Revealing of repeating structures in time series by the structure length eminence

Ts. B. Georgiev

Institute of Astronomy and National Astronomical Observatory,
Bulgarian Academy of Sciences, 72 Tsarigradsko Chaussee Blvd., 1784 Sofia, Bulgaria
`tsgeorg@astro.bas.bg`

Abstract. A conceptually simple method for revealing of periods in a time series is proposed. The method is based on a function of the structure length eminence, whose maxima positions correspond to the structures with maximal internal amplitudes. The method may be regarded as an analogue of the periodogram analysis, but working in the real domain, dealing with the period spectrum. Except the basic period P, the method reveals quasi-periods, such as $2P, 3P/2, 3/4P, P/2$, etc., but they are not obstacles. The method produces the average profile of any specified repeating structure (periodic or quasi-periodic).

The external error of the period estimation is 2-4%, while for the autocorrelation function or structure function it is 3-4 times larger. (The internal error of these 3 methods is ± 1 data point.) The method is applicable on time series with constant point step negligible large scale trend. Otherwise, the TS is resampled and the complicate large scale trend is followed and removed by local moving average. The size of the smoothing window (SW) ought to be 1-2 times larger than the suspected period length. Under this condition the derived period does not depend on the SW, but the external error does.

The test of the method on the monthly Solar Wolf numbers shows the known three period modes at 10.0, 11.0 and 11.7 yrs. The flickering of the symbiotic star AQ Men shows a complicated system of periods and quasi-periods, corresponding well to the system of the significant frequencies of Iłkiewicz et al. (2021), revealed by periodogram analysis.

Key words: time series: periods; Wolf number of Sun spots: periods; Flickering frequencues: AQ Mensae.

Introduction

Recently, we revealed periods and quasi-periods (commonly named hereafter QPs) in time series (TSs) of flickering symbiotic stars Rs Oph, T CrB, and MWP 560 (Georgiev et al., 2022, GBS 22). In such cases, the frequency methods are difficult to apply while the relatively short TS contains significant large scale trend(s) and strong noise. We started by removing the large scale trend globally, using a polynomial fit over the whole TS. Then we detected QPs by the minima positions of the structure function (SF). Later, we resampled the TS with a constant step and detected the same QPs by the maxima positions of the autocorrelation function (ACF). However, these methods derive QPs with a low accuracy and a low resolution, omitting the short QPs.

That is why we propose a method of the structure length eminence (SLE, MSLE). It checks numerous data point lengths m as lengths of repeating structures and it builds a suitable function of the SLE $E(m)$ (Eq. 1, 2, Fig. 1e). The SLE is a period spectrum. Any maximum time position t of the SLE (in the form $E(t)$) is a QP or a real period in the TS. The MSLE also produces the average profile of any specified structure length (Figs. f, g, etc.). The goals of the present paper are to describe the MSLE and to show its applications.

The used abbreviations are as follows:
ACF – autocorrelation function;
AV – average value;

HWHM – half-width at the half of the maximum;
MSLE – Method of the structure length eminence;
TS – time series;
QP – quasi-period;
SD – standard deviation;
SF – structure function;
SLE – structure length eminence;
SS – significant (repeating) structure;
SW – smoothing window size.

1. Method of the structure length eminence (MSLE)

Let us have a TS with n equally spaced data points $(F(t_i)(i = 1, n)$, with zero global average and with negligible large scale trend. Such a TS is the residual TS (Figs. b), derived from the input TS (Figs. a) by a large scale trend removal. Suppose that the TS contains a significant (repeating) structure (SS) with an unknown length m_s data points. We select m_s by checking numerous data lengths $m_1, m_2, ..., m_S, ..., m_M$ (e.g. 50-300 points) as follows.

For m_1 we pull up the first m_1 values of the TS into a set (initially empty) with cell numbers $j = 1, m_1$. Then we add there the next m_1 values from the TS. Later, we add the next m_1, etc. In such a way we perform $k_1 = n/m_1$ (integer number) additions. Then, for every jth cell, containing k_1 additions, we derive the average value (AV) a_{j,m_1} and standard deviations (SD) d_{j,m_1}. The program performs this procedure for every m subsequent data between m_1 and m_M, deriving the relevant mth profiles of the structure in AVs and SDs.

