FROM O-C TO SCUSUM+

Grant Foster
American Association of Variable Star Observers
25 Birch Street
Cambridge, MA 02138

Presented at the October 1993 Meeting

Abstract

Standard O-C (Observed-Computed) analysis assumes that the residual deviations of observed times from linearity are independent random variables. As Koen and Lombard (1993) point out, when a star's period is not constant but fluctuates about a constant mean value, this assumption leads to false confirmation that the period is changing. They supply an alternate test, CUSUM (CUmulative SUMs), which amounts to a series of scaled comparisons of "before" and "after" series averages. It can be extended by computing a separate scale factor for each individual CUSUM, yielding the SCUSUM test. This can be further extended to the SCUSUM+ test by including the effect of lag-1 correlation of the data.

1. Introduction

We define an individual *period* for a periodic phenomenon as the time between consecutive occurrences of some notable event, e.g., maximum or minimum brightness of a variable star. If we know the true times τ_n, n = 0,1,2,..., N for N+1 consecutive events, then we can compute N successive values of the true period

$$\pi_n = \tau_n - \tau_{n-1} \qquad (1)$$

However, we don't know the true times τ_n; rather we obtain estimates T_n, which enable us to compute estimates of the periods

$$P_n = T_n - T_{n-1} \qquad (2)$$

For simplicity's sake, I will assume that all times are observed, none are missing from the sequence.

We can then study the series of observed times T_n or the series of estimated periods P_n, to search for period changes. The result we obtain will depend on the error model we use, and on the question we ask. Searching for period change actually raises two questions: whether the period is constant, and whether the *mean* period is constant. Periods often exhibit random fluctuations about a constant mean value. In this case, although we might be able to demonstrate nonconstancy of individual periods, the strongest tests are required to establish any evolutionary change for the mean period.

2. O-C model

A standard tool for detecting period changes (especially for variable stars) is O-C

(Observed-Computed) analysis. One of the assumptions forming what I will call the *linear O-C null hypothesis* is that the period is truly constant

$$\pi_n = \mu \tag{3}$$

This is the hypothesis for linear O-C regression; in general, O-C assumes that the period is a smoothly varying function of cycle number, i.e., that it has no random fluctuations. The general statistical behavior is the same as that of the linear model.

The true event times are linearly related to cycle number

$$\tau_n = \tau_0 + n\mu \tag{4}$$

The observed times differ from these by some measurement error

$$T_n = \tau_n + \eta_n \tag{5}$$

The deviations η_n have some expected value (which for this analysis I take to be zero) and some variance

$$<\eta_n> = 0 \qquad\qquad <\eta_n^2> = \eta^2 \tag{6}$$

where brackets "< >" indicate the expectation value of the enclosed quantity. The estimated periods exhibit deviations according to

$$P_n = \mu + \eta_n - \eta_{n-1} \tag{7}$$

This is the *linear O-C model.*

Using trial values for the elements τ_0 and μ, we can compare the observed times T_n to the computed times, generating residual deviations, the "observed-minus-computed" values; these are the input data to O-C analysis.

We can then regress the residuals against cycle number, by the method of least squares. Regressing onto a straight line enables us to refine the constants τ_0 and μ; we obtain thereby not only estimates of the regression constants, but also confidence limits for the constants, enabling us to evaluate the validity of the regression. However, as Koen and Lombard (1993) have pointed out, these confidence limits are based on the oft-unwarranted assumption that the phenomenon is periodic with a *constant* period.

Of course we can compute the average period

$$\begin{aligned}\overline{P} &= (1/N)\,\Sigma\, P_n = (1/N)\,(T_N - T_0) \\ &= (1/N)\,(N\mu + \eta_N - \eta_0) = \mu + (\eta_N - \eta_0)/N\end{aligned} \tag{8}$$

The expectation value of the average period, and its variance, are

$$\langle\overline{P}\rangle = \mu$$

$$\langle\overline{P}^2\rangle - \mu^2 = 2\eta^2 N^{-2} \tag{9}$$

In this case the variance of the average period goes as $1/N^2$, less than the variance of an ordinary average, which goes as $1/N$. This is because the period is truly constant; the times T_0 and T_N constitute a direct observation of N times the period.

