Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **401,** 586–596 (2010) doi:10.1111/j.1365-2966.2009.15680.x

The Nyquist frequency for time series with slight deviations from regular spacing

Chris Koen★

Department of Statistics, University of the Western Cape, Private Bag X17, Bellville, 7535 Cape, South Africa

Accepted 2009 September 8. Received 2009 September 8; in original form 2009 May 17

ABSTRACT

The paper is based on the notion that the Nyquist frequency $\nu_{\rm N}$ is a symmetry point of the periodogram of a time series: the power spectrum at frequencies above $\nu_{\rm N}$ is a mirror image of that below $\nu_{\rm N}$. Koen showed that the sum $SS(\nu) = \sum_{k,\ell}[\sin 2\pi\nu(t_k - t_\ell)]^2$ (where t_k and t_ℓ range over the time points of observation) is zero when the frequency $\nu = \nu_{\rm N}$. This property is used to investigate the Nyquist frequency for data which are almost regularly spaced in time. For some configurations, there are deep minima of SS at frequencies $\nu_{\rm P} \ll \nu_{\rm N}$; such $\nu_{\rm P}$ are dubbed 'pseudo-Nyquist' frequencies: the implication is that most of the information about the frequency content of the data is available in the spectrum over $(0, \nu_{\rm P})$. Systematic simulation results are presented for two configurations – small random variations in respectively the time points of observation and the lengths of the intervals between successive observations. A few real examples of CCD time series photometry obtained over several hours are also discussed.

Key words: methods: statistical – stars: individual: HE 0230-4323.

1 INTRODUCTION

A time series of measurements $y_1, y_2, \ldots, y_{\rm N}$ is taken at times $t_1, t_2, \ldots, t_{\rm N}$; the t_i are, in general, not regularly spaced, i.e.

$$\Delta_j = t_{j+1} - t_j \tag{1}$$

is not constant. The notation is simplified if it assumed that the mean of the y_i is zero, hence this non-essential assumption is made. The periodogram of the y_i is defined as

$$I(\omega) = \frac{1}{N}\left[\left(\sum_{k=1}^{N} y_k \cos\omega t_k\right)^2 + \left(\sum_{k=1}^{N} y_k \sin\omega t_k\right)^2\right], \tag{2}$$

where $\omega = 2\pi\nu$ is the angular frequency.

To set the scene, consider the case of constant time spacing between observations, i.e. $\Delta_j = d (j = 1, 2, \ldots, N-1)$. Conventionally, the periodogram (2) is then calculated over the frequency interval $0 \le \nu \le 0.5/d$ (or some subinterval thereof). It is often said that the Nyquist frequency $\nu_{\rm N} = 0.5d^{-1}$ is the highest frequency component which can be extracted from the data. This is not quite true – more accurately, it is the high-frequency limit of *one* interval over which $I(\nu)$ is uniquely defined. It is not difficult to show that exactly the same periodogram is obtained over an infinite number of different frequency intervals. Let $I_0(\nu)$ be defined over $0 \le \nu \le \nu_{\rm N}$, then

$$\begin{aligned}
I(\nu) &\equiv I_0(2\nu_{\rm N} - \nu) \quad \nu_{\rm N} \le \nu \le 2\nu_{\rm N}\\
I(\nu) &\equiv I_0(\nu - 2\nu_{\rm N}) \quad 2\nu_{\rm N} \le \nu \le 3\nu_{\rm N}\\
I(\nu) &\equiv I_0(4\nu_{\rm N} - \nu) \quad 3\nu_{\rm N} \le \nu \le 4\nu_{\rm N}\\
I(\nu) &\equiv I_0(\nu - 4\nu_{\rm N}) \quad 4\nu_{\rm N} \le \nu \le 5\nu_{\rm N}\\
\ldots &\qquad \ldots
\end{aligned} \tag{3}$$

or, more succintly,

$$I(\nu) = \begin{cases} I_0(2k\nu_{\rm N} - \nu) & (2k-1)\nu_{\rm N} \le \nu \le 2k\nu_{\rm N}\\ I_0(\nu - 2k\nu_{\rm N}) & 2k\nu_{\rm N} \le \nu \le (2k+1)\nu_{\rm N}, \end{cases} \tag{4}$$

where k is any positive integer. As pointed out by Eyer & Bartholdi (1999), physical considerations determine which of the intervals is applicable – in practice, observations are usually planned such that any frequency of interest lies in the interval $(0, 0.5d^{-1})$.

Eyer & Bartholdi (1999) considered the case of irregular time spacing, and showed that the Nyquist frequency is given by $\nu_{\rm N} = 0.5d_*^{-1}$, where d_* is the greatest common divisor of the time intervals in (1). For example, if there is an underlying regularity to the times of measurement, characterized by an interval Δt, then $\nu_{\rm N} = 0.5/\Delta t$, even if observations are only obtained sporadically (i.e. if there are many 'missing values'). If the measurements are completely random, then d_* is the accuracy with which time is measured (i.e. the effective unit of time; see Koen 2006).

Koen (2006) supplied a calculation formula to be solved for the Nyquist frequency, for arbitrary time spacings of measurements:

★E-mail: ckoen@uwc.ac.za

the smallest positive root of

$$SS(\nu) = \sum_{\ell=1}^{N-1} \sum_{k=\ell+1}^{N} [\sin 2\pi\nu(t_k - t_\ell)]^2 = 0\,. \quad (5)$$

This will be used below to investigate a situation which is intermediate to the two extremes discussed above, namely the case of small random deviations from regularity in the times between observations. Two plausible theoretical time spacing configurations are discussed in Section 2, and applications to real data follow in Section 3. The contents of the paper are summarized in Section 4.

2 THE PSEUDO-NYQUIST FREQUENCY

Two situations are considered, which are, at a first glance, almost identical:

(i) There is a small random element to the timing of each observation, i.e.

$$t_j = j\Delta t + \epsilon_j, \quad (6)$$

where Δt is fixed and the ϵ_j are random numbers with a small (common) variance.

(ii) There is a small random element to each interval between successive observations:

$$t_j = t_{j-1} + \Delta t + \epsilon_j. \quad (7)$$

Case (i) could be one in which, for example, the intervals between observations are perfectly regular, but there are small recording errors in the times. Case (ii) could, for example, model a situation in which there are small random variations in exposure and/or readout times in time series photometry.

It is assumed that the random elements ϵ_j are Gaussian, with zero mean, and variance σ_ϵ^2. Simulation results presented below all have $\Delta t = 1$; $N = 100$ unless otherwise specified.

2.1 Case (i)

In the first example only, time is specified only to two decimal places, hence the maximum value of the Nyquist frequency, which applies to completely irregular time spacing of measurements, is $\nu_N = 0.5/0.01 = 50$. Fig. 1 shows simulated scatter in $\Delta_j = t_{j+1} - t_j$,

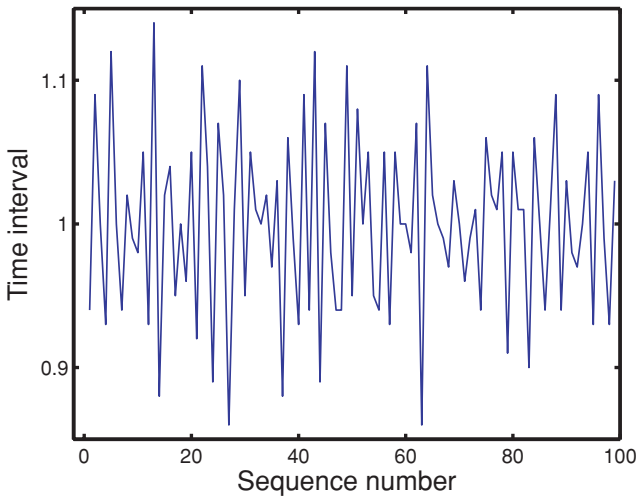

Figure 1. The intervals between successive time points for a Case (i) simulation with $\sigma = 0.05$.