Later, the program derives the average amplitude $A(m)$ and the average SD $D(m)$ for every mth profile:

$$A(m) = < |a_j| >_m, \qquad D(m) = < d_j >_m . \qquad (1)$$

Note that $A(m)$ gathers the absolute AVs $|a_j|$ and it is the average amplitude of the structure with QP P_0 and length m data points. Also, $D(m)$ is the average noise in this structure.

When a SS with length m points is present, it produces a maximum of $A(m_S)$ and minimum of $D(m_S)$ (Fig. 1e). Therefore, a more sensitive indicator of SSs is the ratio $A(m)/D(m)$. For convenience we use it multiplied by 100 and call it **structure length eminence**, SLE:

$$E(m) = 100 \times A(m)/D(m). \qquad (2)$$

Every value of the SLE $E(m)$ may be regarded as a kind of signal-to-noise ratio, expressed by percentages in respect to the noise value. Certainly, the local noise continuum of the SLE must be removed for a correct signal-to-noise ratio estimation.

The MSLE requires a constant data step. In many cases the TS may be resampled by linear interpolation. When the TS resolution is constant, but not sufficient, it may be increased twice by a suitable simple resampling (Fig. 2f).

The MSLS is applicable on TS without significant large scale trend and zero global average. Therefore, a preliminary amplitude decomposition of the input TS F_I into smoothed TS F_S and residual TS, $F_R = F_S - F_I$, is necessary. The subject of the MSLE is the **residual TS** $\Delta F(t)$, expressed in percentages towards the smoothed TS:

$$\Delta F(t) = 100 \times (F_I(t) - F_S(t))/F_S(t). \tag{3}$$

Using such a residual TS the results of various MSLE applications may be easily compared.

The simple MSLF is applicable when the large scale trend may be fitted and removed by a low degree polynomial (Fig. 1).

The basic MSLF is applicable when the large scale trend is complicated, but it may be followed and removed by a moving average (Figs. 3, 5, 6). The practice shows that the SW must be 1-2 times larger than the supposed basic period.

2. Simple MSLE for the Wolf number of the last 6 Solar cycles

The monthly Wolf number W gives a simple and clear test for the MSLE. However, the distribution of W or lg W has strong positive or negative skewness, respectively. Contrariwise, the numbers $W_R = W^{1/2}$ show a quasi-normal distribution (Georgiev 2022, Fig. 1). Therefore, the arithmetic operations (fitting, trend removal) become more appropriate on the numbers W_R. Moreover, the TS of W_R is close to a sinusoid. Also, the bright and faint regions of this TS (Figs. 1b, 5b) and the profiles (Figs. 1f, 5f) pose almost equal scatter.

Figure 1 illustrates the exploring of the simple MSLE on the Wolf number W_R. The time interval is 1950-2017 yr (67 years, 814 months) with 1 step = 1 month. Figure 1a represents the input TS whose large scale trend is fitted by a 2nd degree polynomial (without applying any SW). Figure 1b shows the residual TS (Eq. 3). In this test the trend in (a) is faint and its removal may be omitted. But then the humps of the SS in the SLE should be a little lower and wider. Then the accuracy of the period estimation should be a little lower.

Figures 1c and 1d show the ACF and the SF. In this simple example they recognize well the basic period of about 10.6 yrs. The relative half width at the half of the ACF maxima or SF minima (HWHM), δP, is larger by about 20% or about 15%, respectively.

Figure 1e shows the SLE $E(t_L)$ (Eq. 2). The functions of the amplitudes $A(t_L)$ and of the noise $D(t_L)$ (Eq. 1) are shown too, only for this test. The SLE is derived from the residual TS (b) inside the lengths $m = 70 - 360$ data points (months) or $t_L = 5 - 30$ yrs. The positions of the local maxima (dots) correspond to the QPs of SS. Figures 1f-1i show some SS profiles.

Only the dominant SLE maximum corresponds to the real basic period, $P = 10.83$ yrs ($m = 130$ data poits or months). The basic period is confidently recognized by its high eminence E, 12 times above the local SLE continuum and its single profile shape (f), containing one maximum and one minimum. The P-hump is wide because the structures with data length 131, 132, etc., or 129, 128, etc., cause gradual decrease or increase of the SL0.

