Stat

The ISI's Journal for the Rapid
Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.13

Aligning some Nicholson sheep-blowfly data sets with system input periods

David R. Brillinger[a,*]

Received 2 December 2012; Accepted 10 December 2012

During the 1950s the Australian entomologist Alexander Nicholson studied a sheep pest, *lucilia cuprina*, (L cuprina), the sheep-blowfly. In laboratory experiments blowfly populations were set up in cages. They were supplied with necessary food and water and every other day counts were made of the numbers in their various stages of development. The experiments went on for over a year. Various statistical studies have been carried out on their data. Sadly, the bulk of the data appears to be lost. Recently this author made the discovery of total population counts for ten Nicholson experiments. These data were in a collection of copies of index cards he made during a trip to Australia in 1977. In eight of the experiments the input food was varied cyclically in sawtooth fashion, each experiment having a different period of application. However, and what is the concern of this article, which data set went with which period of application remains unclear. In the present study use is made of periodograms, spectrograms and seasonal adjustment to seek a one-to-one correspondence between series and period. The estimate constructed is consistent under smoothing and limiting conditions. It is time domain based, but confirmed by periodogram and spectrogram computation. Copyright © 2013 John Wiley & Sons, Ltd.

Keywords: A. J. Nicholson; exploratory data analysis; lucilia cuprina; periodogram; sheep-blowfly; spectrogram; triangle/sawtooth wave input

1 Introduction

Alexander Nicholson, then Chief of the Division of Entomology, Commonwealth Scientific and Industrial Research Organisation of Australia, carried out experimental studies in the 1950s of an insect pest *L Cuprina*, the sheep-blowfly. This was part of his work to understand sheep-blowfly development. He maintained laboratory blowfly populations over long periods of time. A variety of combinations and levels of liver, water and sugar were supplied as nourishment to the insects and counts were recorded every other day. Nicholson labeled the cages B, C, D, G, H, I, J, K, L, O. Throughout the experiments a level of food was provided to maintain the populations. Cage I was a control that received a steady supply of food. In the remaining cages liver was supplied with the amount varying up and down in a regular sawtooth fashion. These applications were cyclic with periods 5, 10, 15, 20, 25, 30, 35, and 40 in units of two days. It is known, from Nicholson (1957), that the food for cage C was varied with period 20 and for cage H with period 40. However it is not known which data series from cages B, D, G, K, L, O corresponds to which of the remaining periods (5, 10, 15, 25, 30, 35). The research work in this paper is to infer a one-to-one correspondence between the periods and cages. The result would allow ecologists and applied mathematicians to better study the Nicholson equation, as in Cushing et al. (2003).

[a]**Department of Statistics, University of California, Berkeley, 94720, California, USA**
*__Email: brill@stat.berkeley.edu__

(wileyonlinelibrary.com) DOI: 10.1002/sta4.13

The basic papers concerning these experiments are Nicholson (1954a, 1954b, 1957, 1960). The 1957 paper is mentioned so often here that it will be referred to simply as Nicholson.

The analytic interest lies in estimating a permutation. The practical interest is in carrying out groundwork before building a dynamical system model of the data.

2 Background

Biological subject matter is basic to understanding ecological situations. Some detail concerning blowflies appears in Figure 1 which provides a representation of the insect's stages of development (Brillinger et al., 1980). There is a delayed feedback due to the necessary time between generations.

Cage I was a control experiment in which the nourishment levels were kept constant. Figure 2 shows the time series of successive adult population counts in this case. Examining this figure, one notes that up to about time step 200 there is a continued cycling. (The time step is two days in the computations and the figures of the paper.) The period of the cycling can be related to the length of a blowfly generation. In particular, one sees about 10 cycles in about 200 time steps in Figure 1. For the remainder of the time period oscillations continue, but they vary around a slowly increasing background level. It seems that the situation changes after time step 200. It may be noted that all the series measurements from all of the cages start on the same day. The first 280 observations will be studied here.