Despite this serendipitous accuracy, O-C analysis is even better. This is no surprise; the average period is determined by only the first and last observed times T_0 and T_N, while O-C analysis includes additional information from all the other observations. Therefore when the above assumptions are fulfilled (constant period, only measurement errors), O-C analysis provides an unbiased estimate of the period whose variance goes as $1/N^3$

$$\langle P_{O-C}\rangle = \mu$$

$$\langle P^2_{O-C}\rangle - \mu^2 = \frac{12\eta^2}{N(N+1)(N+2)} \tag{10}$$

A common further analysis is to regress on a quadratic form

$$T_n = \tau_0 + n\mu + Cn^2 \tag{11}$$

where the constant C is the quadratic coefficient. If this constant is nonzero (within its confidence limits), it is taken as evidence that the period is changing, and the quadratic term gives an estimate of the rate of change of the period. Another common practice is to fit straight line segments to pieces of the O-C diagram. Each line segment represents a constant period; each different line segment represents a different period.

3. CUSUM model

Physical systems often exhibit periods which are not constant, but fluctuate about a constant mean value. Therefore the true periods may not be constant, but a random variable with constant mean and unknown variance

$$\langle\pi\rangle = \mu \qquad \langle\pi^2\rangle - \mu^2 = \theta^2 \tag{12}$$

Hence to the n^{th} true period we can assign the form

$$\pi_n = \mu + \theta_n \tag{13}$$

where the deviation θ_n represents a period fluctuation, not an error in measurement. The event times deviate according to

$$\tau_n = \tau_0 + n\mu + \sum_{k=1}^{n} \theta_k \tag{14}$$

If we ignore measurement errors of the critical event times, then the observed times and periods are

$$
\begin{aligned}
P_n &= \pi_n = \mu + \theta_n \\
T_n &= \tau_n = \tau_0 + n\mu + \sum_{k=1}^{n} \theta_k
\end{aligned}
\tag{15}
$$

This is the *CUSUM model.*

Now the problem with O-C analysis emerges: the deviation in T_n includes the *accumulation* of all the preceding period fluctuations. In fact these accumulated deviations are random-walk sums; hence they do not exhibit the simple error structure postulated in (6); if we apply O-C analysis to data with this error structure, the confidence limits on estimated parameters will be much mistaken.

If we use the observed time of cycle zero T_0 as our provisional choice of τ_0, and the average period $\overline{P}$ as our provisional period, then the O-C values are simply

$$
C_k = \sum_{a=1}^{k} (P_a - \overline{P}) \tag{16}
$$

This is identical to Koen and Lombard's (1993) and Williamson's (1985, see also Isles and Saw 1989) statistic to test for a change in the series mean, called CUSUM (CUmulative SUM), also known as the *span test.* The difference between O-C and CUSUM is that O-C values are treated as independent random variables, while CUSUM values are treated as random-walk sums.

To form a test statistic, they scale the C values by the standard deviation of a random-walk sum,

$$
S = \text{Std.Dev.}(\Sigma\epsilon) \approx \theta\sqrt{N} \tag{17}
$$

where N is the number of data, and θ is an estimate of the standard deviation of the periods. S is not the actual standard deviation of any C_k value, but it serves as a perfectly good scale factor (it has the correct N- and θ-dependence). The scaled CUSUM values are

$$
C'_k = C_k/S = C_k/[\theta\sqrt{N}] \tag{18}
$$

and the test statistic is

$$
D_N = \max |C'_k| \tag{19}
$$

For large N it follows the Kolmogorov-Smirnov distribution, given approximately by

$$
\text{Prob.}(D_N > d) \approx 2e^{-2d^2} \tag{20}
$$

For smaller N, critical values of the CUSUM test are calculated by computer simulation (Williamson 1985). Williamson's computations assume as the null hypothesis that the series mean is constant, *and* that the data are a *normally* distributed random variable.

