Medium-term Variation in Times of Minimum of Algol-type Binaries: XZ And, RZ Cas, U Cep, TW Dra, U Sge

Geoff B. Chaplin
Hokkaido, Kamikawa-gun, Biei-cho, Aza-omura Okubo-kyosei, 071-0216, Japan;geoff@geoffgallery.net

Received August 25, 2019; revised October 5, 2019; accepted October 6, 2019

Abstract We use methods of singular spectrum analysis to examine the observed minus expected times of minimum of five Algol-type binaries looking for significant periodicities. All the stars examined show long "waves" of 20 to 50 years but only XZ And and RZ Cas appear to have significant periodicities.

1. Introduction

In this paper we examine the Algol-type variables listed in Table 1 with physical data in Table 2. It is common to attempt to explain secular changes in times of minimum (of the eclipse) in terms of harmonics caused by third bodies in the system, for example Borkovits and Hegedus (1996), or additionally apsidal motion, Hoffman *et al.* (2006). A study by Li *et al.* (2018) of 542 short-period binaries showed that many such stars do indeed show the presence of third bodies. Other causes of periodicity (not necessarily harmonic) are interactions between electromagnetic fields (Hall 1989) or electromagnetic-gravitational interactions (Applegate 1992). A general review of Algol variables is given in Budding (1986) and references therein, and a table of such stars in Budding *et al.* (2004).

Our objective here is to determine if there are oscillations in the observed minus calculated (O–C) times of minimum with "periods" (i.e. non-secular changes) in ranges determined by the available length and density of data but broadly speaking in the range of a few years to 20 years—rather than the long-term variation, and without any presumption as to the cause of such variation. The upper limit of 20 years is driven by the desire to see a minimum of at least 6 or so complete periods in order to be confident of their real nature, and the lower period by the bucketing (binning) size chosen which in turn is driven by the density of data.

XZ And Demicran *et al.* (1995) examined the O–C history and postulated two or three other bodies with periods 137.5, 36.8, and 11.2 years as a possible cause of the variation, although they suggested the third period may be related to cyclic magnetic activity of the secondary star. Borkovits and Hegedus (1996) could not find a good fit for any third body orbit. Manzoori (2016) quote periods of 34.8 ± 2.4 years and 23.3 ± 3.0 years.

RZ Cas This star's variability was discovered by Muller in 1906. The light-curve sometimes shows a flat bottom near primary minimum, which may be caused by the effect of short-term δ-Scuti type variability (Rodriguez *et al.* 2004) or hot and cool spots on the surface of the primary (Olson 1982). Ohshima *et al.* (2001) and Rodriguez *et al.* (2004) found very short-term periodicity of around 22 minutes. Mkrtichian *et al.* (2018) also confirmed the high frequency variability and in addition, based on a 1999–2009 data sample, a 6- to 9-year magnetically induced variation in times of minimum. Using a statistical analysis of times of minimum, Chaplin (2018) postulated a period of 22.5 to 24 years.

U Cep Ceraski (1880) first determined U Cep to be a variable binary star with a period of 2.49 days. It is amongst the most active Algols, and it is known to cycle through periods of extremely active mass exchange and times of relative quiescence and was found to exhibit a 9 ± 4 yr cycle of relative quiescence and high activity (Hall 1975) with period changes related to mass loss and non-conservation of angular momentum.

TW Dra This star forms a visual triple with ADS 9706, complicating both amateur and professional observations. Zejda *et al.* (2008) quote cycles of approximately 20 years caused by electromagnetic-gravitational interaction and 6.5 years caused by a third body.

U Sge Manzoori and Gozaliasl (2007) used polynomial fitting and Fourier analysis of times of minimum to conclude there were cyclic variations of lengths 15.8 and 9.5 years, whereas Simon (1997) reports a roughly defined period of 39 years.

Table 1. Stars analyzed.

Name	*GSC (HD)*	*R.A. (J2000) h m s*	*Dec. (J2000)*[1] *° ′ ″*	*Data History Start*	*Period*[2] *(days)*
XZ And	02824-01360	01 53 48.76	+41 51 24.97	1891	1.35730911
RZ Cas	04317-01793	02 48 55.51	+69 38 03.44	1901	1.19525031
U Cep	04505-00519	01 02 18.44	+81 52 32.08	1880	2.4930911
TW Dra	(139319A)	15 33 51.06	+63 54 25.67	1898	2.80684701
U Sge	01607-00913	19 18 48.41	+19 36 37.72	1901	3.380619331

[1] *Wenger* et al. *(2000).* [2] *Frank and Lichtenknecker (1987).*

2. Methodology

2.1. Data

Data in the form of times of minimum (Tmin) are taken from the Lichtenknecker database (Frank and Lichtenknecker 1987), augmented in the case of RZ Cas by data from Mkrtichian *et al.* (2018). Data and the period (possibly after minor adjustment) are used to calculate observed minus expected (O–C) differences in times of minimum (expressed in days).

2.2. Data cleaning and weighting

Initial relatively isolated observations are ignored together with early very noisy data if any, and analysis is performed on the entire data series (visual and electronically based observations). Where more recent data include a long run of high quality observations (by electric photometer or CCD) we perform the analysis additionally using only the electronically based data.

It is to be noted that the accuracy of Tmin is the result of the combination of the accuracy of individual magnitude estimates and the number of such estimates (as well as a good analytical method to reduce these to a Tmin). In principle many unbiased visual observations can be as accurate as a small number of highly accurate CCD observations although in practice electronic data are far better; further theoretical discussion and analysis is given in Chaplin *et al.* (2018). While accuracy estimates are given for recent CCD results such explicit estimates are generally not available for other observation methods.