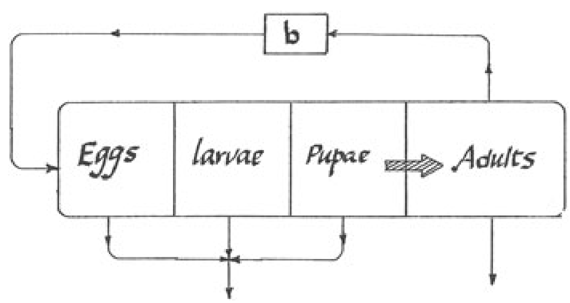


Figure 1. The life cycle of the blowfly. The figure is adapted from one in (Brillinger et al., 1980). The arrows leaving the boxes correspond to deaths.

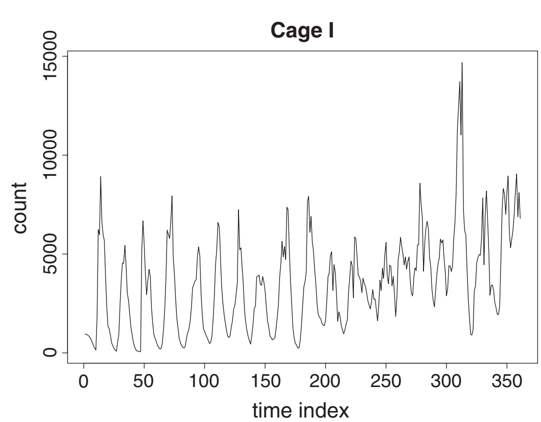


Figure 2. The every-other-day population count of the adults in cage I.

The ISI's Journal for the Rapid Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.13

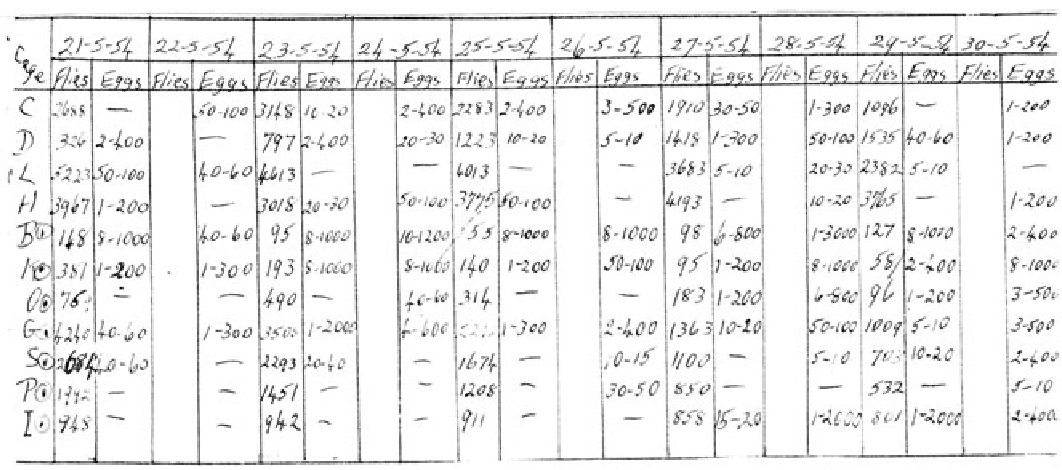
The data graphed in the figure may be found at

`www.stat.berkeley.edu/~brill/~blowfly97I.html`

These Cage I data have been analyzed by a number of researchers including Brillinger et al. (1980), Guttorp (1980), Ionides (2011), Ipkaktchi (1979), Oster & Ipaktchi (1978), Wood (2010), and Xia & Tong (2011).

For the other eight cages, which ran changing input experiments, liver was supplied in a triangle wave/sawtooth fashion. This was Nicholson's attempt to mimic "fluctuating environmental factors". There is some formal discussion of periodically varying environments in Nisbett & Gurney (1976).