4. SCUSUM

CUSUM does not in fact sum random fluctuations ϵ, but their *deviations from average*. The consequences emerge if we cast it in the form

$$\begin{aligned} C_k &= \sum_{a=1}^{k} P_a - kP \\ &= \sum_{a=1}^{k} P_a - k/N \sum_{a=1}^{N} P_a \\ &= \sum_{a=1}^{k} (1-k/N)P_a - k/N \sum_{a=k+1}^{N} P_a \end{aligned} \qquad (21)$$

Hence we have that

$$\frac{NC_k}{k(N-k)} = \sum_{a=1}^{k} (1/k)\, P_a - \sum_{a=k+1}^{N} (1/(N-k))P_a = x_{\text{before}} - x_{\text{after}} \qquad (22)$$

so the k[th] CUSUM value is proportional to the difference between the series averages before and after time t_k. Therefore CUSUM is formally equivalent to a direct comparison of "before" and "after" averages. In this light, the significance of a CUSUM test is crystal clear. A standard computation of the variance inherent in the comparison of two averages reveals that actual standard deviation for a single CUSUM is

$$S_k = \theta\sqrt{k(1-k/N)} \qquad (23)$$

The salient point is that each CUSUM value has a different scaling. In standard CUSUM analysis, all values are assigned the same scale factor, so that values near the beginning or end of the time series are scaled by too large an amount; CUSUM exhibits reduced sensitivity to period changes which occur early or late in a time series.

We can remedy this by computing a set of individually scaled CUSUM values, which I will call *SCUSUM* values

$$\tilde{c}_k = C_k/[\theta\sqrt{k(1-k/N)}] \qquad (24)$$

With this scaling, any single $\tilde{c}_k$ is a normally distributed random variable with mean 0 and variance 1. The test statistic is

$$\tilde{D}_N = \max|\tilde{c}_k| \qquad (25)$$

Note that the cycle corresponding to the maximum $|c_k|$ is likely to be different from the cycle corresponding to the maximum $|C'|$.

Interestingly enough, the individual scale factors for CUSUM statistics were derived over 60 years ago (but not used) by Eddington and Plakidis (1929). Later Sterne (1934) elucidated the same scheme, but rejected the individually scaled CUSUM test in favor of his own test for a change in the series mean. The primary difficulty is that because the statistics for different k values are so strongly correlated, the probability distribution of the maximum of all the $|c_k|$ values is unknown.

This objection is overcome by our ability to test computer-generated data. Table 1 below lists the numerically determined critical values for the SCUSUM test. These values are generated using s^2 to estimate the data variance σ^2. The tabulated value is the average of 25 separate estimates (which permits a calculation of the standard deviation associated with the value), each of which represents the correct percentile limit for 1000 SCUSUM tests in which the raw data are normally distributed. This method is identical to that used by Williamson (1985) to determine critical values for the standard CUSUM test.

Table 1. Critical values for max |SCUSUM|, when estimating σ^2 by s^2.