Various weighting (or unweighted) regimes were tested with relatively minor variation in results. In this paper we analyze using two regimes: weights according to the observational type: 10 (CCD), 3.25 (electric and wedge photometer), or 1 (others) ("fixed" weights), or weighted, by 1/observational-error squared ("inverse error" weighting). Observational error (for Tmin in days) is taken as given by Mkrtichian *et al.* (2018) (applicable to RZ Cas only), 0.0004 for other CCD data, 0.0004 for photometric data, 0.01 for a series of photographic data, and 0.02 for others (primarily visual). For the CCD data the error is a conservative estimate based on Samolyk (2011 and others), for photometric data it is based on a comparison of scatter compared to CCD, and for visual it is derived by assuming 0.1 accuracy in magnitude estimates and between 10 and 20 observations in the series—based on Chaplin *et al.* (2018)). Outlier rejection is performed by repeated local polynomial fitting rejecting outliers at the 4-sigma level.

2.3. Bucketing (binning)

SSA techniques require data at equally spaced time points whereas observations of Tmin are unevenly spaced. We use two alternative methods to obtain the required spacing after defining a bucket size (the time span—150-, 200-, 300-, and 500-day buckets were used). The first method ("date buckets") takes the (weighted) average of data within that interval, filling empty buckets by linear interpolation; the second method ("local poly buckets") fits a local polynomial and takes values from the fitted curve at evenly spaced time intervals. In the case of date bucketing a small bucket size means more empty buckets; where the percentage of empty buckets exceeds about one quarter the results of the analysis are ignored as unlikely to be reliable.

It should be noted that the bucketing process necessarily imposes a limit on the shortest periods that can be reliably detected. For example, with a binning of 500 days periods shorter than this are unlikely to make their presence known in the analysis (although it is theoretically possible if the period and bucketing interval are not simple multiples of each other) and detected periods are likely to be longer than a multiple of the bucketing interval.

2.4. SSA

We use techniques of Singular Spectrum Analysis ("SSA") as explained in detail by Chaplin (2018, 2019 and references therein) and implementations in the R language (R Found. 2018a), CRAN libraries (R Found. 2018b), using RStudio (2018) and in particular the "Rssa," and "simsalabim" (for significance testing, Gudmundsson 2017) libraries.

By way of background SSA can be thought of as a means of calculating averages (called EV series) from a data series—rather like (but much more complicated than) moving averages. These averages have certain properties in common with Fourier series—orthogonality between different series—although the EV series are generally neither periodic nor of constant amplitude. The R language primarily gives access to an extensive library of statistical, mathematical, graphical, and other routines, while RStudio provides an interface to the code, written output, all variable values, graphical output, libraries and help files, and more. Both are free and available for Windows, Mac and a range of Linux systems. Outline code for the analysis is given in Appendix A.

The EV series derived from the SSA analysis are ordered according to the magnitude of the associated eigen value (equivalent to ordering by the strength of the series in the total data), and we refer to them as series 1, 2, 3, etc.

Table 2. Stars' physical data, and sources.

Star	*Spectral Type*	*Masses (solar) (primary, secondary)*	*Radii (solar)*	*Source*
XZ And	A4IV-V, G5IV	3.2, 1.3	2.4, 2.6	Demircan *et al.* (1995)
RZ Cas	A3V[1], carbon[2] or K01V[3]	2.2, 0.7	1.7, 1.9	Maxted *et al.* (1994)
U Cep	B 7/8 V, G 5/8 III–IV[4]	4.2, 2.8	2.8, 4.9	Batten (1974), Singh *et al.* (1995)
TW Dra	A5 V, K0 III	2.2, 0.9	2.6 (primary)	Tkachenko *et al.* (2010)
U Sge	B7III, K1III C	5.7, 1.9	4.2, 5.3	Dobias and Plavec (1985)

[1] *Duerbeck and Hänel (1979).* [2] *Abt and Morrell (1995).* [3] *Maxted* et al. *(1994). Rodriguez* et al. *(2004).* [4] *Tupa* et al. *(2013).*

The given orbital period is adjusted to give a best fit horizontal line to the O–C data. "Sequential SSA" is performed on the bucketed data to determine the trend in the signal, and removing this trend from the O–C data gives the residuals. Trends are defined as EV series which have no identified periodicity. The residuals are then analyzed using SSA again and the reconstructed signals are calculated after grouping residuals according to similarity of pattern and having high correlation. Periodicities of the reconstructed signals are then determined. Only signals which appear under a variety of bucketing time intervals and regimes ("consistency") are considered and tested for significance.

The length of the data series determines the longest period we can reasonable expect to be confident in identifying—our rule of thumb is 6 or more periods in the time series (so 20 years for a 120-year data series). The shortest period is determined by the density of data and how short a bucket size we can take without empty buckets exceeding around 25% of the total. With, for example, a 150-day bucket interval, periods of around 3 years or longer should be apparent from the analysis of the data.

SSA is a "data driven" method of analysis to be contrasted with a "model driven" approach where (as in Hoffman *et al.* 2006), for example, a parabola or sinusoid might be used to remove the trend component of the change in O–C. The strength of the argument for using one method rather than the other depends on the evidence from other observations. If, again for example, a distant companion star is observed, then the calculated orbital period can be used to determine the sinusoidal change (light time effect) in the O–C that its orbit will cause, and this can be used to remove (a component of) the trend. The argument for favoring the data-driven approach is stronger when no such external evidence in favor of a specific long-term change is known. Results regarding any discovered shorter period signals are to some extent dependent of what trend signal has been removed—in other words a model for the trend may lead to different shorter period signals being discovered compared to the purely data-driven approach.

2.5. Significance testing

We use the MCSSA routine from the simsalabim library as a tool to assign a level of significance to the results. Anything below a 95% confidence level is rejected. However, in stating that a result is significant we qualitatively take into account more factors than a single test number—the number of periods observed in the data, changes in amplitude of the signal, etc.

3. O–C charts and data types

Table 3 shows the symbols are used consistently in the O–C charts plotting raw data to identify different equipment used to make the observation leading to the Tmin calculation.

3.1. XZ And

After rejecting 18 outliers, a total of 1,086 times of minimum from year 1891 on are displayed in Figure 1. Analysis using an adjusted period of 1.357283 days showed a trend from signals 1, 2, and 5 across all bucketing and weighting regimes, illustrated in Figure 2.

Table 3. O–C charts and data types.