Figure 3 provides a scan of one of the index cards upon which the data were recorded. The cage labels are listed in the leftmost column. The beginning day is "21-5-56". The adult counts are listed as "Flies". Ranges of egg counts are listed daily.



Cage	21-5-54		22-5-54		23-5-54		24-5-54		25-5-54		26-5-54		27-5-54		28-5-54		29-5-54		30-5-54	
	Flies	Eggs	Flies	Eggs	Flies	Eggs	Flies	Eggs	Flies	Eggs	Flies	Eggs	Flies	Eggs	Flies	Eggs	Flies	Eggs	Flies	Eggs
C	2688	—		50-100	3148	10-20		2-400	2283	2-400		3-500	1910	30-50		1-300	1096	—		1-200
D	326	2-400		—	797	2-400		20-30	1223	10-20		5-10	1418	1-300		50-100	1535	40-60		1-200
L	5223	50-100		40-60	4613	—		—	4013	—		—	3683	5-10		20-30	2382	5-10		—
H	3967	1-200		—	3018	20-30		50-100	3775	50-100		—	4193	—		10-20	3765	—		1-200
B	148	8-1000		40-60	95	8-1000		10-1200	55	8-1000		8-1000	98	6-800		1-3000	127	8-1000		2-400
K	381	1-200		1-300	193	8-1000		8-1000	140	1-200		50-100	95	1-200		8-1000	58	2-400		8-1000
O	76	—		—	490	—		40-60	314	—		—	183	1-200		6-800	96	1-200		3-500
G	4240	40-60		1-300	3500	1-2000		4-600	[illegible]	1-300		2-400	1363	10-20		50-100	1009	5-10		3-500
S	2084	40-60		—	2293	20-40		—	1674	—		10-15	1100	—		5-10	703	10-20		2-400
P	1992	—		—	1451	—		—	1208	—		30-50	850	—		—	532	—		5-10
I	948	—		—	942	—		—	911	—		—	858	15-20		1-2000	801	1-2000		2-400

Figure 3. A copy of one of Nicholson's index cards. The top row provides the recording dates. The leftmost column provides the cage label. The fly and egg counts are in alternating columns.

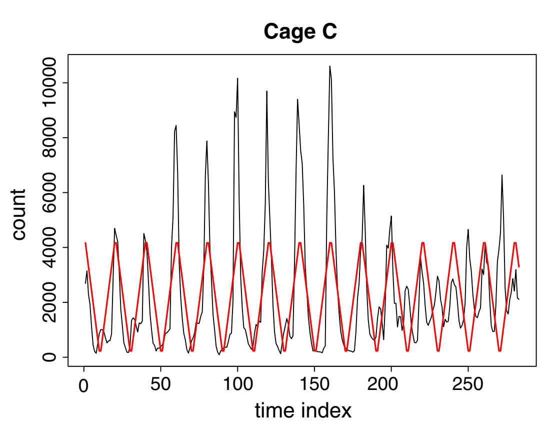


Figure 4. Cage C with the sawtooth wave superposed in red. This corresponds to Nicholson's Figure 8.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.13

Stat
The ISI's Journal for the Rapid Dissemination of Statistics Research

Figure 4 shows the population counts for cage C with the sawtooth wave of food supply superposed in red. It corresponds to Figure 8 in Nicholson. The sawtooth period was matched with the adult fly data by crosscovariance analysis.

3 Some quotes from Nicholson's papers

Nicholson's own words provide a special way to proceed to the analyses. In his seminal 1957 paper he set down goals for his "fluctuating environmental factors" experiments as follows:

> "To demonstrate how populations may accommodate themselves to changed conditions, and maintain themselves in being, ..."

> "In eight out of ten concurrent experimental cultures of *L cuprina* the fairly regular oscillations which took place over a period of about eighteen months began to change in period and in amplitude, and the population levels began to rise appreciably ... All these cultures were governed by adult competition for small quantities of ground liver, the larvae being given more food than they could consume."

