АСТРОФИЗИКА

ТОМ 57 ФЕВРАЛЬ, 2014 ВЫПУСК 1

UNCOVERING PERIODS IN ASTRONOMICAL TIME SERIES WITH FEW DATA

O.CARDONA, M.A.LÓPEZ-CASTILLO, M.REYES-MUÑOZ

Received 1 July 2013
Accepted 6 December 2013

A procedure for discovering periods in astronomical time series containing few observational data using simple mathematical operations is described. By selecting data close to the maxima or minima of the time series differences among these values around the maxima or minima are obtained to produce a set of intervals. Using a technique similar to the least common divisor and applying the maximum common denominator to the set of intervals approximate periods are found. The different ways to improve the periods found are presented. The procedure is applied to a simulated random sinusoidal data set and also to some data from binary and pulsating variable stars to show how the procedure is applied to different type of time series. The procedure is simple to use for any type of data spacing and with gaps and produces results in accordance with other methods.

Key words: *astronomical time series: observational data*

1. *Introduction*. Time series data are ordered sequences of measurements where successive values in the data represent consecutive measurements taken at equally or unequally spaced time intervals and with gaps. There are two main objectives of time series analysis: identifying the nature of the phenomenon represented by the sequence of observations, and predicting future values of the time series variables. Both of these objectives require that the behaviour of the observed time series data is identified and more or less formally described. Once the variation is established, one can use it in the theory of the phenomenon and we can extrapolate the identified variation to predict future events. Most time series patterns can be described in terms of two basic classes of components: trend and seasonality. The former represents a general systematic linear or nonlinear component that changes over time and does not repeat or at least does not repeat within the time range covered by the data. The latter may have a formally similar nature, however, it repeats itself in systematic intervals over time. Those two general classes of time series components may coexist in real-life data. The latter component can have periodic variations that is necessary to characterise to understand the phenomena at hand. The problem of finding periodicities in the time series of many types of observational and experimental data, and from a diversity of other phenomena have been studied in many papers in the past. There exist in the astronomical and time series

analysis a great number of methods and procedures to solve the problem of periodicities in the observations of many types of applications. Petrie [1] wrote at his time that no method exists to determine the correct period of a spectroscopy binary form observations taken many periods apart. Aitken [2] gives some references and recipes to find periods using plots of parts of the data and reversing them with respect to a fixed points to find close coincidences and the interval between two point is equal to the period. The need for precisely determining periods of cyclic phenomena is well known and numerous methods have been produced for evenly spaced data [3-5]. Lately the attention is centered in phenomena observed at irregularly spaced intervals and with gaps [6-9].

To help in the acquisition of new data in changing time series in general it is necessary to have an approximate period to select judiciously the times of further acquisition of new data, especially with few observations, in order to determine a better period. A review of several techniques for uncovering periodicities in variable and binary stars can be found in an article by [10]. A simple procedure using the correlation between the time series and the remainders of the series with respect to the tentative period for equally distributed intervals is described by [11]. There are several period search algorithms in the literature [3,12-14], Least squares methods [15, 16], String Length Statistics [17], Fourier methods [18,6], Periodogram analysis [19,8,9,20,21], Fast Fourier methods for data unevenly spaced and with gaps [7], and Spline methods [22]. Variants of these methods and of some others can be found in the articles given in the references.

In this work we present a simple procedure useful to find approximate periods in astronomical time series that complements the methods mentioned before when the number of observations is small.

2. ***Procedure***. This procedure can be used when one has few points of the time series and it is necessary to have an idea of the period in order to obtain further data to be able to get a better value of the period as mentioned before. Taking values close to the maxima (minima) of a given observed or experimental series for the purposes, in the one hand, to have few values for computational convenience, and in the other hand, to assure that the maxima are taken into account at least approximately in order to define differences among these values. With these differences, it is possible to find with a variant of the least common denominator (LCD) which will satisfy the time intervals between these observations. Any such interval between the approximate maxima is related to the period in an average way. Each pair of corresponding phases gives a relation $t_l - t_m \approx nP$, where $t_l - t_m$ is the interval, n is an integer, and P is the period. Also from those intervals using the LCD one can find submultiples of them, and with the results one obtains a value approximately common to all of the intervals that will give the estimated period. This very

approximate period is used in the next step of our procedure. The intervals found above are also used to find the common greatest divisor (CGD) between any two such intervals. With all the CGD's found before the average of them is calculated, because the points are close to but not necessarily in the maxima (minima) of the series, and that result would represent the tentative period.