The internal error of P is 1 step (1 month or 0.083 yr). The relative HWHM for the SSs here is about 8%, 2-3 times shorter than in the ACF or SF. The
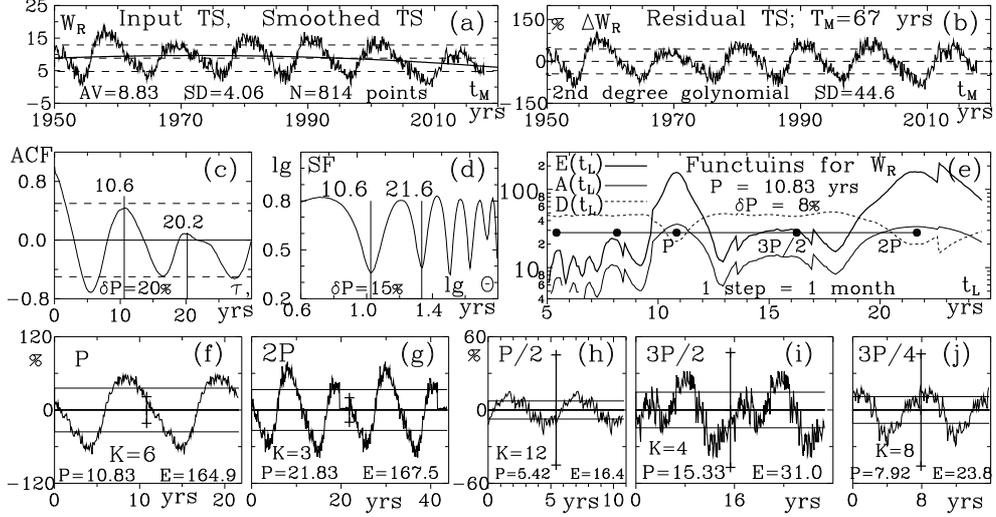
Fig. 1. MSLE analysis of the Wolf number $W_R = W^{1/2}$ with a global polynomial trend removal. **(a)**: Input TS (thin curve) and smoothed TS (thick curve, 2nd degree polynomial) with AV and SD of the input TS. Dashed lines show the levels of AV and AV $\pm$ SD. N is the data number. t_M is the monitoring time; **(b)** Residual TS (Eq. 3, in percentages towards the polynomial levels, with monitoring time interval T_M and the SD of the residual TS. Dashed lines show the levels of 0 and 0 $\pm$ SD; **(c, d)**: ACF and SF of the residual TS (b) with respective steps τ and θ. The extrema positions are denoted by their periods. δP are the HWHM of the humps of ACF and SF, respectively. **(e)**: Graphs of the SLE $E(t_L)$, $A(t_L)$ and $D(t_L)$ (Eqs. 1, 2) derived from the residual TS (b). t_L is the structure length time interval. Dots show the positions, i.e. QPs, of the SSs in the SLE. P is the basic period and the left dots correspond to P/2 and 3P/4. δP is the typical HWHS of the humps of all SS in the SLE; **(f, g, h, i, j)**: Profiles of the SLE maxima positions at $P, 2P, P/2, 3P/2, 3P/4$, their periods P, their SLE values (eminences) E (Eq. 2) and their numbers of the profile additions K. The vertical bar shows the average noise, 0 $\pm$ D. The horizontal lines show the average amplitude 0 $\pm$ E.

right big hump reflects the SS of 2P ($m = 261$ data points or 21.75 yrs). In comparison with the basic P-hump, the 2P-hump is about 2 times wider and 2 times lower. The 2P-profile in (g) contains two basic periods. The 2P-hump does not correspond to a basic period, but to a QP.

The SLE maxima at 3P/2, about $t_L = 16$ yrs, at 3P/4, about $t_L = 8$ yrs, even about $P/2$, $t_L = 5.4$ yrs, are detectable too. In Figs. 1h and 1i, the respective profiles are 6-10 times lower with respect to the basic P-profile (f), corresponding to other QPs. Note that the phase zero points of all profiles correspond to the phase of the beginning of the monitoring.