> "On purely theoretical grounds I had concluded earlier Nicholson (1954b), that oscillations in population density of internal origin should tend to conform to climate changes, the period of which, when cyclical, tends to be impressed upon the population. ... I therefore varied the availability of a prerequisite which is likely to be influenced by climatic change in nature - namely food ... . To check this expectation, ten cultures were each supplied with unlimited quantities of larval food and also of water and sugar for the adults. In eight of the cages the ground liver was regularly and progressively varied from day to day from 0.05 grams of 0.5 grams per day in the way shown ..."

The figure to which Nicholson refers is this paper's Figure 4.

4 Statistical methods employed

Part of the work may be viewed as that of studying a seasonal effect. There are a variety of methods for seasonal estimation and adjustment. Often they are based on models of the form

$$Y_t = T_t + S_t + \epsilon_t \qquad t = 0 \pm 1, \pm 2, \ldots \tag{1}$$

where Y_t is a series of interest, T_t is a trend term, and S_t is a seasonal effect satisfying $S_{t+p} = S_t$ for some period. It is typical to assume that S_t and T_t are smooth functions. The noise term, ϵ_t, will have mean 0.

A traditional way to study cycles/periods in time series analysis is via some form of Fourier/harmonic/spectral analysis. These include periodogram and spectrogram analyses. These methods are often used as exploratory tools.

Let the data for a series be denoted Y_t, $t = 0, \ldots, T-1$. The empirical Fourier transform at frequency λ is defined as

$$d_Y^T(\lambda) = \sum_0^{T-1} Y_t \exp\{-i\lambda t\}.$$

With this definition the periodogram, $I^T(\lambda)$, is $\left|d_Y^T(\lambda)\right|^2 / \{2\pi T\}$. The mean of the Y-values is removed before computing the periodogram.

The spectrogram may be computed by sliding a data window along a series and computing the periodogram of the data within each window. In this way one obtains a quantity depending on both period and time. It may be displayed by contour, image and perspective plots. A lasting periodic component in a series would appear as a constant level

The ISI's Journal for the Rapid Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.13

ridge in a perspective plot or as a line in an image plot. Dahlhaus (1997) and Priestley (1966) have provided formal statistical approaches for studying the spectrogram. Koenig et al. (1946) is a historical reference to the spectrogram.

The presence of periodic behavior is the crux of this present work. A periods is known to be one of the discrete set of sawtooth periods used in Nicholson's experiments. The principal tool to be employed here is "a seasonal-trend decomposition based on loess" (STL), see Cleveland et al. (1990). Both `loess(.)` and `stl(.)` are functions in R. This method is chosen because of its flexibility.

The fuction `stl(.)` leads to estimates $\hat{T}_t$, $\hat{S}_t$ and residuals $\hat{\epsilon}_t = Y_t - \hat{T}_t - \hat{S}_t$. It allows the objective rejection of outliers and the specification of amounts of smoothing in the estimation procedure.

The application of `stl(.)` may lead to a consistent estimate of the parameter $(p_1, p_2, \ldots, p_6)$, some permutation of (5, 10, 15, 25, 30, 35). This then allows aligning the cage series with the food application periods employed by Nicholson. The estimation of the parameter will proceed by looking for the permutation that provide the minimum of the sum of the residuals from model (1), as computed by `stl(.)` summed over the six series for cages B, D, G, K, L, O.

There are advantages to employing a triangle wave. For example, it can crudely mimic the oscillations of the weather, as mentioned in the previous Nicholson quotations. But there are also disadvantages, particularly for the period of length 20. That period is close to the natural period of a blowfly population's oscillations, which is the situation for the data in this paper. The Appendix provides a Fourier expansion for the triangle wave and one sees input power at harmonically related frequencies.