Symbol	*Data Type*
Cross	Visual observation
Tilted cross	Photographic observation
Diamond	Electric photometer
Square	CCD (amateur)
Filled dots	CCD (Mkrtichian *et al.* (2018), RZ Cas only)

Residuals showed a clear and consistent pattern across bucketing and weighting regimes with signals 1 and 2 being responsible for over 90% of the residual, giving rise to a period of approximately 38 years (36–40 years) and are shown in Figure 3. MCSSA testing showed the signal as better than 99% significant. An apparent period of approximately 23 years from signals 3 and 4 is both weak and largely present only when visual observations dominate the data, and does not show as significant.

Figure 4 shows the raw data minus trend value and we note the latest CCD data present a different pattern and indicate a significantly smaller magnitude of variation, casting doubt on possible interpretation as a (single) third body effect. There are indications that earlier turning points also showed more complicated behavior.

3.2. RZ Cas

In total, 4,930 observations were used, and bucketing into 150-day or longer intervals had less than 15% empty buckets with only one empty bucket at 300- and 500-day bucketing. Figure 5 shows the data after rejecting 46 outliers at the 4 standard deviation level.

Consistently across all bucketing intervals (data or polynomial buckets) and weighting regimes, the first two signals describe the trend pattern, and after adjusting the period to 1.1952500 days Figure 6 shows the 500-day bucketed data and the trend from the first two signals.

Decomposition of the residual is again the same for all bucketing periods and data weighting methods—the first two series constitute the signal with a period of 22.7 to 25 years, which is significant to better than 99%. Weaker signals are present in higher series and show (near) harmonics at 11.5 and 5.5 years although they are not significant. The reconstituted series from these two signals together with the residuals is shown in Figure 7 using 500-day bucketing.

CCD and electric photometer data amount to 377 observations after removing two outliers, and are shown in Figure 8.

After period adjustment (to 1.1952500 days again) and removal of the trend (from the first two EV series) signal decomposition of the residual series reveals no surprises; decomposition is very similar across different bucketing intervals and repeats signals at the same periods as found from the long-term analysis. The dominant signal remains the 24-year period and Figure 9 shows the 200-day bucketed data and the primary signal. The signal remains significant at better than 98%.

We note that the long-term variation shown in Figure 7 is of the order of 0.05 day, whereas the variation relative to the long-term trend is of the order of 0.01 day while the accuracy of the

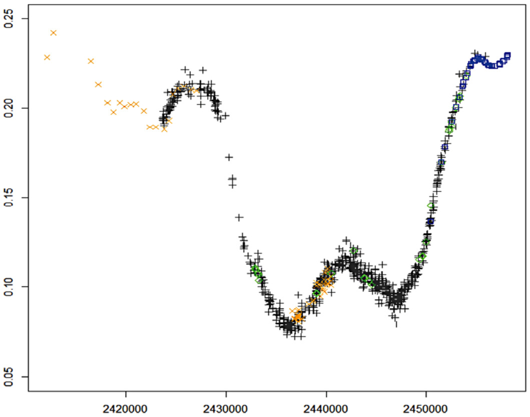

Figure 1. XZ And, O–C data (deviation in days against Julian Date). See Table 3 for key to the symbols.

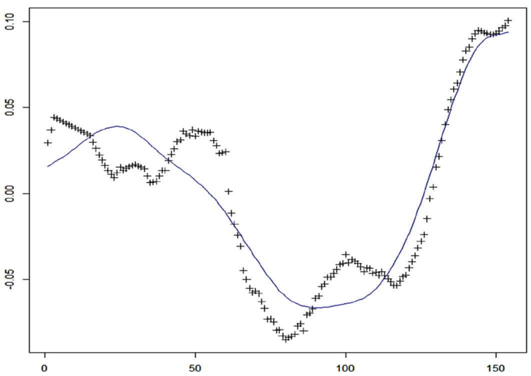

Figure 2. XZ And, 300-day bucketed data and fitted trend (O–C against bucket number).

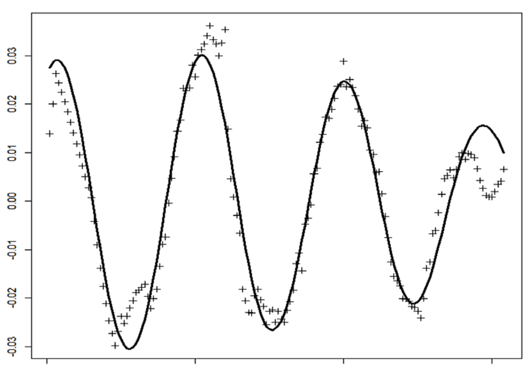

Figure 3. XZ And, residual data and reconstructed signal.

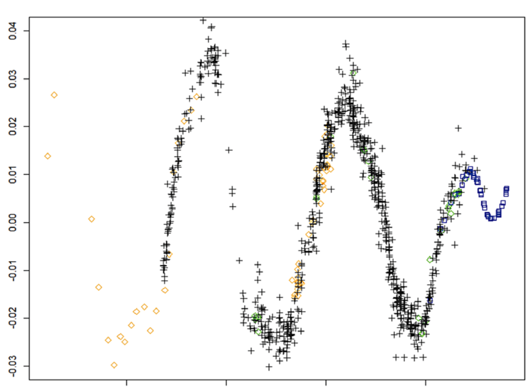

Figure 4. XZ And, raw data minus trend.

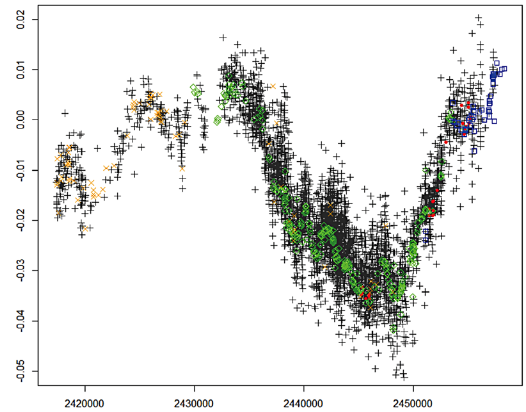

Figure 5. RZ Cas, O–C data (deviation in days against Julian Date).