Figures 2a-2e show the formations of profile shapes from a pure sinusoidal TS. Figures 2a and 2b reflect the case when m corresponds to the half period, $P/2$ or phase $\Delta\varphi = 180^0$. Then every pair of such additions to the average profile contains the sum of the curves (1) and (2). The result is a zero average profile $a_j = 0, j = 1, m)$ or $A(m) = 0$, line (3). Every even number K of additions gives the same result. However, every odd number k contains additionally
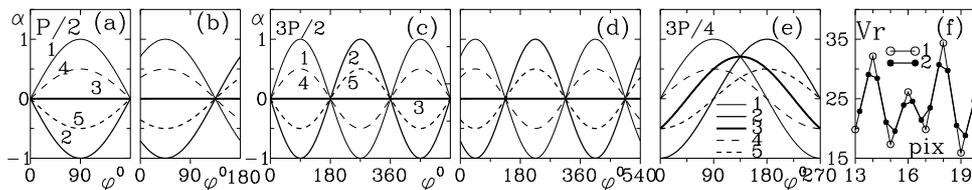
Fig. 2. Formation of SS profiles from a sinusoidal TS. **(a, b)** for $P/2$; **(c, d)** for $3P/2$; **(e)** for $3P/4$. φ and α are the phase and amplitude; **(f)** Method of Two times increase of the TS resolution, applied for Fig. 3.

the curve (1) (for a sinusoid with zero initial shift) or (2) (for a sinusoid with a 180^0 initial shift). Then the results for $a_j, j = 1, m)$ are the curves (4) or (5), respectively, with a single extremum. In Fig. 2b, the beginning of the sinusoid is shifted at 45^0. The results for $A(m)$ in all above-mentioned cases will be the same as in Fig. 1a. In practice, the input TS is not a sinusoid and even at even k the profile poses some low amplitudes. In Fig. 1h the eminence of $P/2$ is notable, but it is 10 times lower in comparison with the P in Fig. 1f.

Figures 2c and 2d show the case when m corresponds to one and a half periods, $3P/2$ and $\Delta\varphi = 540^0$. The interpretation is similar to the above case. Every even number k of such additions gives zero profile. Every odd number k contains curve (1) or (2). The result for $a_j, j = 1, m)$ is the curve (4) or (5), with triple extrema. In Fig. 2d, the beginning of the sinusoid is shifted at 45^0 and the result is the same. The profile contains 2 minima and 1 maximum or 2 maxima and 1 minimum. In Fig. 1i, the eminence of the $3P/2$-profile is about 5 times lower in comparison with the P-profile in Fig. 1f.

Figure 2e shows a part of the model of the more complicated case of $3P/4$, when $\Delta\varphi = 270^0$. Four such additions give a zero profile. Only 2 of them, (1) and (2), are shown. The next +1, +2 or +3 such additions cause different results. When the sinusoid begins from zero, the result of the case "+2" is shown by the curve (3). For the cases "+1" or "+3" the result is (4) or (5). The profile is always single. The eminence of the $3P/4$-profile in (j) is about 6 times lower in comparison with the P-profile in (f). Similar analysis of the more complicated cases for $2P/3$ and $4P/3$ may also be performed.

So, in a TS with one large basic period P (Fig. 1a), the ACF and SF are useful period estimators, but the MSLE estimates the period and its HWHM more accurately. The basic period P poses the highest hump (towards its continuum) and it has a single profile shape. The profiles corresponding to $2P/3, 3P/4$ etc., have a significantly lower eminence, having again single shapes. The profiles corresponding to $3P/2, 2P, 3P$ etc., have lower eminences too, but they have multiple shapes. The relative HWHM is approximately constant in the given SLE.

Moreover, in Fig. 1e the MSLE reveals 1 period and 4 quasi-periods. This "richness" gives a possibility for a weighted estimation of the period P (if needed), using more than one well pronounced hump position.

3. Basic MSLE for the Planetarium visit number from 6 calendar years

The weekly visit number V of the Smolyan Planetarium for the calendar years 2013-2017, gives a more illustrative TS test. It visualizes all particularities of the MSLE. Again the "jagged" distribution of V has strong positive skewness. Like in the case of the Wolf number (Section 2), the number $V_R = V^{1/2}$ shows a quasi-normal distribution (Georgiev, 2022), which is suitable for arithmetic operations.