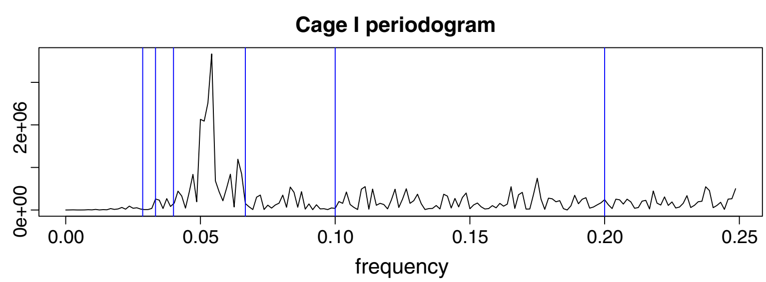


Figure 5. Periodogram of the cage I data. The vertical lines are at Nicholson's periods 5, 10, 15, 25, 30, 35.

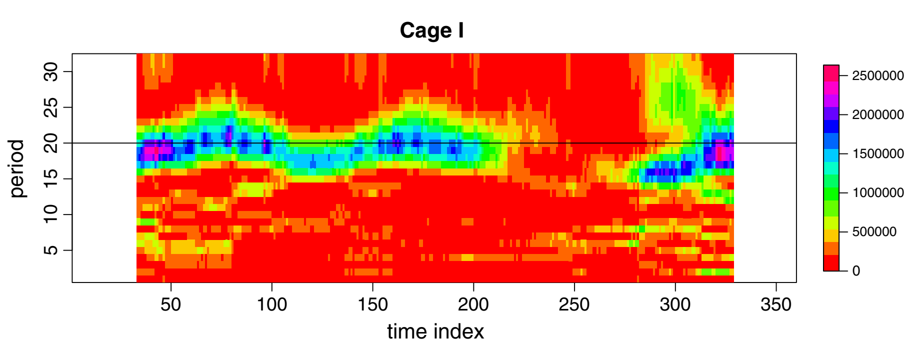


Figure 6. The spectrogram, a plot of variance versus time and period, for the data from cage I. The horizontal line is at period 20. The panels are labeled.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.13

Stat
The ISI's Journal for the Rapid
Dissemination of Statistics Research

5 Results

The presentation begins with results for cages I and C. Cage I is the control, for which the food application does not change, while for cage C the food amount can be deduced to be varying in sawtooth fashion with a period of 20 time steps.

5.1. The control experiment

The data of cage I is studied to obtain a baseline. It was a control in the sense that the levels of input food and water did not change during the experiment. The counts were graphed in Figure 2.

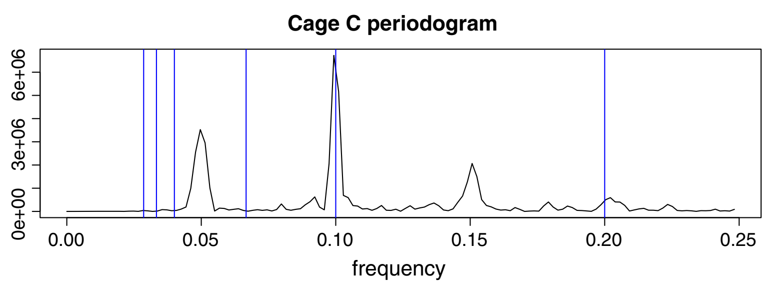


Figure 7. The periodogram for the cage C data. The vertical lines are at Nicholson's known period values.