The mechanics of the procedure is the following, from the given time series data one searches for the maximum (minimum) value of the amplitude and defining a small interval around the maximum (minimum) value that can include enough values to be able to find a good approximate period as was mentioned above. Then calculating the differences of the values found before starting with the first with respect to the rest of the values and then with the second one with respect to the remaining values, except the second, and so on. With these intervals it is possible to find a tentative value for the period by dividing the first set of intervals by two as many times as necessary to obtain a set of numbers and then by three, and so on. The result of those divisions shows the numbers that are similar in size to each other giving the tentative period. This is the approximate period that will be used later as the stopping parameter in the quasi Euclidean procedure used to find the CGD. From the intervals obtained from the differences between all the values with respect to the first one are obtained. The first difference is used with all the other differences in the process of finding the CGD in order to obtain a series of numbers that are used to find the mean of those number that becomes the approximate period. In the process to find the CGD to stop the process, one uses the greatest number found above, in the function for that calculation. The pseudo code of the Euclidean algorithm for integer numbers is given by the following function.

Function CGD (a, b)
While $b \neq 0$
$t := b$
$b := a \bmod b$
$a := t$
Return a

For real numbers the stoping factor is different from zero in the While statement of this function and should be chosen carefully to have the appropriate range of values around the value found before in the LCD. This procedure can be carried out by simple hand calculations.

3. *Improvement of the Period.* There are several ways to improve the tentative period found above. The first and simplest procedure that we have used is that with the value of the period found before a phase diagram is plotted to see if it represents the observations correctly and changing the period slightly in such a way that one can appreciate the changes in the diagram until one

is satisfied with the plot, for example when the minimum of the curve is close to half of the phase. This procedure is the best for finding periods when one has few values of the astronomical time series, independent of the type of operations used in the calculation of the period. The periods found are good for the purpose of choosing the subsequent times of acquisition of new data and if one has more points of the time series one can refine the period by an approximate least squares method [23] to find a better period through a simple iteration process that we propose and is described in the following section. Also in this case, using more sophisticated methods one can find a better period [3,12,14], or using Spline [22], or Period 04 [24]. Of course, one can obtain more data to improve the period but sometimes that is not possible in observational data.

4. *Examples for the use of the Procedure.* In this section some analyses are made of some time series with the purpose of showing how the procedure is applied to some numerical simulations and several real observations of some known binary and variable stars reported in the literature.

4.1. *Numerical Simulation.* The numerical simulation is made for a sinusoidal variation with a given period with random data generated using a gaussian distribution [21]. The sinusoid is given by

$$y = R\sin(wt+\varphi) \qquad (1)$$

with

$$w = \frac{2\pi}{P}, \qquad (2)$$

where P is the period, φ is the phase and R is the amplitude.

We have used a period of 2.5 days and an amplitude of 1.0 for the sinusoidal variation. Following our procedure we can recuperate the period without any problem, as we will show in this case. The data close to the maximum value are seven and are given in Table 1, the Julian Date TJ is given in days. The differences between the first value and the other six, and of the second value with the other five and so on are given in Table 2. Dividing the first six numbers

Table 1

VALUES CLOSE TO THE MAXIMA FOR SINUSOIDAL

	TJ	Amplitude
1	31.61425	0.9996352
2	16.63571	0.9996378
3	1.614091	0.9996241
4	19.13149	0.9998671
5	54.11985	0.9999163
6	59.13433	0.9997246
7	84.12874	0.9999558