In this case, the TS resolution is increased 2 times. The method of resampling is shown in Fig. 2f. The initial data number is $n = 313$ (weeks) and the resampled data number is $2(n-1) = 624$ (half-weeks). This transformation decreases the slope of the SF, which is not of interest here, but it ensures better representation of the short structure QPs.
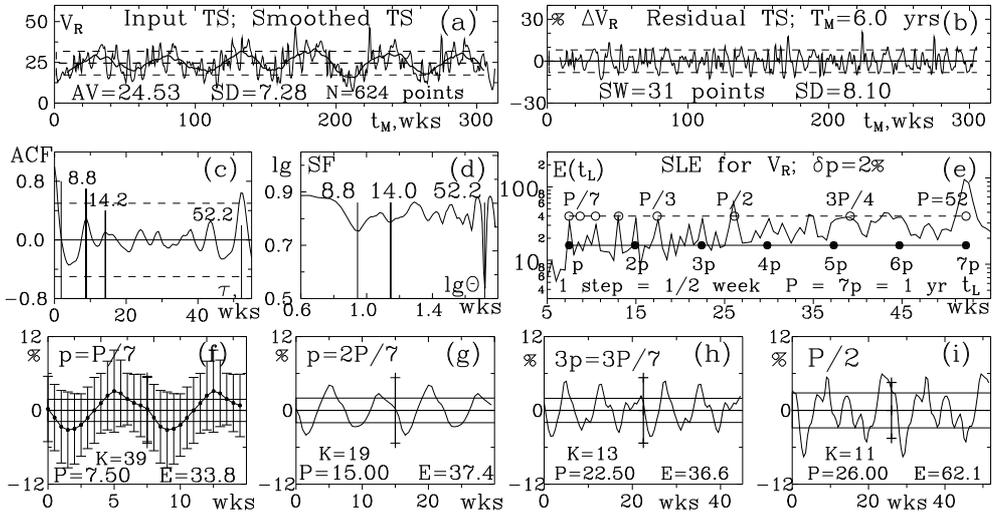
Fig. 3. MSLE analysis of the number $V_R = V^{1/2}$ with trend removal by sliding average. See also Fig. 1. **(a)**: Input TS (thin curve) with its AV and SD, plus smoothed TS (thick curve); **(b)**: Residual TS with its SW and SD toward the smoothed TS; **(c, d)**: ACF and SF of the residual TS with QPs estimations; **(e)**: SLE from the residual TS with basic short period $p = 7.5$ weeks and cumulative profile $P = 7p = 52$ weeks. The respective profiles are shown in Fig. 3f and 4a. The dots mark the SS for p, $2p$, ..., $7p$. The circles mark the SS for P, $P/2$, ..., $P/7$. **(f, g, h, i)**: Average profiles of SSs.

Figure 3 illustrates the exploring of the MSLE on the TS of the visit number V_R. A short period of about 8 weeks (16 points) in the TS is preliminarily suspected to be the basic one. Figure 3a represents the input TS and the smoothed TS. The large-scale trend is built by a moving average with SW of 31 points. Figure 3b shows the relevant residual TS. Note that the natural period $P = 1$ yr is well protruded in (a), but it is not visible in (b). The reason is the short SW.
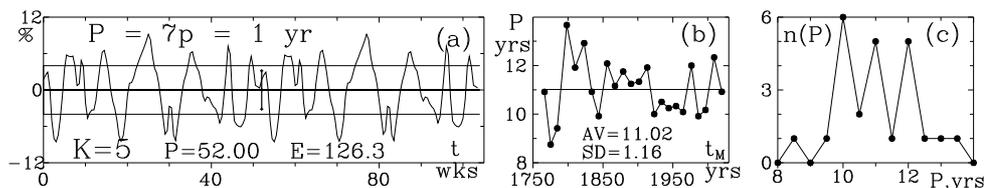
Fig. 4. **(a)**: Average profile of the SS for $P = 7p = 1$ yr from the SLE in Fig. 3e; **(b)**: Behaviour of the Sun cycle periods over the years; **(c):** Histogram of the Sun periods.