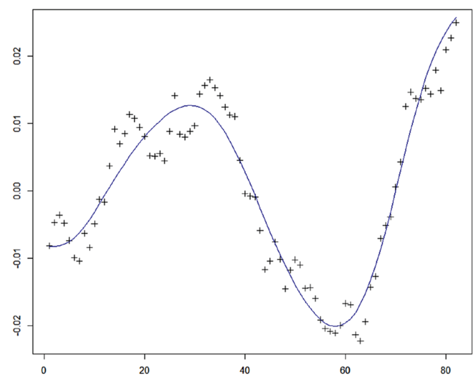

Figure 6. RZ Cas, data after period adjustment with fitted trend signal (O–C against bucket number).

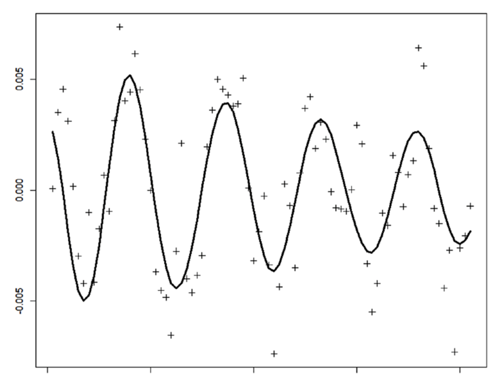

Figure 7. RZ Cas, residuals (500-day buckets) and reconstructed signal.

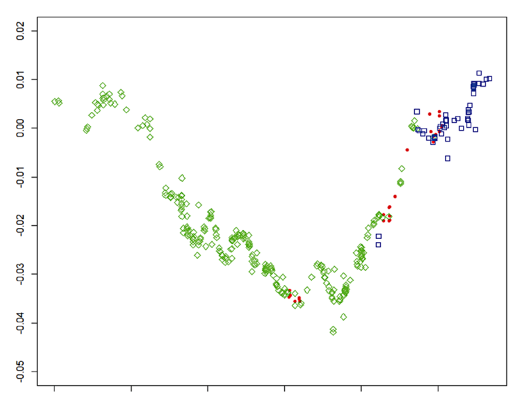

Figure 8. RZ Cas, electronic data (O–C against JD).

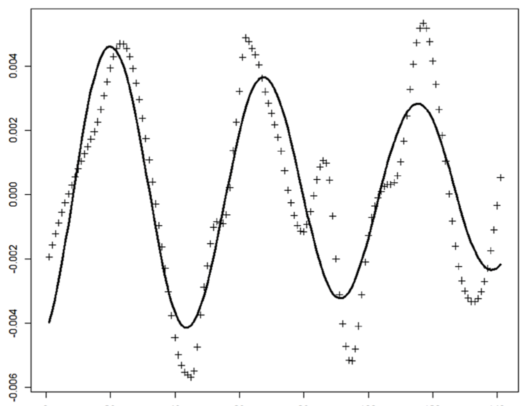

Figure 9. RZ Cas, 200-day bucketed electronic data and long-term trend (O–C vs bucket number).

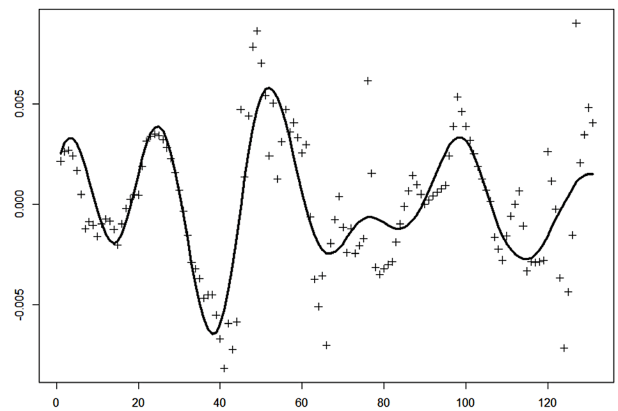

Figure 13. U Cep, 200-day bucketed data and reconstructed signal.

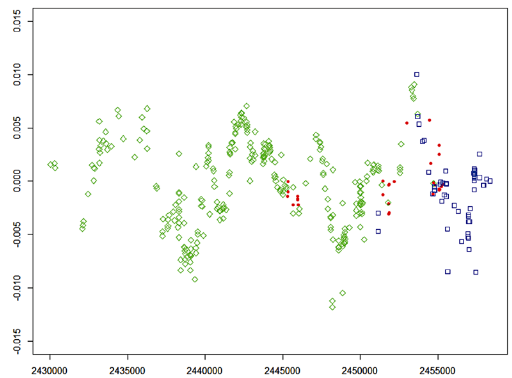

Figure 10. RZ Cas, electronic data minus the trend values.

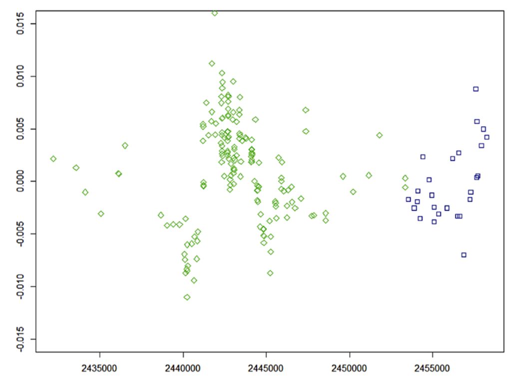

Figure 14. U Cep, electronic data minus the trend values.

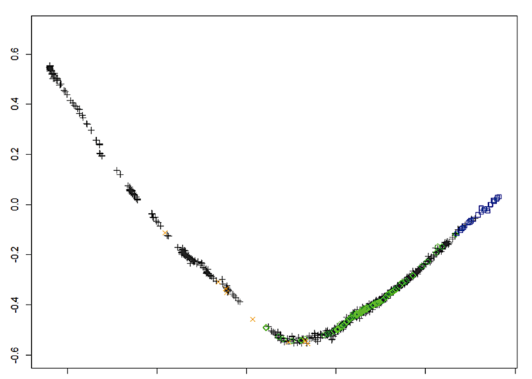

Figure 11. U Cep, O–C data (deviation in days against Julian Date).