Table 2

DIFFERENCES BETWEEN THE VALUES FORMING GROUPS

Differences			
1	14.97854	12	17.51740
2	30.00016	13	52.50576
3	12.48277	14	57.52024
4	22.50560	15	82.51465
5	27.52008	16	34.98837
6	52.51449	17	40.00285
7	15.02162	18	64.99725
8	2.495779	19	5.014484
9	37.48414	20	30.00889
10	42.49863	21	24.99440
11	67.49303		

Table 3

THE FIRST SIX VALUES OF Table 2 IN THE FIRST ROW, IN THE SECOND ROW THE FACTORS, AND IN THE THIRD ROW THE RESULTS

14.97854	30.00016	12.48277	22.5056	7.520008	52.51449
6	12	5	9	11	21
2.49642	2.500013	2.496554	2.500622	2.501825	2.50069

of Table 2 by two and then by three, then four, five, and so on, the same could be done with the other numbers to produce a list of numbers where some of them are almost equal. One way to carry out this process for simplicity is with the first and second numbers of Table 2 that can be divided by two and then by three and so on and then the other numbers are divided with

Table 4

RESULTS OF THE CGD IN THE SECOND COLUMN AND THE SUM IN COLUMN THREE

	CGD	SUM		CGD	SUM
1	2.523373	2.523373	11	2.590164	28.10745
2	2.539366	5.062738	12	2.613091	30.72054
3	2.562292	7.625031	13	2.649681	33.37022
4	2.598892	10.22392	14	2.543236	35.91345
5	2.542722	12.76664	15	2.566154	38.47961
6	2.495779	15.26242	16	2.602753	41.08236
7	2.543236	17.80566	17	2.518705	43.60107
8	2.566154	20.37181	18	2.555319	46.15639
9	2.602745	22.97456	19	2.532393	48.68878
10	2.542723	25.51728			

those results to find the number of times they are divisible and then divide the number by those factors giving in Table 3. The greatest of these similar numbers in this case is around 2.5018. But at first sight one can see in Table 2 that the period could be close to 2.495779. Applying the maximum common multiple with the pseudo Euclidean algorithm using the value of 2.5018 found in the last step multiplied by 0.8 as the stopping parameter to examine the results up to this quantity in order to have numbers of the order of 2.5018 in this procedure, Table 4 is generated. The average value using the last value of column three of Table 4 divided by 19 is 2.562567. Plotting a phase diagram with this value for the period Fig.1 is obtained. This phase diagram shows some scatter of the points around the theoretical curve which means that the period is close to the theoretical one but must be corrected.

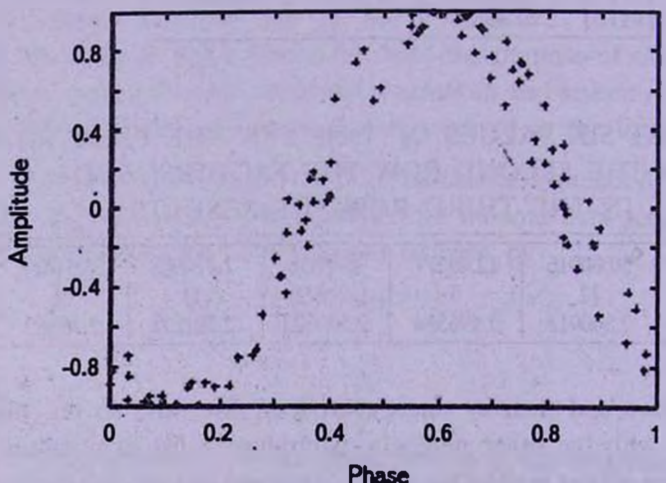

Fig.1. Phase diagram for the sinusoidal with a tentative period of 2.562567.