Figures 3c and 3d show the ACF and the SF of the residual TS. In this complicated example, they recognize well only the long period, $P = 1$ year. The noticeable structure length at 8.8 weeks corresponds to the humps df $P/7$ and $P/5$ in (e) together.

Figure 3e shows the SLE of the residual TS (b) in the interval 5-55 weeks (10-110 points). The shortest well pronounced period, that occurs the basic one, is $p = 7.5$ weeks (15 points, 1/7 yr). The natural large period, having the highest hump, is $P = 1$ yr (52 weeks, 104 points). The respective profiles are shown in Fig. 3f, with 1 basic period, and in Fig. 4a, with 7 basic periods.

The complicated SLE in Fig. 3e poses many humps that may be regarded as members of two systems. One of them with origin from p (dots) contains humps at $p, 2p, ..., 7p = P$ (without $5p$). Another, with origin from P (circles), contains the humps at $P, 3P/4, P/2, P/3, ..., P/7$ (without $P/6$).

Figures 3f, 3g, 3h, and 3i show the SS profiles corresponding to the remarkable humps. The basic period is $p = P/7$. It is explained with the scholar schedules. Its shape in (f) is very smooth, because it is the result of $K = 39$ additions. The profile for $2p$ in (g) and $3p$ in (h) contain 2 and 3 p-profiles, respectively. Note that the profile for $P/2$ in (i) contains 3.5 p-profiles.

In this test, the ACF and SF are useless for short period estimations. MSLE confidently shows the basic period p, cumulative period P and two systems of QPs distinguished also by their specific profile shapes.

4. Basic MSLE for the Wolf number of all 24 Sun cycles

All Wolf numbers W (Fig. 5, top) give an important test of the MSLE. Figure 5 illustrates the processing of the numbers $W_R = W^{1/2}$ (see Section 2) for all solar cycles (267 years, 3226 months).

Preliminarily, Figs. 4b and 4c represent the statistics of the cycle periods. The minima positions are derived after a suitable smoothing of the TS. Figure 4b shows the behaviour of the period over the years. Figure 4c shows a histogram of the periods where the triad of period modes protrudes clearly.

Figures 5a and 5b show the smoothed and residual TS of W_R. The SW size is 265 points (months), while the preliminary estimated period is 130 points (months). Dalton minimum (1790-1840) is well expressed in (a), but not visible in (b).
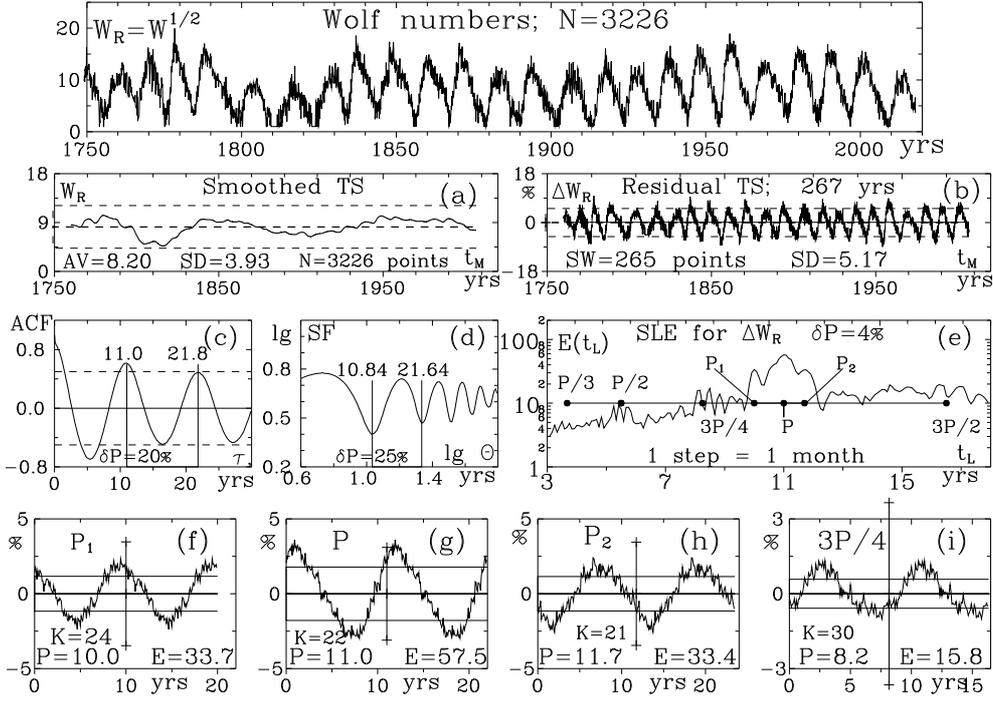
Fig. 5. MSLE analysis of whole Wolf numbers $W_R = W^{1/2}$ with local trend removal. See also Fig. 1. Top graph: Input TS; **(a,b)**: Smoothed TS and residual TS; **(c,d)**: ACF and SF of the smoothed TS in (b) and detected periods; **(e)**: SLE $E(t_M)$ for the residual TS (b) and scheme of the revealed periods. See the text. **(f, g, h, i)**: Average profiles for P_1, P, P_2 and $3P/4$.