	False-alarm probability							
	----10%----		----5%----		----1%----		----0.5%----	
N	Ave	σ_{Ave}	Ave	σ_{Ave}	Ave	σ_{Ave}	Ave	σ_{Ave}
5	1.613	0.002	1.693	0.002	1.810	0.004	1.846	0.003
10	2.079	0.004	2.226	0.005	2.453	0.006	2.520	0.007
15	2.277	0.004	2.452	0.005	2.748	0.009	2.842	0.011
20	2.392	0.005	2.588	0.008	2.931	0.011	3.037	0.013
25	2.476	0.005	2.678	0.007	3.052	0.010	3.183	0.015
30	2.518	0.007	2.730	0.008	3.125	0.010	3.261	0.016
40	2.594	0.006	2.816	0.009	3.227	0.013	3.372	0.019
50	2.635	0.006	2.868	0.008	3.299	0.017	3.465	0.018
60	2.684	0.005	2.920	0.008	3.341	0.015	3.481	0.021
70	2.708	0.008	2.953	0.009	3.409	0.011	3.577	0.019
80	2.735	0.005	2.966	0.007	3.443	0.014	3.582	0.015
90	2.767	0.009	3.001	0.010	3.451	0.013	3.644	0.019
100	2.775	0.006	3.012	0.008	3.474	0.020	3.680	0.027
100	2.775	0.006	3.052	0.008	3.502	0.015	3.649	0.019
140	2.826	0.006	3.065	0.009	3.530	0.013	3.709	0.022
160	2.850	0.007	3.086	0.008	3.545	0.017	3.744	0.022
180	2.851	0.006	3.092	0.009	3.577	0.015	3.745	0.021
200	2.868	0.005	3.113	0.008	3.583	0.018	3.787	0.027
300	2.906	0.006	3.152	0.008	3.646	0.012	3.825	0.018
400	2.950	0.005	3.200	0.010	3.686	0.022	3.899	0.027
500	2.989	0.006	3.231	0.010	3.704	0.015	3.913	0.015

5. SCUSUM+

The standard deviation responsible for the random walk is that of the periods themselves, not that of the time estimates. For variable stars, we expect not only period fluctuations as in (13), but also measurement errors of the critical times as in (5). Thus the observed times are

$$T_n = \tau_0 + n\mu + \sum_{k=1}^{n} \theta_k + \eta_n \tag{26}$$

and the observed periods deviate according to

$$P_n = \mu + \theta_n + \eta_n - \eta_{n-1} \tag{27}$$

This is the *SCUSUM+* model. Note that the limit $\eta \to 0$ is ordinary SCUSUM; the limit $\theta \to 0$ is pure O-C.

The required scaling for SCUSUM+ is no longer (23) but

$$S_k = \sqrt{k\theta^2 (1\text{-}k/N) + 2\eta^2 (1-k/N+k^2/N^2)} \tag{28}$$

The coefficient of $2\eta^2$ is always close to 1, especially for k small or large (when it counts the most), so we may use this approximation and define the SCUSUM+ values as

$$c = \frac{C_k}{\sqrt{k\theta^2 (1-k/N) + 2\eta^2}} \tag{29}$$

The test statistic is again max $|\vec{c}_k|$ and the probability distribution is the same as SCUSUM.

This of course requires calculation of both θ and η. One way is

$$\begin{aligned} \eta^2 &= -\gamma_1 \\ \theta^2 &\approx s^2 + 2\gamma_1 = s^2 - 2\eta^2 \end{aligned} \tag{30}$$

where γ_1 is the *serial covariance at lag 1*

$$\gamma_1 = 1/(N-1) \sum_{k=1}^{n} (P_n - \bar{P})(P_{n+1} - \bar{P}) \tag{31}$$

and s^2 is the estimated variance of the observed periods

$$s^2 = 1/(N-1)\, [\Sigma(P_n - \bar{P})^2] \tag{32}$$

Another method is to compute variances for multiple periods (Blacher and Perdang

1988); the sum Q_k of k consecutive periods has a variance given by

$$<Q_k^2> - (k\mu)^2 = k\theta^2 + 2\eta^2 \tag{33}$$

By plotting the estimated variance of k-fold sums against k, and fitting a straight line, we can estimate both the period variance θ^2 and the variance of the time measurements η^2. These values are interesting in themselves: estimates of θ and η have been computed by this method for 391 Mira-type variable stars by Percy *et al.* (1990).