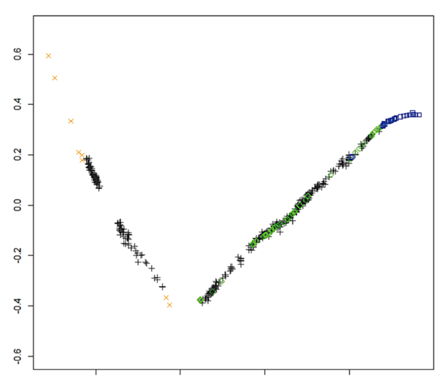

Figure 15. TW Dra, O–C data (deviation in days against Julian Date).

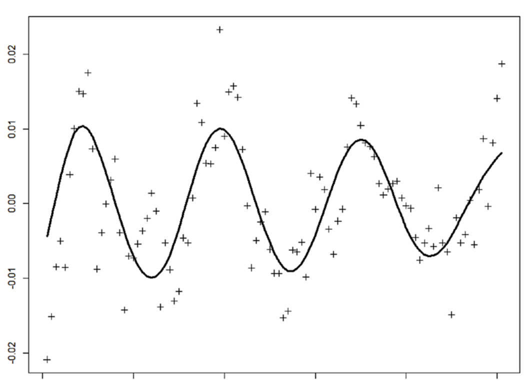

Figure 12. U Cep, residuals and reconstructed signal.

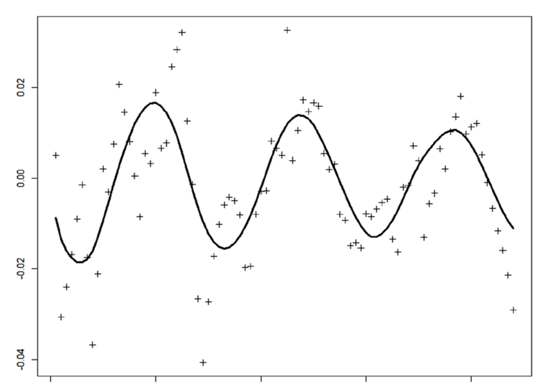

Figure 16. TW Dra, residuals and reconstructed signal.

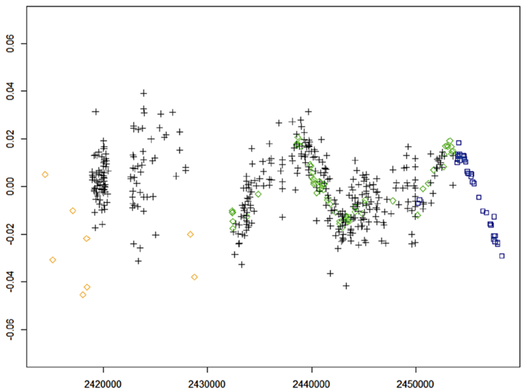

Figure 17. TW Dra, data minus trend.

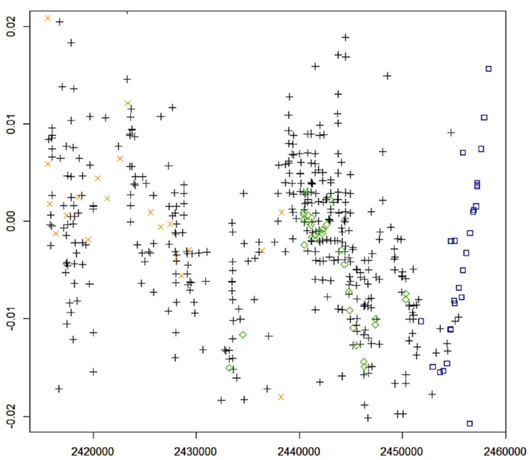

Figure 18. U Sge, data after eliminating outliers.

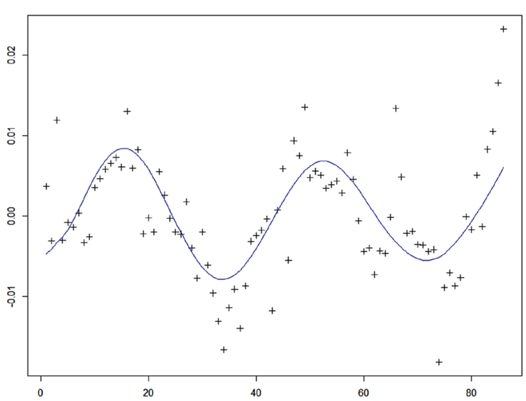

Figure 19. U Sge, 500-day bucket data and trend.

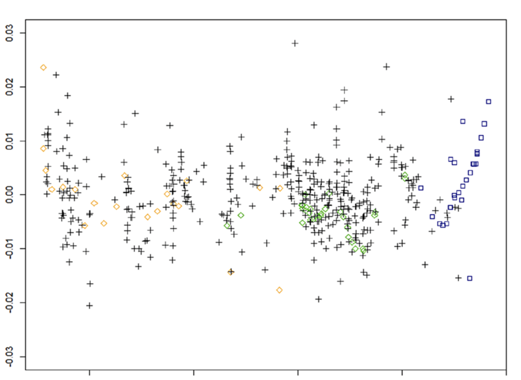

Figure 20. U Sge, data minus trend.

electronic data is generally much better than 0.001. Figure 10 shows data minus the trend values. It can be seen that RZ Cas regularly shows very short-term changes in period of the order of 0.01 day in between 3 and 6 years.

3.3. U Cep

1,242 observed minus calculated times of minimum after rejecting 37 outliers from 1,880 to 2,018 using a period of P = 2.493009 are shown in Figure 11.

Between 4 and 7 signals constitute trends (period shown as infinity), depending on bucketing method and length. After calculating residuals the reconstructed signal repeatedly shows periodicities of 11.7–12.1 years, 20.5–24 years, with a longer period of 40 years occasionally appearing—however, none of these appear to be significant. Only at the 500-day interval does the signals analysis agree for both bucketing methods (and weighting regimes). Figure 12 shows the residuals under 500-day bucketing with the strongest signal of 40 years from EV series 1 and 2.