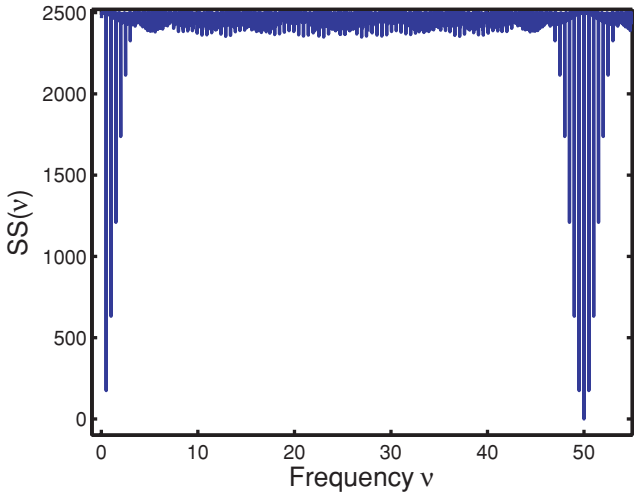

Figure 2. The function *SS* in equation (5), for the observation times corresponding to Fig. 1.

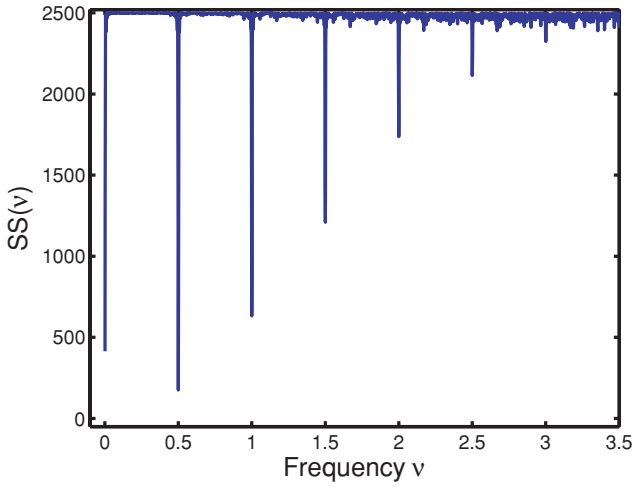

Figure 3. The low-frequency part of Fig. 2.

assuming $\sigma_\epsilon = 0.05$. The corresponding function *SS* in (5) is plotted in Fig. 2. The function is zero at $\nu = 50$ – the Nyquist frequency. Note though, the presence of deep local minima, particularly at small frequencies ($\nu < 5$). The low-frequency interval is shown in more detail in Fig. 3. Although the sum-of-squares *SS* is not zero in the point $\nu = 0.5$, it is quite small: the ratio $D = SS_{\rm min}/SS_{\rm max}$ is 0.07.

Consider a candidate Nyquist frequency ν_*, and let I_0 be the periodogram, calculated over the interval $0 \le \nu \le \nu_*$. The sum-of-squares $SS(\nu_*)$ defined in (5) is then a measure of the difference between $I(\nu)$ ($\nu_* \le \nu \le 2\nu_*$) and $I_0(2\nu_* - \nu)$ (cf. the first line of equation 3; Koen 2006). If $\nu_* = \nu_N$, then the two periodograms are identical and $SS = 0$. On the other hand, if *SS* is non-zero but small, it means that there is good, albeit not perfect, agreement between the two spectra. The point is illustrated by Fig. 4, which compares spectra over $0 \le \nu \le \nu_* = 0.5$ and $0.5 \le \nu \le 1$: the spectra are almost mirror images, but there are some small differences. (The 'measurements' were 100 Gaussian random values, taken at the time points giving rise to Figs 1–3.)

The correlation between the two periodograms defined above is 0.96 – it would have been unity if $\nu = 0.5$ had been the true Nyquist frequency. Given the very high correlation, a frequency such as

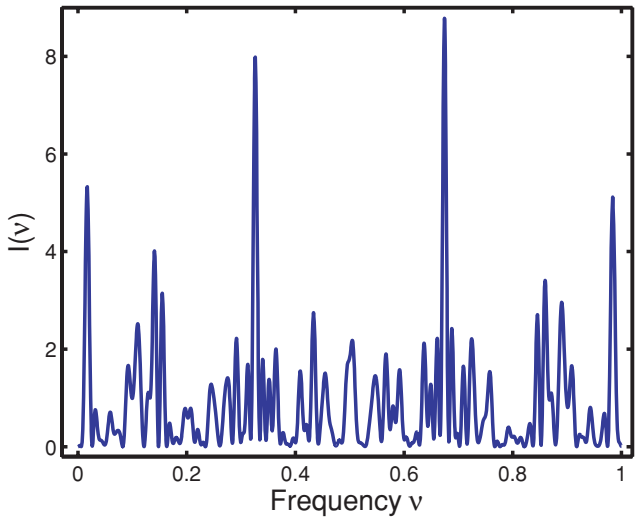

Figure 4. The periodogram, as defined in (2), for white noise defined on the time points giving rise to Figs 1–3. Note the near symmetry around the pseudo-Nyquist frequency $\nu_P = 0.5$.

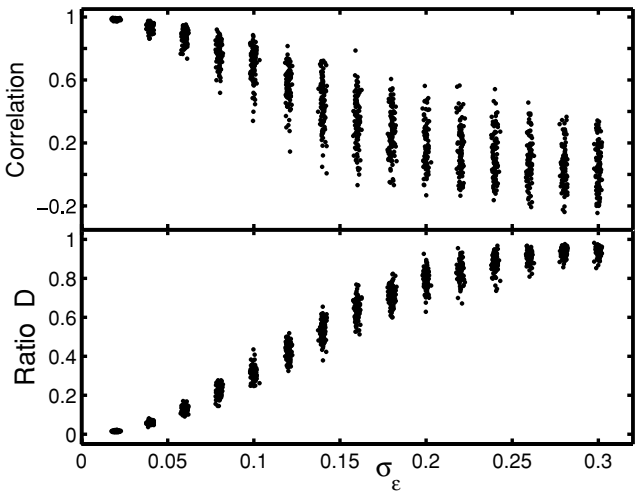

Figure 5. The top panel shows the correlation between periodograms calculated over the frequency intervals $(0, \nu_P)$ and $(2\nu_P, \nu_P)$, as a function of deviation from regularity of the time points. The bottom panel shows $SS(\nu_P)$, normalized by the maximum of SS over $(0, 2\nu_P)$. Calculations were performed for $\sigma_\epsilon = 0.02(0.02)0.30$. At a given σ_ϵ, plotted values have been jittered slightly to show the vertical scatter of points more clearly. For each value of σ_ϵ, 100 white noise samples, each of size $N = 100$, were generated.

$\nu = 0.5$ in this example will be referred to as a 'pseudo-Nyquist frequency' ν_P in what follows.

It may be expected that the smaller σ_ϵ (compared to Δt) the higher the correlation between the two spectra in the first line of (3) will be. A simulation experiment was conducted to obtain quantitative results: 15 values of σ_ϵ were chosen in the range 0.02 to 0.30. For each of these 100 sets of time points, with associated white noise 'observations' y_j were generated. The pseudo-Nyquist frequency ν_P near 0.5 was determined by locating the local minimum of $SS(\nu)$, and the periodograms over $(0, \nu_P)$ and $(2\nu_P, \nu_P)$ were correlated. The results can be seen in the top panel of Fig. 5. (Values of σ_ϵ have been jittered slightly horizontally to make the vertical spread of values in the diagram clearer.) Note that the spread is caused by variations in the configurations of time points t_j, variations in the

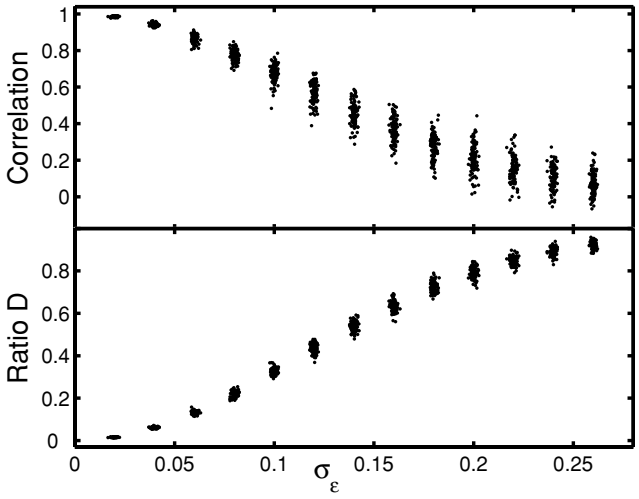

Figure 6. As for Fig. 5, but for sample sizes $N = 500$.