4.2. *Improvement of the period.* The period can be improved in several ways as was mentioned before. The method of minimum least squares in its simplest formulation [23] can be applied to find a better period starting with the tentative period found before. For a three-parameter model, following the notation of [23], given by

$$x_i = \mu + A\cos(wt_i) + B\sin(wt_i) + \varepsilon_i , \tag{3}$$

where x_i and t_i denote the i^{th} values of the observations, w is the frequency and ε_i is the residual. The approximate solutions of the equations of the estimates of least squares for the model are

$$\tilde{\mu} = \bar{x} = \frac{\sum x_i}{n}, \tag{4}$$

$$\tilde{A} = 2\sum(x_i - \bar{x})\cos(wt_i), \tag{5}$$

$$\widetilde{B} = \sum (x_i + \bar{x}) \sin (wt_i). \tag{6}$$

To find R and φ, the amplitude and phase we solve the above equations with

$$A = -R\sin(\varphi) \tag{7}$$

and

$$B = -R\cos(\varphi), \tag{8}$$

therefore

$$R = \sqrt{A^2 + B^2} \tag{9}$$

and

$$\varphi = \arctan\left(-\frac{B}{A}\right). \tag{10}$$

In these equations the frequency w is regarded as known. The method is extended to include the estimation of w following a simple iteration procedure starting with a smaller value than the approximate value found in the first part of the procedure presented in this article and defining the sum of squares of the residuals [23] as

$$e = \frac{n}{2} R^2 \tag{11}$$

to carry out the iterations over frequency for all the equations given above we use the following expression

$$w_{n+1} = w_n + e \times 10^{-3}. \tag{12}$$

The criteria for stopping the iterations is the value found for the approximate period. This iteration procedure gives good results in this case with respect to the theoretical period of 2.5.

4.3. *Analysis of Some Observations.* Some real observational time series are presented to show how the procedure to find the periods is implemented and to see how the principal steps involved in the procedure exhibit the intricacies of the mechanical flow of the operations.

4.3.1. *Binaries.* The analyses of three binary stars of different periods are presented to show the procedure for these type of time series.

The spectroscopic binary star 26 Aquilae has high orbital eccentricity where the primary component is of type G8 III-IV. There are fifty-one spectroscopic observations covering a 20 years interval. Fig.2 shows the plot of the 51 radial velocities [25]. Table 5 gives the six values close to the minima of the time series of 26 Aquilae, where the times are in days. The differences between the first value and the other five, and of the second value with the other four and so on are given in Table 6. Divide the first five numbers of Table 6 using

Table 5

VALUES CLOSE TO THE MAXIMA FOR 26 AQUILAE

	DJ	V_R		DJ	V_R
1	22951.695	-11.100	4	32036.832	-9.070
2	30908.668	-10.030	5	33372.023	-10.870
3	32015.820	-10.740	6	33397.984	-10.720

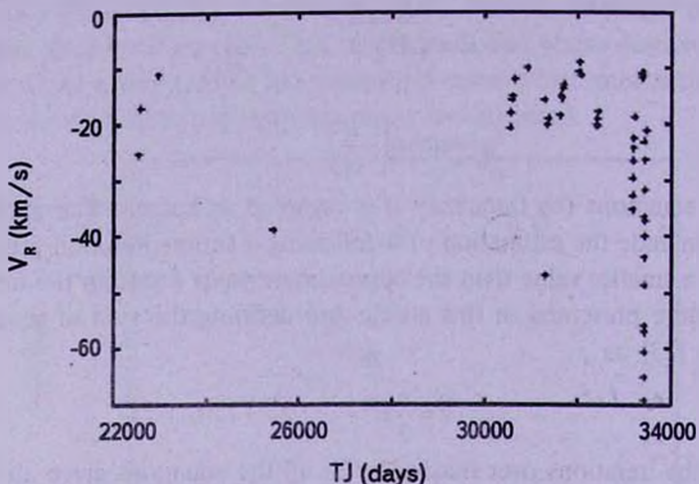

Fig.2. Observational radial velocity curve for 26 Aquilae.

the same procedure mentioned before in the sinusoidal case to produce a list of numbers where some of them are almost equal are given in Table 7. The greatest of the similar numbers in this case is around 267.85. Applying the maximum common multiple with the pseudo Euclidean algorithm using the value of 267.85 found in the last step multiplied by 0.8 as the stopping parameter to have numbers of the order of 267.85 in this procedure, Table