There are 190 electronic (CCD and electric photometer) data after removal of 5 outliers. Trend and residual signal analysis is clearer and more consistent than with the entire data series, although the strongest periodicity in the residual signal is identified as 17–19 years and several other signals with periods as short as 12 years also appear, but none of these appear to be significant. Figure 13 shows the 200-day bucketing residual series together with the reconstructed signal from vectors 1 to 4.

At 200-day bucketing there are approximately 30% empty buckets, while at 500 days there are still 20% empty buckets and the data points are much sparser. Although testing indicates the period is significant (over 95%) the lack of data with the resulting interpolated points, together with the lack of a match to results found from the longer data series, indicates not much confidence can be placed in this result.

Figure 14 shows the data minus the trend values and shows the clumping of observations around the middle of the series and to a lesser extent at the end. The accuracy of the observations is considerably smaller than the range of variation so we can conclude that U Cep undergoes irregular changes in period of the order of 0.01 to 0.02 day—increasing and decreasing—taking place within a timespan of the order of 7 years.

3.4. TW Dra

After rejecting 10 outliers 534 times of minimum from year 1901 on were bucketed into 150-day or longer buckets, with at best 20% of the buckets being empty using a 500-day bucket. The data are shown in Figure 15.

Four EV series constituted tends and after adjusting the period to 2.806772 days analysis of the residuals showed periodicity 36.5–37.5 years (signals 1 and 2 accounting for 40%), see Figure 16. In addition there is a weak signal of 13 to 17 years but the amplitude of the signal is far greater during the time of visual observations and virtually disappearing during the latter interval of electronic observations. Neither of the periodicities appears to be significant.

Figure 17 shows the deviations of the observations from the trend and we see the above trend pattern repeated and period changes of the order of 0.04 day occurring over 14 years or so.

3.5. U Sge

After rejecting 16 outliers, 480 times of minimum data from year 1901 on were analyzed; the data are shown in Figure 18 using a period of 3.380619 days. The raw data show the evidence of changes in the period of the order of 0.03 day—both increasing and decreasing—taking place over a timespan varying in the range of 10 to 25 years, in particular evidenced by the recent series of CCD observations. Unlike previous examples U Sge does not show a clear secular pattern.

Initial analysis presented a confusing picture with periodicities of 42 to 68 years occurring but with no regular repetition or pattern and, unlike the previous stars, no obvious long-term pattern. After including EV signals with periods of 40 years or longer with signals having undefined periods, Figure 19 using 500-day bucketing shows a wave of roughly 48 years, but with only just over two such cycles occurring in the data series we feel it more appropriate to regard this as a trend rather than genuine periodicity.

Analysis of the residuals using 500-day bucketing produced a dominant signal of 22.5 to 24 years although this was not confirmed with shorter bucketing intervals. Periods around half this occurred both with 500-day bucketing and with shorter bucketing intervals. However, testing showed no significant signals. Figure 20 shows the data minus the trend and confirms the sharp changes of 0.02 day in 9 years or so.

4. Summary and conclusions

It should be born in mind that a data-driven method of analysis may produce periodicities different from those which would arise if a model-driven approach for removing the trend had been used.

XZ And has a secular pattern spanning about 0.02 day (change in Tmin) and shows a clear and consistent residual signal across the range of bucketing intervals, methods, and weighting regimes with a highly significant signal of approximately 38 years but with much lower amplitude in the recent cycle and evidence of more complicated behavior at turning points.

RZ Cas has a secular pattern spanning about 0.06 day and shows a fairly clear and consistent residual signal, with a significant signal of approximately 24 years repeating 5 cycles throughout the data history. Analyzed separately, electronic data reveal the same signal. RZ Cas exhibits rapid shifts in Tmin of around 0.01 day over periods of 3 to 6 years.

U Cep has a secular pattern spanning about 1.1 days does but not show well defined or significant signals of any periodicity, although a long wave of 40 years is often revealed (as are 12- and 24-year periods) and is visually present in Figure 12. Electronic data suggest a different period; they also shows changes in Tmin of the order of 0.02 day occurring over a period of 7 years or so.

TW Dra has a secular pattern spanning about 1.0 day and shows a clear and consistent breakdown of the residual signal with a dominant wave of 37 years, although this does not show as significant. Electronic data show shifts in Tmin taking place over a timespan of 14 years or so.

U Sge exhibits no long-term trend, unlike the other stars discussed here, with no clear breakdown of the signals. The only period revealed is a long wave of approximately 48 years. There is evidence of changes in the period of the order of 0.02 day taking place over a timespan varying in the range of 9 years and longer, in particular evidenced by the recent series of CCD observations.

All the above stars exhibit long waves from 24 to 48 years but with only 2 to 5 cycles appearing in the data history we can not feel completely confident that these are stable periods despite what statistical significance tests might say. The strongest candidates for such stable periods (and which show as strongly significant) are XZ And, additionally because of the clarity of the signal breakdown and the magnitude of the wave, and RZ Cas, with 5 cycles present in the data and the clarity of the breakdown of the signal.

The value of amateur CCD observations is apparent from the analysis of these stars, and continued observations with accurate reduction of the results to a time of minimum are strongly encouraged for these and other close binaries.

5. Acknowledgements

This research has made use of the Lichtenknecker-Database of the BAV, operated by the Bundesdeutsche Arbeitsgemeinschaft für Veränderliche Sterne e.V. (BAV). This paper has benefited in terms of readability and content through the constructive comments of an anonymous referee.

References

Abt, H. A., and Morrell, N. 1995, *Astrophys. J., Suppl. Ser.*, **99**, 135.

Applegate, J. H. 1992, *Astrophys. J.*, **385**, 621.

Batten, A. H. 1974, *Publ. Dom. Astrophys. Obs. Victoria*, **14**, 191.

Borkovits, T., and Hegedus, T. 1996, *Astron. Astrophys., Suppl. Ser.*, **120**, 63.

Budding, E. 1986, *Astrophys. Space Sci.*, **118**, 241.

Budding, E., Erdem, A., Çiçek, C., Bulut, I., Soydugan, F., Soydugan, E., Bakiş, V., and Demircan, O. 2004, *Astron. Astrophys.*, **417**, 263.