'observations' y_j and in the values of ν_P. The corresponding ratios $D = SS_{min}/SS_{max}$ are plotted in the bottom panel of the figure.

An implication of Fig. 5 is that for $\sigma_\epsilon < 0.05$, ν_P functions effectively as a true Nyquist frequency, while for $\sigma_\epsilon > 0.2$ or so, the time points are essentially randomly distributed, and $D = D(\nu)$ is small only in the true Nyquist frequency $\nu = \nu_N$.

Fig. 6, for $N = 500$, shows that the spread in correlation values for given σ_ϵ decreases with increasing sample size. There is also a suggestion that the mean correlation decreases to zero faster with increasing σ_ϵ.

A technical issue is mentioned in passing: for the largest values of σ_ϵ in Figs 5 and 6, the simulated values of ϵ are often larger than unity in absolute value: such data were rejected to avoid ambiguity in the simulated time points. Strictly speaking, random values were, therefore, drawn from a truncated, rather than a standard, normal distribution.

If ν_P is fixed at ν_* (the exact Nyquist frequency obtained when $\sigma_\epsilon = 0$), then relatively simple analytical expressions for the mean and variance of $SS(\nu_*)$ can be derived. It is shown in Appendix A that

$$\begin{aligned} \mathrm{E}SS(\nu_*) &= \frac{1}{4}N(N-1)(1-\mathrm{e}^{-4q}), \\ \mathrm{var}[SS(\nu_*)] &= \frac{1}{16}N(N-1) \\ &\quad \times \left[(1-\mathrm{e}^{-8q})^2 + 2(N-2)\mathrm{e}^{-4q}(1-\mathrm{e}^{-4q})^2\right], \end{aligned} \tag{8}$$

where $q = (2\pi\nu_*\sigma_\epsilon)^2$. Fig. 7 shows the dependence of $\mathrm{E}SS(\nu_*)/N^2$ and $\mathrm{var}[SS(\nu_*)]/N^3$ on σ_ϵ, for ν_* fixed at 0.5, and three different values of N. Clearly, $\mathrm{E}SS(\nu_*) \propto N^2$ and $\mathrm{var}[SS(\nu_*)] \propto N^3$ to very good accuracy, at least for $N \geq 100$. These results may have been anticipated from the dependences on N in (8).

The results in (8) can be related to those in the lower panels of Figs 5 and 6 by noting that

$$D = SS_{min}/SS_{max} \approx SS(\nu_*)/[0.25N^2] .$$

2.2 Case (ii)

Plots of $SS(\nu)$ around $\nu_P = 0.5$ show that, for given σ_ϵ, the local minima are less deep than for simulated Case (i) data. Also, ν_P is not as sharply defined as for Case (i) data. Fig. 8 illustrates this

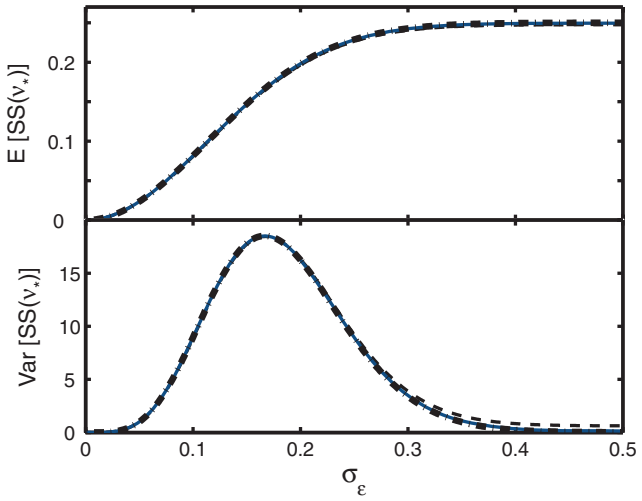

Figure 7. The expected value (top panel) and variance (bottom panel) of the sum of squares in equation (5), evaluated in a fixed pseudo-Nyquist frequency $\nu_* = 0.5$, for Case (i) of Section 2. Although not readily distinguishable, results are shown for three sample sizes: $N = 100$ (broken lines); $N = 500$ (solid lines); and $N = 5000$ (dotted lines). Expected values have been scaled by N^2, and variances by N^3.

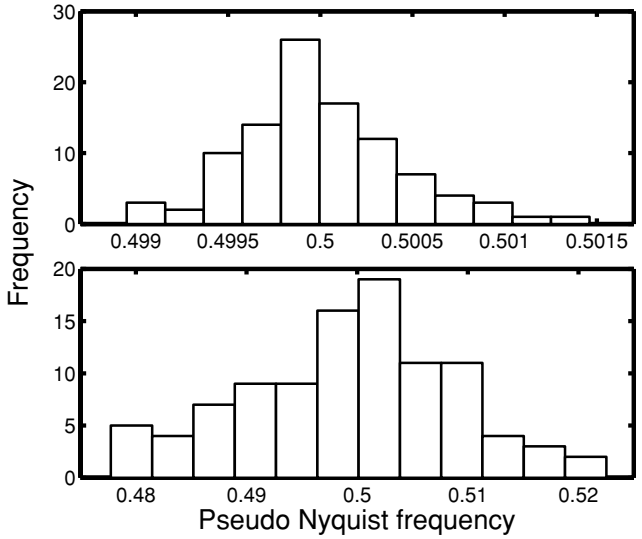

Figure 8. The distribution of pseudo-Nyquist frequencies for 100 Case (i) data sets with $\sigma_\epsilon = 0.2$ (top) and 100 Case (ii) data sets with $\sigma_\epsilon = 0.1$.

for 100 data sets for each of Case (i) ($\sigma_\epsilon = 0.2$) and Case (ii) ($\sigma_\epsilon = 0.1$). The correlation between spectra with symmetry point ν_P is 0.23 for both these sets of data. Note that the spread of ν_P values is more than an order of magnitude larger for Case (ii) data sets.

The Case (ii) analogue of Fig. 5 is given in Fig. 9. For a given value of σ_ϵ, the mean correlation is smaller, and the spread greater, for the simulated Case (ii) data. Also, the minima in *SS*, as measured by the ratio *D*, are shallower for Case (ii) data.

Analytical formulae for $ESS(\nu_*)$ and $\mathrm{var}[SS(\nu_*)]$, analogous to (8), are derived in Appendix B (equations B3, B7 and B8), but these are not transparent. It is again found that $ESS(\nu_*) \propto N^2$, but the dependence of the variance on N is stronger than for Case (i) – $\mathrm{var}[SS(\nu_*)] \propto N^4$. Some results are plotted in Fig. 10, for $\nu_* = 0.5$, and three different values of N.

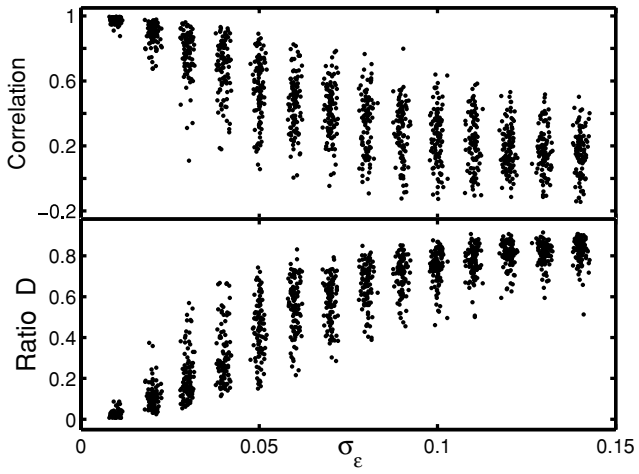

Figure 9. As for Fig. 5, but for Case (ii) models with $\sigma_\epsilon = 0.01(0.01)0.14$.