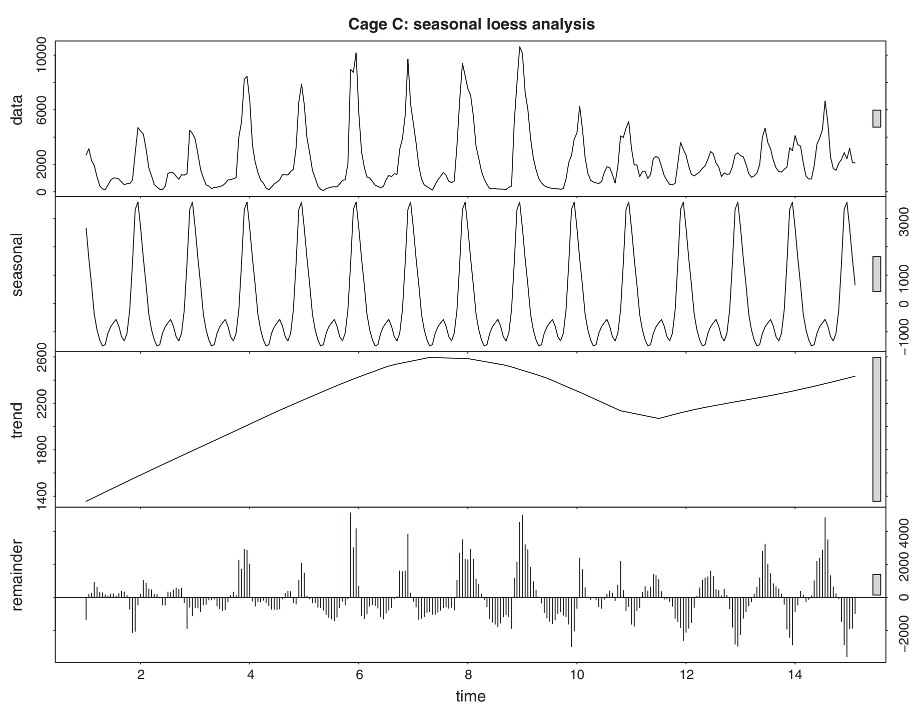


Figure 8. The result of employing the R trend-seasonal modeling function `stl(.)`.

The ISI's Journal for the Rapid
Dissemination of Statistics Research
(wileyonlinelibrary.com) DOI: 10.1002/sta4.13

Figure 5 shows the periodogram of the first differences of the data of cage I. One sees a pronounced peak around frequency 0.05 (period 20), amongst many smaller peaks.

Figure 6 provides a spectrogram for the same data. One notes a broad ridge around period 20 until just after time mark 200. This may correspond to a natural period relating to the blowfly's life cycle. The situation changes after time 225, as evidenced in the right-hand side of Figure 6. In particular, the natural period becomes indistinct. A horizonal line has been added at level 20, corresponding to a natural period.

5.2. *The experiments with changing input*

The experiments in the eight cages B, C, D, G, H, K, L and O vary the level of ground liver provided to the blowflies in a sawtooth wave fashion (see Figure 4), with differing periods of application.

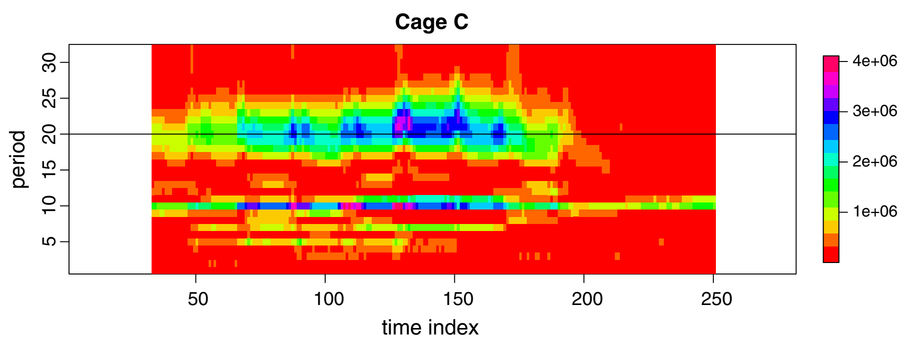


Figure 9. The spectrogram, a plot of variability versus time and period, for the data for cage C.