Table 6

DIFFERENCES BETWEEN THE VALUES FORMING GROUPS

Differences			
1	7956.973	9	2489.316
2	9064.125	10	21.01172
3	9085.137	11	1356.203
4	10420.33	12	1382.164
5	10446.29	13	1335.191
6	1107.152	14	1361.152
7	1128.164	15	25.96094
8	2463.355		

Table 7

THE FIRST FIVE VALUES OF Table 6 IN THE FIRST ROW, IN THE SECOND ROW THE FACTORS, AND IN THE THIRD ROW THE RESULTS

7956.973	9064.125	9085.133	10420.33	10446.29
30	34	34	39	39
265.23	266.592	267.21	267.19	267.85

Table 8

RESULTS OF THE CGD IN THE SECOND COLUMN AND THE SUM IN COLUMN THREE

	CGD	SUM		CGD	SUM
1	279.5273	279.5273	8	213.3477	2452.113
2	395.0742	674.6016	9	21.01172	2473.125
3	281.9219	956.5234	10	321.6719	2794.797
4	307.8828	1264.406	11	347.6328	3142.430
5	279.5273	1543.934	12	300.6602	3443.090
6	300.5391	1844.473	13	326.6211	3769.711
7	394.2930	2238.766	14	25.96094	3795.672

8 is generated. The average value using the last value of column three of Table 8 divided by 14 is 271.1194. Plotting a phase diagram with this value for the period Fig.3 is obtained. This phase diagram shows some scatter of the points with a well defined curve which means that the period is close to the real one but even so it must be corrected. The period can be improved in as was

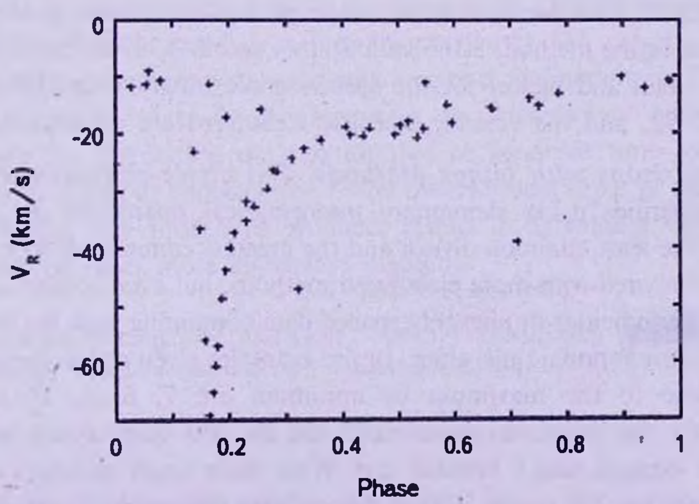

Fig.3. Phase diagram for 26 Aquilae for radial velocities with the period 271.1194.

mentioned before. With the curve fitting method one finds a period of 266.995 days and the period given by [25], and [26] is 266.544 and 266.7 by Spline.

The spectroscopic binary star HD 145425 is located in Serpens Caput with magnitude 9.5 and spectral type K0 with forty-six radial velocities observed [27]. There are eight values close to the maxima. Following the procedure, the greatest of the similar numbers in this case is around 564.813. Applying the maximum common multiple with the pseudo Euclidean algorithm using the value of 564.813 found in the last step multiplied by 0.7 as the stopping parameter to obtain numbers of the order of 564.813 in this procedure. The average value is 550.2391. The period can be improved using the curve fitting method giving 550.963 and 550.134 with Spline. The value given by [27] is 549.9.

The spectroscopic binary star HD 217792 of magnitude $V = 5.10$ and spectral type F0V has fifty-two radial velocity observations [28]. The data close to the maximum value are 14. Continuing with the procedure, the greatest of the similar numbers in this case is around 181.794. Applying the maximum common multiple with the pseudo Euclidean algorithm using the value of 181.794 found in the last step multiplied by 0.8 as the stopping parameter to have numbers of the order of 181.794 in this procedure. The average value is 177.7688. The period can be improved giving 178.053 with the curve fitting method and 178.316 with Spline. The value given by Bopp et al. is 178.3177.