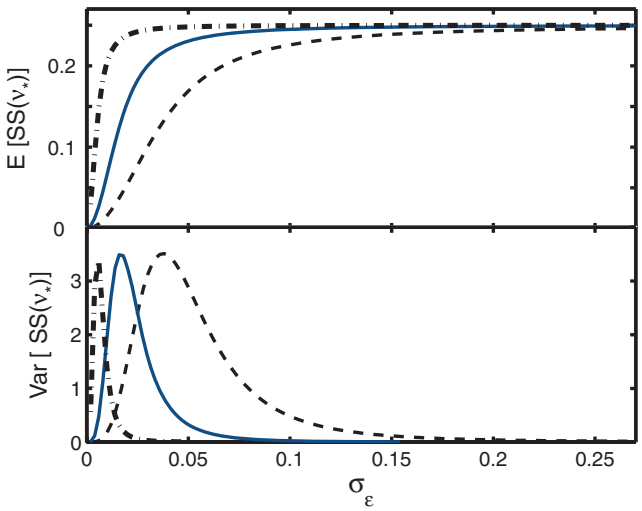

Figure 10. As for Fig. 7, but for Case (ii) of Section 2. The scaling of expected values by sample size N is the same as for Case (i) (i.e. N^2), but variances have been scaled by N^4, and then multiplied by 1000 (to facilitate plotting).

This section of the paper is closed with a word on how to ascertain which case a particular observed data set might belong to. This is easily done by comparing observed and predicted time points in an 'observed-calculated' $(O - C)$ diagram:

$$(O - C)(j) = t_j - j\overline{\Delta t}, \tag{9}$$

where $\overline{\Delta t}$ is the mean interval between successive observed time points. Simulated $O - C$ diagrams for the two data types are compared in Fig. 11: the graph for Case (ii) data is a random walk, considerably smoother than that for Case (i).

3 REAL EXAMPLES

The first set of illustrations is based on data acquired during three observing runs on the hot subdwarf star HE 0230-4323 (Koen 2007). The times between successive *B*-band exposures are plotted in Fig. 12. Exposures through a *V* filter were interleaved with those in *B*, and filter changes also required a few seconds. The top panel of the figure demonstrates the results of an uneventful run – only slight irregularities in the times between exposures are evident.

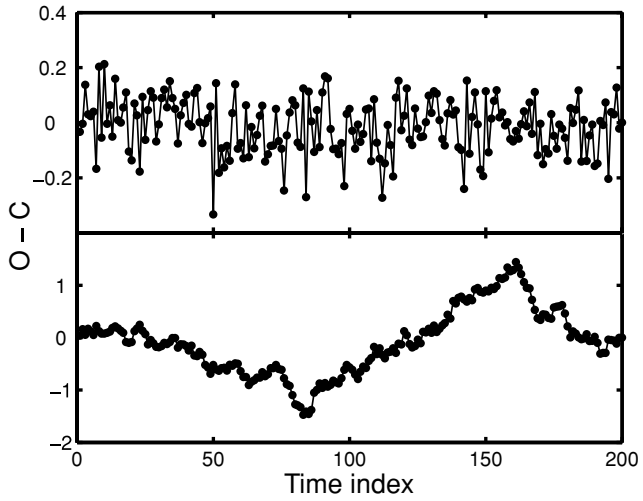

Figure 11. 'Observed minus predicted' diagrams for Case (i) (top panel) and Case (ii) (bottom panel) time points, respectively. The variance $\sigma_\epsilon^2 = 0.01$, and $N = 200$, for both data sets.

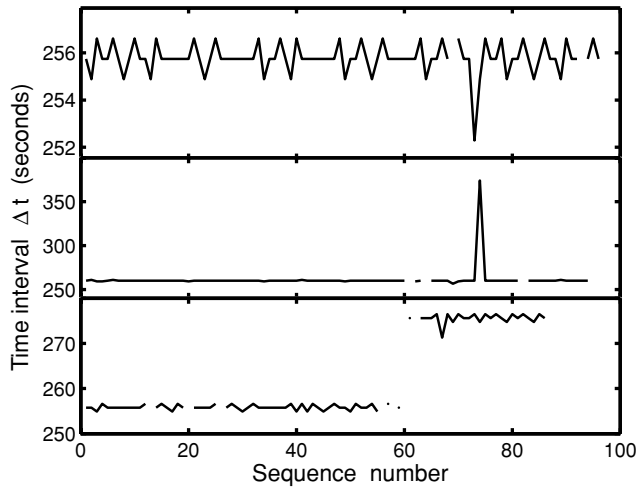

Figure 12. Three typical time interval configurations encountered in CCD time series photometry. Top panel: slight irregularities in the time intervals between measurements. Middle panel: as for the top panel, but with a single aberrant irregularity due to a software restart. Bottom panel: as for the top panel, but with a permanent change of exposure times. Breaks in the plots are due to missing observations: these do not affect the Nyquist frequency.

It is interesting that, within the accuracy of the times recorded, only a few discrete time interval values occur. The effect of a ∼100 s interruption in the observing sequence is illustrated in the middle panel. The time intervals plotted in the bottom panel were generated by a switch in exposure times, necessitated by changes in weather conditions.

Results for the three data sets are given in Table 1: note that the brightness measurements of HE 0230-4323 were used in the calculations. For the time spacing plotted in the top panel of Fig. 12, ν_P can essentially be treated as a true Nyquist frequency. In the case of the spacing in the bottom panel of Fig. 12, the correlation between spectra on either side of ν_P is too small for the periodogram to be considered unique over $(0, \nu_P)$. Results for the data in the middle panel are apparently intermediate.

It is interesting to compare ν_P with the frequency

$$\nu_1 = 0.5/\overline{\Delta t}\ , \tag{10}$$

Table 1. Parameters derived for the data sets in Fig. 12.

Data set	ν_P	ν_1	ρ	D
Slight irregularities	168.91	163.78	1.00	0.0007
One outlier	165.40	158.66	0.92	0.46
Change of spacing	169.00	151.60	0.22	0.53

Note. ν_P is a low frequency candidate substitute for the true Nyquist frequency; ν_1 is based on the mean time spacing (see equation 10); ρ is the correlation between periodograms defined on intervals separated by ν_P; and D measures the depth of the sum of squares SS in the frequency ν_P.

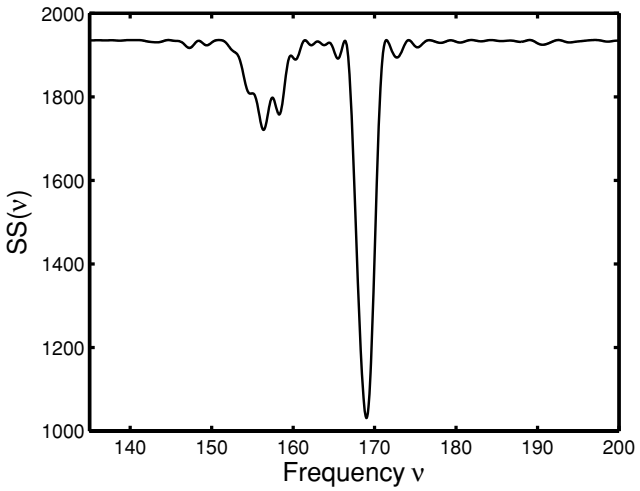

Figure 13. The sum of squares defined in equation (5), for the data set with a change in times between successive observations (bottom panel of Fig. 12). Two closely spaced pseudo-Nyquist frequencies are obtained.

where $\overline{\Delta t}$ is the mean time spacing between measurements. It has been speculated in the astronomy literature that this is an appropriate 'Nyquist frequency' for irregularly spaced data. It is clear from the entries in the table that ν_1 as defined in (10) is an inaccurate estimator for ν_P.

For the first data set, which has an underlying time spacing which is close to regular, ν_1 can be adjusted to take account of the fact that some of the Δt are artificially long because a few poor data points have been excised. The result is a frequency of 168.90 – very close to ν_P.