6. Sparse data

One of the great advantages of O-C analysis is that it does not require a complete observational sequence: it can be applied with only a scattering of observed times, as long as we are able to compute the cycle number of each observation. SCUSUM, as it stands, cannot be so applied, because it requires estimating the variance inherent in the periods themselves. A very incomplete observational sequence may not have any two consecutive times, and therefore may not provide any direct estimates of individual periods on which to base a variance computation. However, any two observed times representing *k* cycles will provide an estimate of the sum of *k* periods. On this basis we can estimate of the variance inherent in the periods.

Suppose we have *n+1* observed times T_a, a=0,1,2,...,n, together with their corresponding cycle numbers E_a. Let the first observed time correspond to cycle 0, the last to cycle *N* (so $E_0=0$, $E_n=N$). Then we can compute *n* period estimates *P* and their corresponding cycle counts *k*

$$\begin{aligned} k_a &= E_a - E_{a-1} \\ P_a &= (T_a - T_{a-1})/k_a \end{aligned} \tag{34}$$

Because the entire observational interval covers *N* cycles, we have

$$\Sigma\, k_a = N \tag{35}$$

Since each P_a is the average of k_a cycles,

$$\begin{aligned} <P_a> &= \mu \\ <P_a^2> &= \mu^2 + \theta^2/k_a + 2\eta^2/k_a^2 \end{aligned} \tag{36}$$

The average period is given, as before, by

$$P = (T_n - T_0)/N = \Sigma\, k_a P_a/N \tag{37}$$

and we can estimate the variance of the periods by

$$s^2 = \frac{\Sigma k_a P_a^{\,2} - NP^2}{n - 1} \tag{38}$$

The actual expected value of this estimate is

$$\langle s^2 \rangle = \theta^2 + 2\eta^2\,[\Sigma(1/k_a) - 1/N]/(n-1) \tag{39}$$

The more sparse the data, the smaller is the second term; s^2 becomes a good estimate of the period variance θ^2.

For sparse data, the maximum absolute SCUSUM value does *not* follow the same probability distribution as outlined above. The SCUSUM test is too severe, because most of the SCUSUM values are unobserved. For very sparse data, each separate SCUSUM can be treated as a normally distributed random variable with mean 0 and variance 1. However, the false alarm probability must be adjusted to account for the fact that several separate tests are made (one for each SCUSUM value). Usually the false alarm probability should be divided by the number of independent tests

$$\text{Prob(effective)} = \frac{\text{Prob(desired)}}{\text{(\# tests)}} \tag{40}$$

Because of the strong correlation among SCUSUMs, the number of "independent" tests (for adjusting the false alarm probability) is less than the number of SCUSUM values, and never more than 10. Further investigation is required on this particular.

References

Blacher, S., and Perdang, J. 1988, *Testing for Chaos in Long Period Variables*, in *Multimode Stellar Pulsations*, Konkoly Observatory, p. 283.

Eddington, A. S., and Plakidis, S. 1929, *Mon. Not. Roy. Astron. Soc.*, **90**, 65.

Isles, J. E., and Saw, D. R. B. 1989, *J. Brit. Astron. Assoc.*, **99**, 121.

Koen, C., and Lombard, F. 1993, *Mon. Not. Roy. Astron. Soc.*, **263**, 287.

Koen, C., and Lombard, F. 1993, *Mon. Not. Roy. Astron. Soc.*, **263**, 309.

Percy, J. R., Colivas, T., Sloan, W.B., and Mattei, J.A. 1990, *Long-term changes in Mira variables*, in *Confrontation between Stellar Pulsation and Evolution*, ASP conference series.

Pringle, J. 1975, *Mon. Not. Roy. Astron. Soc.*, **170**, 633.

Sterne, T. 1934, *Harvard Circ.* 386 and 387.

Williamson, R. J. 1985, *The Statistician*, **34**, 345.