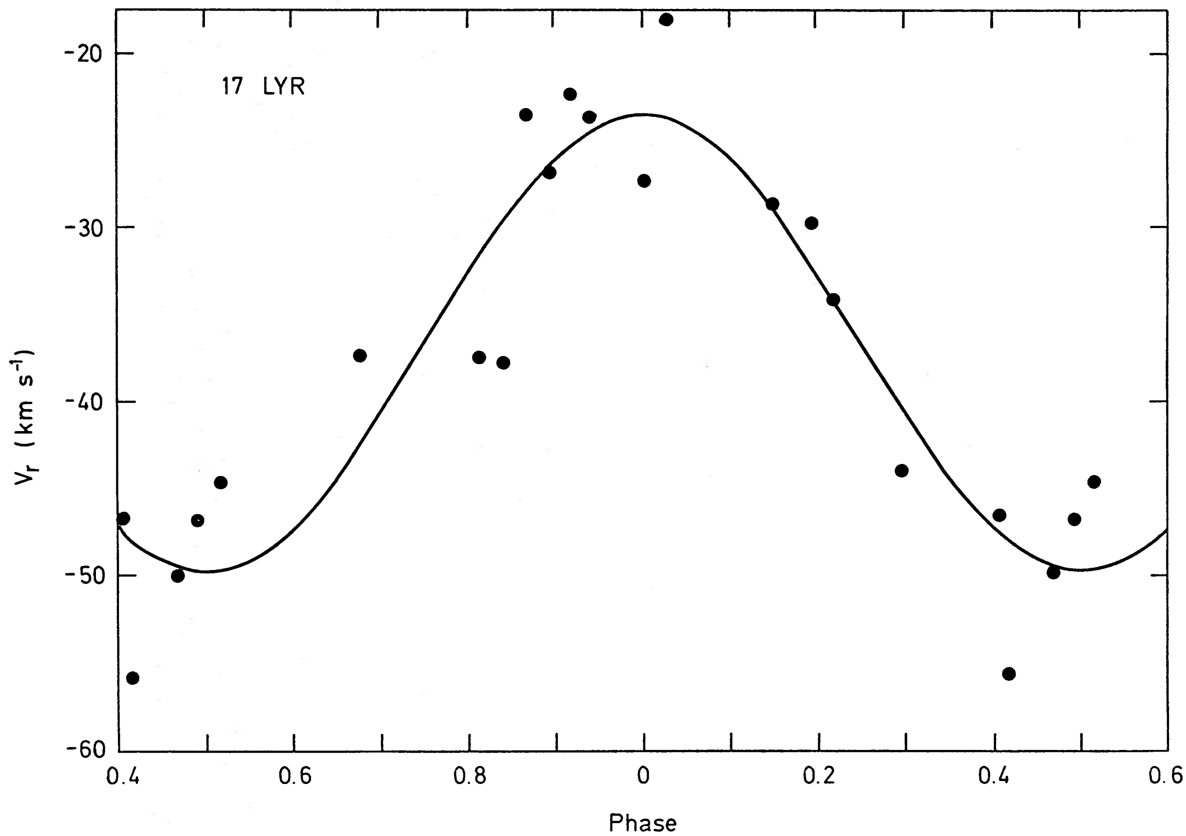

Figure 2. Velocity curve for the revised orbit of 17 Lyr from data of Abt & Levy (1976). See Table 2.

5 Conclusions

The string-length method has several advantages when used for finding periods. It is conceptually straightforward and is a natural complement to other techniques which are variants of the Lafler–Kinman method, and to Fourier transform methods. It is especially useful in the limit of a very small number ($N_{obs} \gtrsim 20$) of randomly spaced observations of periodic phenomena. Results are insensitive to the shape of the variation.

Criteria have been established for predicting the expected value L of the minimum string-length in the presence of a known amount of noise in the data. Given a sufficient amount of data the method will always present the correct period as one of the alternative minima resulting from the calculation. It is important to sample a wide range of frequency space in order to check for aliases due to the timing of the observations.

It is essential when applying this method to ensure that the frequency sampling step Δf is adequately small. The step size Δf must be chosen so that the maximum phase error $\Delta\phi$ is also small. Experience has shown that the value $\Delta\phi = 0.10$ used in this work is sufficiently small.

One practical application is in estimating the period of a spectroscopic binary at an early stage in the observing programme. This would assist the observer in planning further observations at critical times, which would distinguish between possible alternative periods and provide more uniform phase coverage, with a net increase in observing efficiency.

Acknowledgments

I an grateful to E. N. Walker for illuminating discussions of various problems connected with period-finding techniques.

© Royal Astronomical Society • Provided by the NASA Astrophysics Data System

1983MNRAS.203..917D

References

Abramowitz, M. & Stegun, I. A., 1964. *Handbook of Mathematical Functions*, U.S. National Bureau of Standards.

Abt, H. A. & Biggs, E. S., 1972. *Bibliography of Stellar Radial Velocities*, Kitt Peak National Observatory.

Abt, H. A. & Levy, S. G., 1976. *Astrophys. J. Suppl.*, **30**, 273.

Batten, A. H., Fletcher, J. M. & Mann, P. J., 1978. *Publs Dom. Astrophys. Obs.*, **15**, 121.

Burke, E. W., Rolland, W. W. & Boy, W. R., 1970. *J. R. astr. Soc. Canada*, **64**, 353.

Fahlman, G. G. & Ulrych, T. J., 1982. *Mon. Not. R. astr. Soc.*, **199**, 53.

Gray, D. F. & Desikachary, K., 1973. *Astrophys. J.*, **181**, 523.

Hube, D. P., 1976. *J. R. astr. Soc. Canada*, **70**, 27.

Hube, D. P. & Lowe, A., 1980. *J. R. astr. Soc. Canada*, **74**, 342.

Lafler, J. & Kinman, T. D., 1965. *Astrophys. J. Suppl.*, **11**, 216.

Morbey, C. L., 1978. *Publs Dom. Astrophys. Obs.*, **15**, 105.

Petrie, R. M., 1962a. In *Astronomical Techniques*, p. 63, ed. Hiltner, W. A., University of Chicago Press.

Petrie, R. M., 1962b. In *Astronomical Techniques*, p. 560, ed. Hiltner, W. A., University of Chicago Press.

Pike, C. D., Lloyd, C. & Stickland, D. J., 1978. *Acta astr.*, **28**, 235.

Renson, P., 1978. *Astr. Astrophys.*, **63**, 125.

Schlesinger, F., 1915. *Astrophys. J.*, **41**, 162.

Stellingwerf, R. F., 1978. *Astrophys. J.*, **224**, 953.

Tanner, R. W., 1948. *J. R. astr. Soc. Canada*, **42**, 177.

Tanner, R. W., 1949. *Publs David Dunlap Obs.*, **1**, 473.

Tidy, G. H., 1940. *Publs David Dunlap Obs.*, **1**, 191.

© Royal Astronomical Society • Provided by the NASA Astrophysics Data System