There are two issues of particular interest raised by the information in the table. The first is the fact that the pseudo-Nyquist frequency of the third data set differs little from the first two, despite the change in times between measurements. Fig. 13 shows that two distinct minima in SS are in fact obtained – but the one near $\nu = 169$ is considerably deeper.

The second point is an apparent discrepancy between the values of D and the correlation coefficients for the data with a single outlying time spacing: the correlation is high, despite the fact that D is relatively large. Inspection of Fig. 14, a plot of $I(\nu_P + \nu)$ against $I(\nu_P - \nu)$ $(0 < \nu < \nu_P)$, reveals the reason for the large correlation coefficient: a reasonable correspondence between the largest peaks in the two periodograms. If both spectra are truncated so that only periodogram values below 300 are taken into account, the correlation coefficient drops to 0.53. A short simulation experiment was also conducted using white noise as 'measurements' – only modest periodogram peaks are expected for such data. Correlation coefficients in the range 0.22–0.77 were obtained, for 100 different 'data

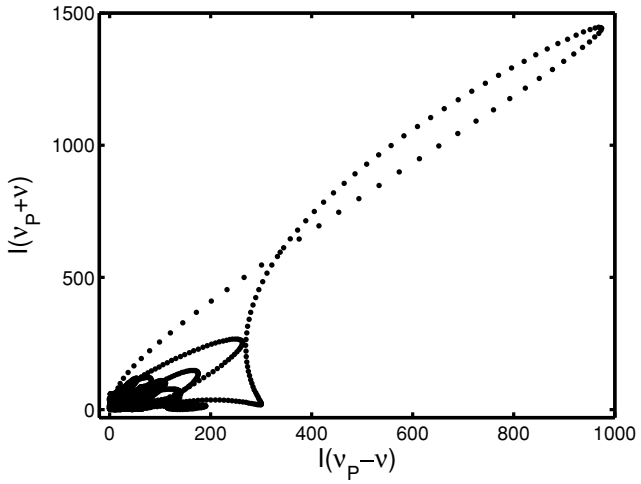

Figure 14. The relationship between periodograms over (ν_P, 0) and (ν_P, $2\nu_P$), for the data set with a single outlying time spacing (middle panel of Fig. 12).

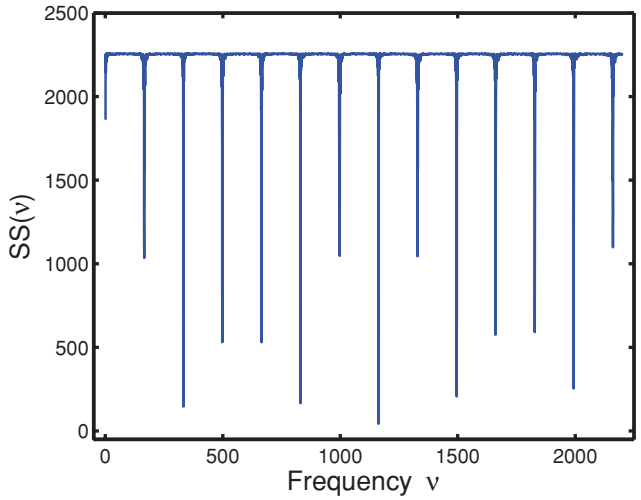

Figure 15. The sum of squares defined in (5), for the time spacings in the middle panel of Fig. 12; see also the second line in Table 1, and Table 2.

sets'. In general, the ratio D is therefore probably a better measure of whether a candidate pseudo-Nyquist frequency is useful than a single correlation coefficient.

A broader search for pseudo-Nyquist frequencies for the 'one outlier' data set gives interesting results. In Fig. 15, the sum of squares SS is plotted over a wide frequency interval. There is an array of deep local minima, spaced roughly 165 d^{-1} apart. Furthermore, there is a symmetrical arrangement of depths of the minima, with symmetry point $\nu \sim 1200$. The first seven minima are listed in Table 2.

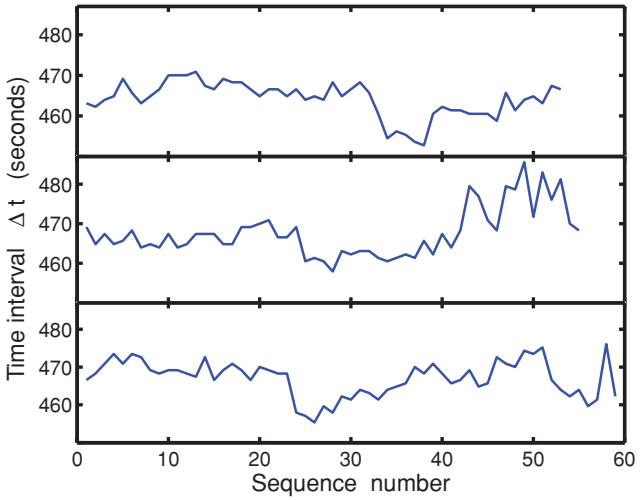

Figure 16. The times between successive observations of fields in the open cluster IC 2391. Data for three different nights are shown.

Table 3. Parameters derived for the data sets in Fig. 16.

Panel in Fig. 16	ν_P	ν_1	ρ	D	$\sigma(\Delta t)$
Top	93.10	93.08	0.86	0.13	4.3
Middle	92.64	92.38	0.84	0.15	6.1
Bottom	92.62	92.50	0.93	0.072	4.7

Note. The correlation coefficients are mean values derived from 100 synthetic white noise data sets. The last column gives a measure of the scatter in the plotted time intervals; other symbols are defined in Table 2.

Scrutiny of the table, and Fig. 15, suggests that $\nu = 1163$ d^{-1} is very close to being a true Nyquist frequency. Using $\nu = 332.2$ d^{-1} would lead to a slight loss of information.

Data for the second set of illustrations are taken from observations of stars in the cluster IC 2391 (Koen & Ishihara 2006). Telescope pointing was rotated between two fields in the cluster throughout each of the three observing nights – this accounts for the rather erratic intervals between measurements (Fig. 16).

The results of the analysis are presented in Table 3. The correlations were derived by associating random white noise 'data' with each of the time points; the values in the table are means over 100 simulated data sets.

The agreement between ν_1 and ν_P is much better than for the data in Table 1: one reason is that there are no censored time points amongst the IC 2391 data. It is curious that the correlation coefficient for the bottom data set in Fig. 16 is larger than for the top data set, despite the fact that the scatter in the time intervals is slightly larger.

4 SUMMARY

(i) If measurements are regularly spaced in time, an interval Δt apart, then the Nyquist frequency is $\nu_N = 0.5/\Delta t$. This frequency is

Table 2. Properties of the local minima in Fig. 15.

ν_P	165.40	332.54	497.93	665.07	830.47	997.60	1163.01
D	0.46	0.063	0.23	0.23	0.073	0.47	0.019
Real data ρ	0.92	0.96	0.87	0.82	0.96	0.67	0.99
White noise ρ	0.55	0.94	0.76	0.76	0.93	0.52	0.98

Note. Two correlation coefficients ρ were calculated: one for the actual HE 0230-4323 observations, the other being the mean obtained for 100 simulated white noise data sets.

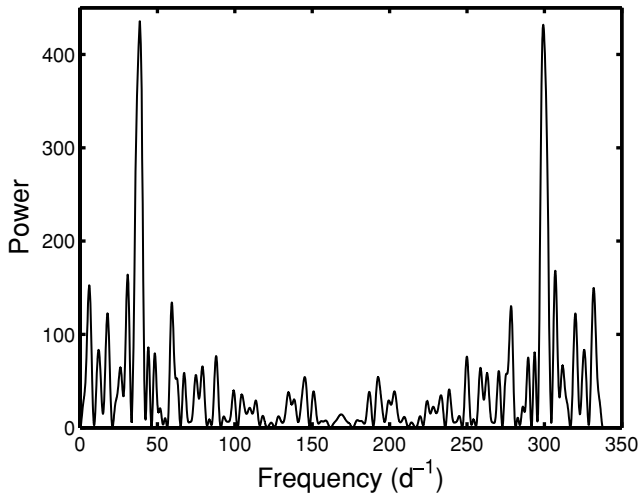

Figure 17. Periodogram of observations of the star HE 0230-4323. The time spacing of the measurements is plotted in the top panel of Fig. 12; see also the first line of Table 1.

not affected by missing values: provided that there is an underlying regularity, large gaps in the data make no difference to the Nyquist frequency.