Ceraski, W. 1880, *Astron. Nachr.*, **97**, 319.

Chaplin, G. B. 2018, *J. Amer. Assoc. Var. Star Obs.*, **46**, 157.

Chaplin, G. B. 2019, *J. Amer. Assoc. Var. Star Obs.*, **47**, 17.

Chaplin, G. B., Samolyk, G., and Screech, J. T. 2018, *J. Br. Astron. Assoc.*, **128**, 167.

Demircan, O., Akalin, A., Selam, S., Derman, E., and Mueyesseroglu, Z. 1995, *Astron. Astrophys., Suppl. Ser.*, **114**, 167.

Dobias, J. J., and Plavec, M. J. 1985, *Publ. Astron. Soc. Pacific*, **97**, 138.

Duerbeck, H. W., and Hänel, A., 1979, *Astron. Astrophys., Suppl. Ser.*, **38**, 155.

Frank, P. and Lichtenknecker, D. 1987, *BAV Mitt.*, No. 47, 1 (Lichtenknecker database http://www.bav-astro.eu/index.php/veroeffentlichungen/service-for-scientists/lkdb-engl).

Gudmundsson, L. 2017, Singular System/Spectrum Analysis (SSA; https://r-forge.r-project.org/projects/simsalabim).

Hall, D. S. 1975, *Acta Astron.*, **25**, 1.

Hall, D. S. 1989, *Space Sci. Rev.*, **50**, 219.

Hoffman, D. I., Harrison, T. E., McNamara, B. J., Vestrand, W. T., Holtzman, J. A., and Barker, T. 2006, *Astron. J.*, **132**, 2260.

Li, M. C. A., *et al.* 2018, *Mon. Not. Roy. Astron. Soc.*, **480**, 4557.

Manzoori, D., and Gozaliasl, G. 2007, *Astron. J.*, **133**, 1302.

Manzoori, D. 2016, *Astron. Lett.*, **42**, 329.

Maxted, P. F. L., Hill, G., and Hilditch, R. W. 1994, *Astron. Astrophys.*, **282**, 821.

Mkrtichian, D. E., *et al.* 2018, *Mon. Not. Roy. Astron. Soc.*, **475**, 4745.

Muller, G. 1906, *Astron. Nachr.*, **171**, 357.

Ohshima, O., *et al.* 2001, *Astron. J.*, **122**, 418.

Olson, E. C. 1982, *Astrophys. J.*, **259**, 702.

The R Foundation for Statistical Computing. 2018a, R: A language and environment for statistical computing (https://www.R-project.org).

The R Foundation for Statistical Computing. 2018b, CRAN: The Comprehensive R Archive Network (https://cran.r-project.org).

RStudio. 2018, RStudio software (https://www.rstudio.com).

Rodriguez, E., *et al.* 2004, *Mon. Not. Roy. Astron. Soc.*, **347**, 1317.

Samolyk, G. 2011, *J. Amer. Assoc. Var. Star Obs.*, **39**, 177, and numerous other similar reports.

Simon, V. 1997, *Astron. Astorphys.*, **327**, 1087.

Singh, K. P., Drake, S. A., and White, N. E. 1995, *Astrophys. J.*, **445**, 840.

Tkachenko, A., Lehmann, H., and Mkrtichian, D. 2010, *Astron. J.*, **139**, 1327.

Tupa, P. R., DeLeo, G. G., McCluskey, G. E., Kondo, Y., Sahade, J., Giménez, A., and Caton, D. B. 2013, *Astrophys. J.*, **775**, 46.

Wenger, M., *et al.* 2000, *Astron. Astrophys., Suppl. Ser.*, **143**, 9.

Zejda, M., Mikulášek, Z., and Wolf, M. 2008, *Astron. Astrophys.*, **489**, 321.