Figures 5c and 5d show ACF and SF for the residual TS in (b). These functions estimate confidently the dominant period of 10.8-11.0 years, but they do not distinguish the triad of the period modes. The HWHMs of the triad is 20-25% and ACF and SF do not detect shorter quasi-periods. The MSLE gives more accurately the system of the periods.

Figure 5e shows the SLE (Eq. 2) for structure lengths 3-20 yr (36-240 points). The highest maximum at $P = 11.0$ years (132 points), with HWHM about 4%, is about 10 times higher than its continuum. Two satellite period modes appear at $P_1 = 10.0$ yr (120 points) and $P_2 = 11.7$ yr (140 points). Note that the triads of QPs appears about $3P/2, 3P/4$ and $P/2$. A hint of a triad may be seen even at $P/3$.

Figures 5f, 5g, 5h and 5i show the profiles of SS. All shapes are close to sinusoids but with higher tops and deeper valleys, plus positive skewness. The phase zero point corresponds to the beginning of the residual TS.

We note again, that the period estimation P by the means of the MSLE does not depend on the SW, if it lays in the bounds of $1P - 2P$. However, the height and the HWHM of the SLE humps do.

5. Basic MSLE for AQ Men

Recently Iłkiewicz et al. (2021, ISC21), using TESS data, analyzed the flickering of the cataclysmic star AQ Men, Sector 13. In the periodograms, they revealed 24 significant frequencies. Four of them, corresponding to the orbital period (3.4 h), positive superhump (3.6 h), negative superhump (3.2 h), and superorbital period (57 h), are the basic ones. Twenty high frequencies, corresponding to combinations of the basic ones, are noticeable too (ISC21, Table. 1).
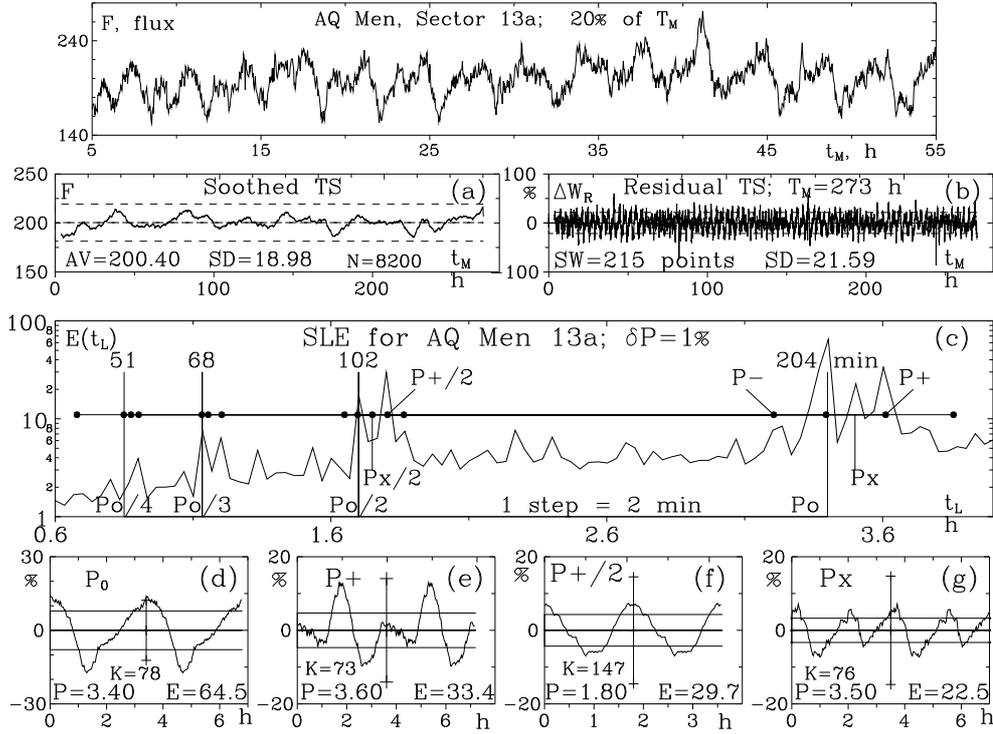
Fig. 6. MSLE analysis of the flickering of AQ Men with local trend removal. See also Figs. 1, 3 and the text. Top graph: A part of the Input TS; **(a,b)**: Smoothed TS and residual TS; **(c)**: SLE from the residual TS in (b) and scheme of 8 periods corresponding to the frequencies of ISC21; **(d, e, f, g)**: Average profiles for $P_0, P+, P+/2$ and Px.