(ii) At the other end of the scale, for completely random observations, the Nyquist frequency is also determined by the largest common divisor of the Δt_j: in this case, it would usually be the smallest unit of time recorded. As an example, if the Julian date of observation is recorded to five decimal places, then $\nu_{\rm N} = 50\,000$ d^{-1}.

(iii) The results in Section 2 address intermediate situations, in which there are small variations in the Δt_j. Two cases were distinguished: random variations in the time points of measurement, and random variations in the intervals between measurements. For each, the potential usefulness of the pseudo-Nyquist frequency $\nu_{\rm P} \sim 0.5/\overline{\Delta t}$ was investigated. For Case (i), $\sigma_{\rm P}$ could be used as a Nyquist frequency for scatter less than about 5 per cent of $\overline{\Delta t}$; for Case (ii) data the limitation is more severe. These results were derived for $N = 100$; numerical results for $N = 500$ were similar.

(iv) Real data sets with small irregularities were discussed in Section 4. Provided there are no discrepant time points, such observations may have pseudo-Nyquist frequencies which may be used as true Nyquist frequencies (e.g. the first HE 0230-4323 data set – see also below). For somewhat larger scatter, as in the IC 2391 data, the pseudo-Nyquist frequency does not seem to be useful.

(v) An intriguing discovery of this paper is the result that, for some time spacing configurations, pseudo-Nyquist frequencies which could function as true Nyquist frequencies may be found at approximate multiples of $0.5/\overline{\Delta t}$ (Fig. 15). This seems to be associated with data containing time points which are aberrant, i.e. do not belong to the approximately regular spacing of the rest. None of the data studied in this paper, aside from the 'one outlier' set in Table 1, showed this behaviour: for the rest, the dependence of *SS* on ν resembled that shown in Fig. 3.

(vi) The combined results of Sections 2 and 3 show that for a given data set with timing irregularities, it may not be possible to draw definite conclusions about the location of pseudo-Nyquist frequencies, nor about the degree of symmetry around these, without careful investigation. It is recommended that, where feasible, simulation be used to this end.

The paper is closed with a very real application. It was noted in the previous section that $\nu_{\rm P} = 168.9$ d^{-1} is essentially a real Nyquist frequency for the first of the HE 0230-4323 data sets. The periodogram of the data is plotted in Fig. 17, over an interval of twice $\nu_{\rm P}$. The excellent agreement between the spectra on either side of the Nyquist frequency is evident. Koen (2007) only studied the low-frequency part of the spectrum, and concluded that HE 0230-4323 is a slowly pulsating star with a prominent $\sim$39 d^{-1} mode. Subsequent investigation, using very short exposures (10 s), revealed that the wrong alias had been identified: the star is in fact a rapid pulsator with periods of the order of a few minutes.

REFERENCES

Eyer L., Bartholdi P., 1999, A&AS, 135, 1
Koen C., 2006, MNRAS, 371, 1390
Koen C., 2007, MNRAS, 377, 1275
Koen C., Ishihara A., 2006, MNRAS, 369, 846

APPENDIX A: SOME STATISTICAL PROPERTIES OF *SS* FOR CASE (i) OF SECTION 2

Consider a frequency ν_* which would have been a Nyquist frequency if the time spacings had been regular (i.e. if $\epsilon_j \equiv 0$ had held in equation 6). Then,

$$\sin 2\pi\nu_*(k-\ell) = 0 \qquad \cos 2\pi(k-\ell)\nu_* = \pm 1$$

for all integer k, ℓ. It follows that

$$SS(\nu_*) = \sum_{\ell=1}^{N-1} \sum_{k=\ell+1}^{N} [\sin 2\pi\nu_*(\epsilon_k - \epsilon_\ell)]^2 \, . \tag{A1}$$

The primary interest is in the mean and variance of $SS(\nu_*)$. One way of obtaining these is through the probability density function (PDF) of $SS(\nu_*)$: it is not difficult to show that the PDF of the individual terms

$$y = \sin^2 2\pi\nu_*(\epsilon_k - \epsilon_\ell) \tag{A2}$$

in (A1) is

$$f(y) = \frac{1}{\sigma_*\sqrt{y(1-y)}} \left\{ \phi\left(-\frac{\arcsin\sqrt{y}}{\sigma_*}\right) + \sum_{j=1}^{\infty} \left[\phi\left(\frac{j\pi + \arcsin\sqrt{y}}{\sigma_*}\right)\right] + \left[\phi\left(\frac{j\pi - \arcsin\sqrt{y}}{\sigma_*}\right)\right] \right\}, \tag{A3}$$

where $\sigma_*^2 = 2(2\pi\nu_*\sigma)^2 = \mathrm{var}[2\pi\nu_*(\epsilon_k - \epsilon_\ell)]$ and

$$\phi(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right).$$

However, the PDF of the sum of terms in (A1) is not readily written down, due to the interdependence of the terms in (A1). We therefore proceed to evaluate the mean and variance of $SS(\nu_*)$ directly from the Gaussian distribution of the arguments $2\pi\nu_*(\epsilon_k - \epsilon_\ell)$.

Some useful results are first given. Using the standard integral

$$\int_0^\infty \mathrm{e}^{-ax^2} \cos bx dx = \frac{1}{2}\sqrt{\frac{\pi}{a}}\mathrm{e}^{-b^2/4a}$$

and the trigonometric identities

$$\begin{aligned}
\sin^2 x &= \frac{1}{2}(1 - \cos 2x) \\
\cos^2 x &= \frac{1}{2}(1 + \cos 2x) \\
\sin^2 x \cos^2 x &= \frac{1}{8}(1 - \cos 4x) \\
\sin^4 x &= \frac{1}{8}(3 - 4\cos 2x + \cos 4x) \\
\cos^4 x &= \frac{1}{8}(3 + 4\cos 2x + \cos 4x)
\end{aligned}$$

it is not difficult to show that, for y as defined in (A2),

$$\begin{aligned}
\mathrm{E}\sin^2 y &= \left(1 - \mathrm{e}^{-2\sigma_*^2}\right)/2 \\
\mathrm{E}\cos^2 y &= \left(1 + \mathrm{e}^{-2\sigma_*^2}\right)/2 \\
\mathrm{E}\sin^2 y \cos^2 y &= \left(1 - \mathrm{e}^{-8\sigma_*^2}\right)/8 \\
\mathrm{E}\sin^4 y &= \left(3 - 4\mathrm{e}^{-2\sigma_*^2} + \mathrm{e}^{-8\sigma_*^2}\right)/8 \qquad \text{(A4)} \\
\mathrm{E}\cos^4 y &= \left(3 + 4\mathrm{e}^{-2\sigma_*^2} + \mathrm{e}^{-8\sigma_*^2}\right)/8.
\end{aligned}$$

Furthermore, expectations of variables such as $\sin y$, $\sin^3 y$, $\sin y \cos y$, etc., are zero, since these are uneven functions of y. Finally, since the $\epsilon_j (j = 1, 2, \ldots, N)$ are assumed independent, it follows that for distinct ℓ, k, r, s

$$\mathrm{cov}[\sin^2 \omega_*(\epsilon_k - \epsilon_\ell), \sin^2 \omega_*(\epsilon_s - \epsilon_\ell)] = 0,$$

where $\omega_* = 2\pi\nu_*$.