4.4. *Variable Star*. This classical cepheid star BK Centaurus has 49 observations that show a beat period [29]. The data close to the maximum value are 12. The greatest of the similar numbers in this case is around 3.218. Applying the maximum common multiple with the pseudo Euclidean algorithm using the value of 3.218 found in the last step multiplied by 0.8 as the stopping parameter to have numbers of the order of 3.0 in this procedure. The average value is 3.152659. The period can be improved in several ways giving 3.218 with the curve fitting method, 3.166 with Spline, and 3.17389 by Leotta-Janin.

Results, tables and figures for the spectroscopic binary stars HD 145425 and HD 217792, and the variable star BK Centaurus are given in [30].

5. *Comparisons with Other Methods*. This simple procedure produces approximate periods using elementary mathematical operations as are the analogues of the least common divisor and the greatest common divisor, hence can not be compared with more elaborated methods, but even so one can find approximate periodicities in unevenly spaced data containing gaps for few data points of the observational time series. In the examples given above the number of points close to the maximum or minimum are 7, 6, 8, 14 and 12 respectively for the numerical experiment and for real observations of three spectroscopic binaries and a variable star. With those small numbers of data points this procedure can handle without any problem the search for periodicities, something that most of the methods mentioned in the body of the article can

not solve. Therefore, this procedure is the best for finding periods when one has few values of the astronomical time series, independent of the type of operations used in the calculation of the periods. The approximate periods can be found by hand calculations something that can not be done with most of the other methods mentioned in this work. When one has at hand more points of the time series and with the approximate periods found one can use any of the other methods to improve these tentative periods as we have done with the curve fitting procedure by the approximate least squares method. The results for the cases considered in this article compare well with the results obtained with other methods. And also when one has more data points one can consider the procedure as complement to some of the other methods because the period found could be used as the starting search for periodicities for those more elaborated methods. This procedure can be used to find other periods in the same time series eliminating the period found before from the numbers obtained in the first part of the process and then one can repeat the procedure again to search for another period, and so on for multi-periodic data.

6. *Conclusion and Commentaries.* The procedure for finding periods for few observational data uses the values close to the maxima of the observational time series to apply something like the least common divisor to find an approximation of the period that can be used in a procedure similar to the Euclid's procedure to find the greatest common divisor but with a stopping parameter different from zero that can be obtained from the approximate period found before multiplied by a small fraction to produce values close to the approximate period. As the points close to the maxima are approximations to the maxima (minima) of the series and can fall in either side of the maximum, therefore it is necessary to take an average of the values found with the CGD to obtain a value close to the true period. This value can be improved with different techniques as mentioned previously. We use the curve fitting method by least squares with a three-parameter model iteratively. The procedure produces good results for predicting in an approximate way the periods necessary for forecasting the evolution of an observed time series with the purpose of aiding in choosing the future observational times of the phenomena under study. The procedure produces results in agreement with the results produced by other more elaborated methods.

Instituto Nacional de Astrofísica, Óptica y Electrónica, México,
e-mail: ocardona@inaoep.mx   angeles@inaoep.mx   mreyes@inaoep.mx

ВОССТАНОВЛЕНИЕ ПЕРИОДОВ АСТРОНОМИЧЕСКИХ ВРЕМЕННЫХ РЯДОВ С ПОМОЩЬЮ МАЛОГО КОЛИЧЕСТВА ДАННЫХ

О.КАРДОНА, М.А.ЛОПЕЗ-КАСТИЛЛО, М.РИЕС-МУНОЗ

Описывается процедура нахождения периодов астрономических временных рядов, содержащих малое количество наблюдательных данных, с использованием простых математических операций. Выбирая данные, близкие к максимумам и минимумам временных рядов, с целью создания набора интервалов, вычислены разницы между этими значениями вокруг максимумов и минимумов. Приблизительные значения периода получены с помощью метода, похожего на метод наименьшего общего делителя, определив для полученного набора наибольший общий знаменатель. Представлены разные пути для уточнения полученных периодов. С целью иллюстрации применения метода для различных временных рядов, эта процедура применена к случайным синусоидальным данным, а также к некоторым данным двойных и пульсирующих переменных звезд. Процедура проста для использования к любым данным с разрывами и дает результаты, которые согласуются с другими методами.