Figure 6 shows the application of the MSLE on the Sector 13a of AQ Men. Because of a few interruptions, the original TS is resampled to have a constant step, again 2 min. Because of the peculiar tail of the original TS, the first 8200 points (273 h from 330 h) are only used. About 20% of the TS is shown in the top of Fig. 6.

128

Figures 6a and 6b show the smoothed and residual TS of the flux F. The SW has 215 points (430 min, 7.17 h).

Figure 6c shows the SLE (Eq. 2) for structure lengths 0.6-4 h (16-120 points). The HWHM of the humps is about 1%. Sixteen frequencies of ISC21, recalculated in periods, are shown by dots. The SLE resolution is not sufficient and the shortest 8 periods are omitted. The basic period is $P_0 = 3.4$ h (204 min, 102 points). Two neighbouring periods appear, $P_+ = 3.6$ h (positive superhump) and $P_- = 3.2$ h (negative supehump). Relevant periods appear (or may be suspected) at about $P_0/2, P_0/3$ and $P_0/4$.

Figures 6d, 6e, 6f, and 6g show the shapes of the remarkable SS. Note that in (e), the profile of P_+ described by ISC21 as a peculiar one contains two components with different shapes. However, its counterpart $P++/2$, well revealed by ISC21 and by the SLE in (c), shows a single shape in (f).

Note also that the P_x, well pronounced in the SLE (c) at 3.5 h, and having double shape in (g), is absent in ISC21. However, its counterpart $F_x/2$, missing in (c), is well pronounced in ISC21. The frequency of the last one is noted to be $2(P_0 - N)$, where N is the superorbital period. Therefore, the period P_x corresponds to the frequency of $(P_0 - N)$.

The test on AQ Mem shows that the MSLE confidently recognizes the basic periods and the most remarkable of their shorter counterparts.

Summary

1. We described the MSLE which may be regarded as an analogue of the periodogram analysis, but giving the spectrum of the periods. This method is applicable on time series with constant step and negligible large scale trend. The basic MSLE includes local smoothing of the TS by a moving average. The period estimation does not depend on the smoothing when the SW size is 1-2 times larger than the suspected basic period.

2. The MSLE is illustrative. The basic period poses the highest SLE maximum (above the local continuum) and it has single period shape (with one maximum and one minimum). The shortest quasi-periods have single shapes, but a few times lower SLE maxima. The largest quasi-periods have multiple shapes and lower maxima. The HWHM of all significant humps of the SLE is typically 2-4%.

3.In various tests (Section 3, 4, 5) the MSLE, including local smoothing of the TS, reveals confidently and accurately the periods and quasi-periods.

References

Georgiev, Ts. 2022, *Bulg. Astron. J.* 36, 70
Georgiev, Ts., Boeva, S., Stoyanov, K. et al. 2022, *Bulg. Astron. J.* 37, 62 (GBS22)
Iłkiewicz, K., Scaringi, S., Court, J. M. C.. et al. 2021 *MNRAS* 503,4050 (ISC 21)