The expected value (mean) of $SS(\nu_*)$ follows immediately from the first relation in (A4) as

$$\mathrm{E}SS(\nu_*) = \frac{1}{4}N(N-1)\left(1 - \mathrm{e}^{-2\sigma_*^2}\right). \qquad \text{(A5)}$$

Calculation of the variance is more involved:

$$\begin{aligned}
\mathrm{var}[SS(\nu_*)] &= \mathrm{cov}\left[\sum_{\ell=1}^{N-1}\sum_{k=\ell+1}^{N} \sin^2 \omega_*(\epsilon_k - \epsilon_\ell), \sum_{r=1}^{N-1}\sum_{s=r+1}^{N} \sin^2 \omega_*(\epsilon_s - \epsilon_r\right] \\
&= \sum_{\ell=1}^{N-1}\sum_{k=\ell+1}^{N} \mathrm{cov}\left[\sin^2 \omega_*(\epsilon_k - \epsilon_\ell), \sum_{s=\ell+1;s\neq k}^{N} \sin^2 \omega_*(\epsilon_s - \epsilon_\ell) + \sum_{s=k+1}^{N} \sin^2 \omega_*(\epsilon_s - \epsilon_k)\right. \\
&\quad \left. + \sum_{r=1;r\neq\ell}^{k-1} \sin^2 \omega_*(\epsilon_k - \epsilon_r) + \sum_{r=1}^{\ell-1} \sin^2 \omega_*(\epsilon_\ell - \epsilon_r) + \sin^2 \omega_*(\epsilon_k - \epsilon_\ell)\right]. \qquad \text{(A6)}
\end{aligned}$$

A typical covariance term is

$$\begin{aligned}
&\mathrm{cov}[\sin^2 \omega_*(\epsilon_k - \epsilon_\ell), \sin^2 \omega_*(\epsilon_s - \epsilon_\ell)] \\
&\quad = \mathrm{E}[\sin^2 \omega_*(\epsilon_k - \epsilon_\ell) \sin^2 \omega_*(\epsilon_s - \epsilon_\ell)] - [\mathrm{E}\sin^2 \omega_*(\epsilon_k - \epsilon_\ell)][\mathrm{E}\sin^2 \omega_*(\epsilon_s - \epsilon_\ell)] \\
&\quad = \mathrm{E}[\sin \omega_*\epsilon_k \cos \omega_*\epsilon_\ell - \cos \omega_*\epsilon_k \sin \omega_*\epsilon_\ell]^2[\sin \omega_*\epsilon_s \cos \omega_*\epsilon_\ell - \cos \omega_*\epsilon_s \sin \omega_*\epsilon_\ell]^2 - \frac{1}{4}(1 - \mathrm{e}^{-4q})^2 \\
&\quad = \mathrm{E}[\sin^4 \omega_*\epsilon_k \cos^2 \omega_*\epsilon_\ell \cos^2 \omega_*\epsilon_s + \cos^4 \omega_*\epsilon_k \sin^2 \omega_*\epsilon_\ell \sin^2 \omega_*\epsilon_s + \sin^2 \omega_*\epsilon_k \cos^2 \omega_*\epsilon_\ell \cos^2 \omega_*\epsilon_k \sin^2 \omega_*\epsilon_s \\
&\qquad + \cos^2 \omega_*\epsilon_k \sin^2 \omega_*\epsilon_\ell \sin^2 \omega_*\epsilon_k \cos^2 \omega_*\epsilon_s] - \frac{1}{4}(1 - \mathrm{e}^{-4q})^2,
\end{aligned}$$

where $q \equiv \sigma_*^2/2$. Using (A4), and simplifying,

$$\mathrm{cov}[\sin^2 \omega_*(\epsilon_k - \epsilon_\ell), \sin^2 \omega_*(\epsilon_s - \epsilon_\ell)] = \frac{1}{8}\mathrm{e}^{-4q}(1 - \mathrm{e}^{-4q})^2 \quad s \neq k \tag{A7}$$

is obtained. It is straightforward to also show that, if $k = s$,

$$\mathrm{cov}[\sin^2 \omega_*(\epsilon_k - \epsilon_\ell), \sin^2 \omega_*(\epsilon_k - \epsilon_\ell)] = \frac{1}{8}(1 - \mathrm{e}^{-8q})^2. \tag{A8}$$

Substitution of (A7) and (A8) into (A6), then leads to

$$\begin{aligned}\mathrm{var}[SS(\nu_*)] &= \sum_{\ell=1}^{N-1}\sum_{k=\ell+1}^{N}\left[\frac{(2N-4)}{8}\mathrm{e}^{-4q}(1-\mathrm{e}^{-4q})^2 + \frac{1}{8}(1-\mathrm{e}^{-8q})^2\right] \\ &= \frac{1}{16}N(N-1)\left[(1-\mathrm{e}^{-8q})^2 + 2(N-2)\mathrm{e}^{-4q}(1-\mathrm{e}^{-4q})^2\right].\end{aligned} \tag{A9}$$

APPENDIX B: SOME STATISTICAL PROPERTIES OF *SS* FOR CASE (ii) OF SECTION 2

The analogue of equation (A1) is

$$SS(\nu_*) = \sum_{\ell=1}^{N-1}\sum_{k=\ell+1}^{N}[\sin \omega_*(e_k - e_\ell)]^2, \tag{B1}$$

where

$$e_j \equiv \sum_{i=1}^{j}\epsilon_i\ .$$

It follows that

$$e_k - e_\ell = \sum_{i=\ell+1}^{k}\epsilon_i$$

is Gaussian with zero mean and variance $(k - \ell)\sigma^2$ (σ^2 being the variance of the ϵ_j). Consequently, by (A4),

$$\mathrm{E}\sin^2 \omega_*(e_k - e_\ell) = \frac{1}{2}\left(1 - \mathrm{e}^{-2(k-\ell)q}\right), \tag{B2}$$

where, as in Appendix A, $q = (\omega_*\sigma)^2$.

The expected value of $SS(\nu_*)$ can be obtained with the help of (B2):

$$\begin{aligned}\mathrm{E}SS(\nu_*) &= \frac{1}{2}\sum_{\ell=1}^{N-1}\sum_{k=\ell+1}^{N}\left(1 - \mathrm{e}^{-2(k-\ell)q}\right) \\ &= \frac{1}{4}N(N-1) - \frac{1}{2}\sum_{\ell=1}^{N-1}\sum_{k=\ell+1}^{N}a^{k-\ell},\end{aligned}$$

where $a = \exp(-2q)$. The last term is

$$\begin{aligned}\sum_{\ell=1}^{N-1}\sum_{k=\ell+1}^{N}a^{k-\ell} &= \sum_{\ell=1}^{N-1}a\frac{1-a^{N-\ell}}{1-a} \\ &= \frac{a(N-1)}{1-a} - \frac{a}{1-a}[a + a^2 + \cdots + a^{N-1}] \\ &= \frac{a}{1-a}\left[(N-1) - a\frac{1-a^{N-1}}{1-a}\right].\end{aligned}$$

The final result is

$$\mathrm{E}SS(\nu_*) = \frac{1}{4}N(N-1) - \frac{a}{2(1-a)}\left[(N-1) - a\frac{1-a^{N-1}}{1-a}\right]. \tag{B3}$$

In order to evaluate the variance of $SS(\nu_*)$, covariances of the form

$$C(\theta_1, \theta_2, \theta_3) = \mathrm{cov}\left[\sin^2(\theta_1 + \theta_2), \sin^2(\theta_1 + \theta_3)\right],$$

where $\theta_j \sim N(0, q_j)$ are independent, are required. Expanding the sinusoidal functions, setting expected values of odd functions equal to zero, and re-arranging,

$$\begin{aligned} C(\theta_1,\theta_2,\theta_3) &= \left[\mathrm{E}\cos^2\theta_2\right]\left[\mathrm{E}\cos^2\theta_3\right]\mathrm{var}(\sin^2\theta_1) + \left[\mathrm{E}\sin^2\theta_2\right]\left[\mathrm{E}\sin^2\theta_3\right]\mathrm{var}(\cos^2\theta_1) \\ &\quad + \left\{\left[\mathrm{E}\cos^2\theta_2\right]\left[\mathrm{E}\sin^2\theta_3\right] + \left[\mathrm{E}\sin^2\theta_2\right]\left[\mathrm{E}\cos^2\theta_3\right]\right\}\mathrm{cov}(\sin^2\theta_1,\cos^2\theta_1) \\ &= \left\{\left[\mathrm{E}\cos^2\theta_2\right] - \left[\mathrm{E}\sin^2\theta_2\right]\right\}\left\{\left[\mathrm{E}\cos^2\theta_3\right] - \left[\mathrm{E}\sin^2\theta_3\right]\right\}\mathrm{var}(\sin^2\theta_1). \end{aligned}$$