Appendix A: Outline code for SSA analysis and significance testing

This code outline shows the steps in the analysis but for a single bucketing period, method, and weighting regime, with user entering the trend and signal components after viewing the relevant SSA analysis output. Comments are in *italics*

```
rm(list=ls(all=TRUE))
#Load User-Defined Functions
setwd("D:/Documents/R/GBC defined functions")
source("astro_udf.R")
#load Rssa R library from Install Packages
library(Rssa)
library(simsalabim) # used for MCSSA

# USER INPUT
run = "findTrends" # "check O–C" "findTrends" "findSignals" "final"
star = "TW Dra"
allData = "ALL" # choose ALL or CCD for electronic only
dataMethodForSSA = "bucket" # "localPoly" or "bucket"
weightsMethod = "coded" # "inverseError" or "coded"
ylimRange = c(-0.6,0.7) # "y" range for some charts - initially a guess
periodAdjustment = -0.00008 # sometimes needed because epoch changes, try 0 at start
bucketSize = 500 # (days,) for example

# END USER INPUT

# STEP 1: Read in data and calculate OminusC

setwd("D:/Documents/EBdata")
tsIn<-read.csv(paste0(star,".csv"))
# data is named "TW Dra.csv", csv column data with header line as follows
# Tmin,N,period,OminusC,Visual,error
# N is not used; period column not used except first line gives stated period and second the epoch
# Visual is a flag as below; error is the data accuracy
# X Mkrkitchian data (RZ Cas only)
# C CCD
# E electric photometer
# K wedge photometer
# P single density photoplate
# F series of exposures

period = tsIn$period[1]; period = period + periodAdjustment
epoch = tsIn$period[2]
n1 = nrow(tsIn)
tsIn = na.omit(tsIn) # eliminate data rows where NA values exist
cat(paste0(n1-nrow(tsIn)," incomplete data rows eliminated"))
if (allData == "CCD") tsIn = tsIn[(tsIn[,5] == "C ") | (tsIn[,5] == "X ") | (tsIn[,5] == "E "),] # for electronic data only
ts = tsIn[-c(1:dropFirstNRows),]
ndata = nrow(ts)

# calculate expected Tmin and OminusC, and plot

expectedTmin = expectedTmin_udf(ts$Tmin,period,epoch) # library code calculates expected Tmin
ts$OminusC = ts$Tmin - expectedTmin
ts = cbind(ts,Weights=seq(1,1,length.out=ndata)) # add bad data flag column
if(weightsMethod == "inverseError") ts$Weights=1/ts$error else if (weightsMethod == "coded") {
 ts$Weights = ifelse(ts$Visual == "X " | ts$Visual == "C ",10,
 ifelse(ts$Visual == "E " | ts$Visual == "K ",3.25,1))
 }

ts = cbind(ts,badFlag=seq(0,0,length.out=ndata)) # set bad data flag
ts = cbind(ts,bucketNumber=seq(0,0,length.out=ndata)) # add column to say which bucket data is
ts <- within(ts,badFlag <- ifelse(!is.na(OminusC),0,1)) # bad flag for missing values in OminusC
ts$OminusC[is.na(ts$OminusC)] <- 0 # convert NAs to zero
xAxis = ts$Tmin
xlimRange = c(min(xAxis),max(xAxis))
# the following chart allows the user to check whether O–C data is based on two or more epochs
# user adjusts period to get a continuous set of data
plot(y=ts$OminusC,x=ts$Tmin,xlim=xlimRange,ylim=ylimRange,xlab="",ylab="",type="p",col=" grey20",pch=3, main=paste(star,": O–C raw data"))
if (run == "check O–C") stop()

# STEP 2: do a Local Poly Regression and reject outliers

ndataOld = ndata
if (allData == "CCD") {
 sigLevels = c(4,4,4,4,4,4)
 polySpan = 0.2
} else if (allData == "ALL") {
 sigLevels = c(4,4,4,4)
 polySpan = 0.03
}
ts = badRawDataMethodLP_udf(ts,polySpan,sigLevels,ylimRange) # library code eliminates outliers
```

```
ndata = nrow(ts)

# STEP 3: calculate bucketed data and find trends

# STEP 3: bucket data
if (dataMethodForSSA == "localPoly"){
 maxBuckets = floor((ts$Tmin[ndata]-ts$Tmin[1])/bucketSize)
 bucketOmC = seq(0,0,length.out=maxBuckets)
 cat("maxBuckets ",maxBuckets,"\n")
 bucketOmC = localPolyBuckets_udf(ts,bucketSize,maxBuckets,ylimRange,
xlimRange) # library  code

} else if (dataMethodForSSA == "bucket") {
 tmp = bucketDataAndFlag_udf(bucketSize,ts,drawPlot=FALSE) # library code
 maxBuckets = unlist(tmp[2])
 emptyBuckets = unlist(tmp[3])
 avgFilledBucketCount = unlist(tmp[4])
 bucketOmC = unlist(tmp[1])
 bucketOmC[is.na(bucketOmC)] <- 0
 ts$bucketNumber = unlist(tmp[6])
 flag = unlist(tmp[5])
}

# adjust period so best fit line is horizontal
Tmin <- ts$Tmin
x = c(1:maxBuckets)
fit = lm(bucketOmC~x)
slope = fit$coefficients[["x"]]
newPeriod = period + slope / bucketSize
bucketOmC = bucketOmC - slope * x
mu = mean(bucketOmC)
bucketOmC = bucketOmC - mu
plot(bucketOmC,xlim=c(1,maxBuckets),xlab="", ylab=paste(""),
type="p",col="black",
 lwd=2, pch=19, main=paste(star,"bucket O–C adjusted period"))

# find trends
L = floor(maxBuckets / 2)
s<-ssa(bucketOmC,L,kind="1d-ssa") #run Rssa
outputVecCount = 10
plot(s,type="vectors",idx=1:outputVecCount,xlim=c(1,L),col="black",lwd=2)
# vector data plots
plot(w<-wcor(s,groups=c(1:outputVecCount)),title=paste(star,"1d-ssa
correlation matrix"))
if (run == "findTrends") stop()
trendEV = c(1,2,5) # for example after inspection of the above charts
trend <- reconstruct(s, groups = list(EV = trendEV))
trend = unlist(trend[1])
plot(bucketOmC,xlim=c(1,maxBuckets),xlab="", ylab=paste(""),
type="p",col="black",
 lwd=1, pch=3, main=paste(star,"bucket O–C and trend"))
lines(xAxis=c(1,maxBuckets),trend,type="l",col="blue",lwd=1,lty=1)

# STEP 4: find signals

residual = bucketOmC - trend
s<-ssa(residual,L,kind="1d-ssa") #run Rssa
outputVecCount = 10
plot(s,type="vectors",idx=1:outputVecCount,xlim=c(1,L),col="black",lwd=2)
plot(w<-wcor(s,groups=c(1:outputVecCount)),title=paste(star,"1d-ssa
correlation matrix"))
if (run == "findSignals") stop()

# STEP 5: get reconstructed signals and frequencies

SSSAsignalEV = c(1:2) # for example after review of the above charts
r2 <- reconstruct(s, groups = list(EV = SSSAsignalEV))
signal = unlist(r2[1])
plot(signal,xlim=c(1,length(signal)),xlab="", ylab=paste(""),
type="l",col="black",
 lwd=2, main=paste(star,"signal1 SSSA"))
spectrum = spectral_udf(signal,drawPlot=TRUE,paste(star,"signal1 SSSA
spectrum"),smoothing="ar")
SSSAspecPeaks1 = unlist(spectrum[1])*bucketSize/365.25
spectrum2 = spectral_udf(signal,drawPlot=TRUE,paste(star,"signal1 SSSA
spectrum"),smoothing="pgram")
plot(residual,xlim=c(1,length(trend)),xlab="bucket", ylab=paste("O–C"),
type="p",col="black",
 pch=3, main=paste(star,"residual series"))
lines(signal,col="black",lwd=3,lty=1)

# STEP 6: MCSSA analysis

y = MCSSA(s, residual, 1000, conf = 0.99, keepSurr = FALSE, ar.method="mle")
plot(y, by = "freq", normalize = FALSE, asFreq = TRUE,
 lam.pch = 1, lam.col = "black", lam.cex = 1, sig.col = "black",
 sig.pch = 19, sig.cex = 1, conf.col = "darkgray", log = "xy",
 ann = TRUE, legend = TRUE, axes = TRUE)
```