Ключевые слова: *астрономические временные ряды: наблюдательные данные*

REFERENCES

1. *R.M.Petrie*, Astron. Techniques, ed. W.A.Hiltner, Chicago: Univ. of Chicago Press, 1962, p.560.
2. *R.G.Aitken*, The Binary Stars New York: Dover, 1962.
3. *J.Lafler*, *T.D.Kinman*, Astrophys. J. Suppl. Ser., **11**, 216, 1965.
4. *R.B.Blackman*, *J.W.Tukey*, The Measurement of Power Spectra New York: Dover, 1959.
5. *G.G.Fahlman*, *T.J.Urlych*, Mon. Notic. Roy. Astron. Soc., **199**, 53, 1982.
6. *D.F.Gray*, *K.Desikachary*, Astrophys. J., **181**, 2523, 1973.
7. *T.J.Deeming*, Astrophys. J. Suppl. Ser., **36**, 137, 1975.
8. *N.R.Lomb*, Astrophys. J. Suppl. Ser., **39**, 447, 1976.
9. *J.H.Scargle*, Astrophys. J., **263**, 835, 1982.
10. *A.W.Fullerton*, The Study of Variable Stars Using Small Telescopes, ed. J.B.Percy, New York: Cambridge Univ. Press, 1986, p.201.
11. *E.Whittaker*, *G.Robinson*, The Calculus of Observations 4th edition, London:

Blackie and Sons, 1944.
12. *I.Jurkevich*, Astrophys. J. Suppl. Ser., **13**, 154, 1971.
13. *R.F.Stellingwerf*, Astrophys. J., **224**, 953, 1978.
14. *H.G.Marraco*, *J.C.Mussio*, Publ. Astron. Soc. Pacif., **92**, 700, 1980.
15. *F.J.M.Barning*, Bull. Astron. Inst. Neth, **17**, 22, 1963.
16. *P.Vaniček*, Astrophys. J. Suppl. Ser., **12**, 10, 1971.
17. *M.M.Dworetsky*, Mon. Notic. Roy. Astron. Soc., **203**, 917, 1983.
18. *W.Wehlau*, *K.-C.Leung*, Astrophys. J., **139**, 843, 1964.
19. *A.Shuster*, Terrestrial Magnetism, **3**, 24, 1898.
20. *W.H.Press*, *G.B.Rybicki*, Astrophys. J., **338**, 277, 1989.
21. *W.H.Press*, *S.A.Teukolsky*, *W.T.Vetterling*, *B.P.Flannery*, Numerical Recipes New York: Cambridge Univ. Press, 1992.
22. *C.Akerlof et al.*, Astrophys. J., **436**, 787, 1994.
23. *P.Bloomfield*, Fourier Analysis of Time Series: An Introduction New York: Wiley&Sons, 1976.
24. *P.Lenz*, *M.Breger*, Comm. in Asteroseismology, **146**, 1, 2005.
25. *K.L.Franklin*, Astrophys. J., **216**, 383, 1952.
26. *R.H.Wolfe*, *H.G.Horak*, *N.W.Storer*, Modern Astrophysics Paris: Gauthier-Villars, M.Hack, Ed., 1967, p.251.
27. *R.F.Griffin*, The Observatory, **114**, 231, 1994.
28. *B.W.Bopp*, *D.S.Evans*, *J.D.Laing*, Mon. Notic. Roy. Astron. Soc., **147**, 355, 1970.
29. *C.Leotta-Janin*, Bull. Astr. Inst. Neth., **19**, 169, 1967.
30. *O.Cardona*, *M.A.López Castillo*, *M.Reyes Muñoz*. Technical Report INAOE-181013, Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, Mexico, 2012.