The last line of the result follows because

$$\mathrm{var}(\sin^2\theta_1) \equiv \mathrm{var}(\cos^2\theta_1) \equiv -\mathrm{cov}(\sin^2\theta_1,\cos^2\theta_1)\ .$$

Substituting results from (A4),

$$C(\theta_1,\theta_2,\theta_3) = \frac{1}{8}\left(1-\mathrm{e}^{-4q_1}\right)^2\mathrm{e}^{-2(q_2+q_3)}. \tag{B4}$$

The required variance is then

$$\begin{aligned} \mathrm{var}[SS(\nu_*)] &= \sum_{\ell=1}^{N-1}\sum_{k=\ell+1}^{N}\mathrm{cov}\left[\sin^2\omega_*(e_k-e_\ell),\sum_{r=1}^{N-1}\sum_{s=r+1}^{N}\sin^2\omega_*(e_s-e_r)\right] \\ &= \sum_{\ell=1}^{N-1}\sum_{k=\ell+1}^{N}\mathrm{cov}\left\{\sin^2\omega_*\left[\sum_{i=\ell+1}^{k}\epsilon_i\right],\sum_{r=1}^{k-1}\sum_{s=\ell+1}^{N}\sin^2\omega_*\left[\sum_{j=r+1}^{s}\epsilon_j\right]\right\}. \end{aligned} \tag{B5}$$

The last sum is conveniently split into four parts, which differ in their overlap θ_1 with the terms in $S_0 = \sin^2(\omega_*\sum_{i=\ell+1}^{k}\epsilon_i)$:

$$\sum_{r=1}^{k-1}\sum_{s=\ell+1}^{N}\sin^2\omega_*\left(\sum_{j=r+1}^{s}\epsilon_j\right) = S_1+S_2+S_3+S_4,$$

where

$$\begin{aligned} S_1(\ell,k) &\equiv \sum_{r=1}^{\ell}\sum_{s=\ell+1}^{k}\sin^2\left(\omega_*\sum_{j=r+1}^{s}\epsilon_j\right) \quad \theta_1=\omega_*\sum_{i=\ell+1}^{s}\epsilon_i \quad \theta_2=\omega_*\sum_{i=s+1}^{k}\epsilon_i \quad \theta_3=\omega_*\sum_{j=r+1}^{\ell}\epsilon_i \\ S_2(\ell,k) &\equiv \sum_{r=1}^{\ell}\sum_{s=k+1}^{N}\sin^2\left(\omega_*\sum_{j=r+1}^{s}\epsilon_j\right) \quad \theta_1=\omega_*\sum_{i=\ell+1}^{k}\epsilon_i \quad \theta_2=0 \quad \theta_3=\omega_*\left(\sum_{i=\ell+1}^{s}\epsilon_i+\sum_{j=r+1}^{\ell}\epsilon_i\right) \\ S_3(\ell,k) &\equiv \sum_{r=\ell+1}^{k-1}\sum_{s=r+1}^{k}\sin^2\left(\omega_*\sum_{j=r+1}^{s}\epsilon_j\right) \quad \theta_1=\omega_*\sum_{i=r}^{s}\epsilon_i \quad \theta_2=\omega_*\left(\sum_{i=\ell+1}^{r-1}\epsilon_i+\sum_{j=s+1}^{k}\epsilon_i\right) \quad \theta_3=0 \\ S_4(\ell,k) &\equiv \sum_{r=\ell+1}^{k-1}\sum_{s=k+1}^{N}\sin^2\left(\omega_*\sum_{j=r+1}^{s}\epsilon_j\right) \quad \theta_1=\omega_*\sum_{i=r+1}^{k}\epsilon_i \quad \theta_2=\omega_*\sum_{i=\ell+1}^{r}\epsilon_i \quad \theta_3=\omega_*\left(\sum_{i=k+1}^{s}\epsilon_i\right). \end{aligned}$$

In these formulae, θ_2 are the terms which occur only in S_0, and θ_3 those only in S_1– S_4. Using (B4),

$$\begin{aligned} \mathrm{cov}(S_0,S_1) &= \frac{1}{8}\sum_{r=1}^{\ell}\sum_{s=\ell+1}^{k}\left(1-\mathrm{e}^{-4q(s-\ell)}\right)^2\mathrm{e}^{-2q(k+\ell-r-s)} \\ \mathrm{cov}(S_0,S_2) &= \frac{1}{8}\sum_{r=1}^{\ell}\sum_{s=k+1}^{N}\left(1-\mathrm{e}^{-4q(k-\ell)}\right)^2\mathrm{e}^{-2q(\ell-r+s-k)} \\ \mathrm{cov}(S_0,S_3) &= \frac{1}{8}\sum_{r=\ell+1}^{\ell}\sum_{s=r+1}^{k}\left(1-\mathrm{e}^{-4q(s-r-1)}\right)^2\mathrm{e}^{-2q(r+k-\ell-s-2)} \\ \mathrm{cov}(S_0,S_4) &= \frac{1}{8}\sum_{r=\ell+1}^{\ell}\sum_{s=k+1}^{N}\left(1-\mathrm{e}^{-4q(k-r)}\right)^2\mathrm{e}^{-2q(r+s-k-\ell)}. \end{aligned} \tag{B6}$$

The required variance can be calculated from (B5) as

$$\mathrm{var}[SS(\nu_*)] = \sum_{\ell=1}^{N-1}\sum_{k=\ell+1}^{N}\mathrm{cov}[S_0(\ell,k),S_1(\ell,k)+S_2(\ell,k)+S_3(\ell,k)+S_4(\ell,k)] \tag{B7}$$

with the covariances given by (B6). Computer summation over four indices is slow for large N, and can be speeded up by explicit evaluation of (B6):

$$\begin{aligned}
\mathrm{cov}(S_0, S_1) &= \frac{a^{k+\ell} - a^k}{8(1-a)} \left[\frac{a^{-\ell} - a^{-k} + 2a^{-\ell+1} - 2a^{k-2\ell+1}}{1-a} + \frac{a^3}{1-a^3}(a^{3k-4\ell} - a^{-\ell}) \right] \\
\mathrm{cov}(S_0, S_2) &= \frac{a}{8(1-a)^2}(a^{N+\ell-k} - a^{\ell})(1-a^{-\ell})(1-2a^{2(k-\ell)} + a^{4(k-\ell)}) \\
\mathrm{cov}(S_0, S_3) &= \frac{a^{k-\ell}}{8} \left\{ \frac{1}{1-a} \left[-(k-\ell-1) + \frac{a^{\ell-k+1} - 1}{1-a} \right] - \frac{2a}{1-a} \left[(k-\ell-1) + \frac{a^{k-\ell} - a}{1-a} \right] \right. \\
&\quad \left. + \frac{a^3}{1-a^3} \left[(k-\ell-1) + \frac{a^{3(k-\ell)} - a^3}{1-a^3} \right] \right\} \\
\mathrm{cov}(S_0, S_4) &= \frac{(a^k - a^N)}{8} \left[\left(\frac{a}{1-a} \right)^2 (a^{-k} - a^{-\ell-1}) + 2\frac{a^{k-\ell+1}}{(1-a)^2}(a^{-\ell} - a^{-k+1}) - \frac{a^{3k-4\ell+1} - a^{-\ell+4}}{(1-a)(1-a^3)} \right],
\end{aligned} \tag{B8}$$

where $a = \exp(-2q) = \exp[-2(2\pi\nu_*\sigma)^2]$, as before.

It is, of course, also possible to sum the expressions in (B8) explicitly over k and ℓ, but the resulting equations are tediously long, and do not seem amenable to simplification.

This paper has been typeset from a TEX/LATEX file prepared by the author.