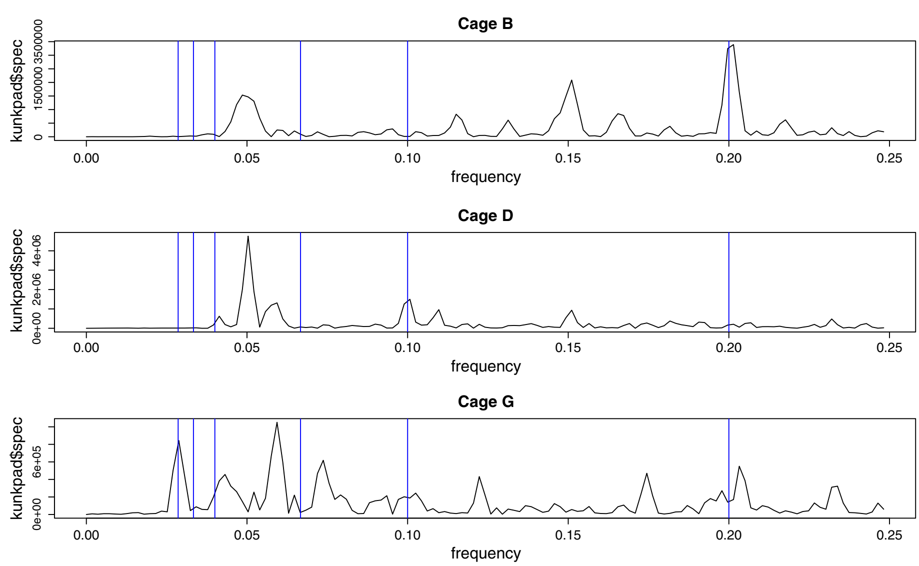


Figure 10. Periodograms for the first differences for the series from cages B, D, and G. The vertical lines correspond to periods 35, 30, 25, 15, 10, and 5, respectively.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.13

Stat
The ISI's Journal for the Rapid Dissemination of Statistics Research

Figure 7 shows the periodogram of the first differences for the cage C data. One sees a high peak at frequency 0.1 and lesser peaks at other frequencies. The peak near 0.05 is perhaps the natural frequency showing itself. The peak near 0.15 may be an interaction between the natural period and the liver input period.

Figure 8 is the result of applying the function `stl(.)` to the series for cage C. The seasonal effect, seen in the second panel of the figure, has two peaks. This would lead analytically to harmonics in the manner of the expansion in Appendix A. The final panel shows that the unexplained variability is much reduced by including the trend and seasonal components.

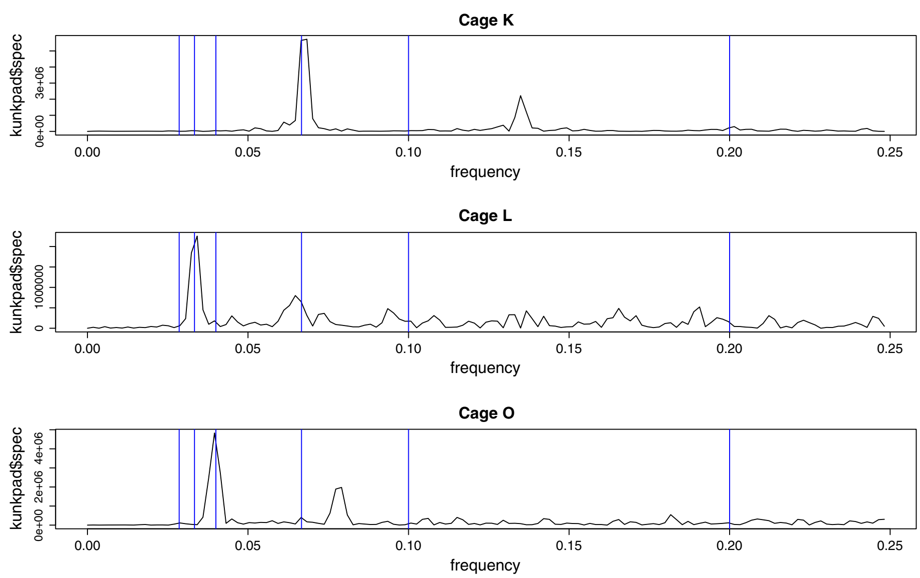


Figure 11. Periodograms for the first differences for the series from cages K, L, and O. The vertical lines correspond to periods 35, 30, 25, 15, 10, and 5, respectively.

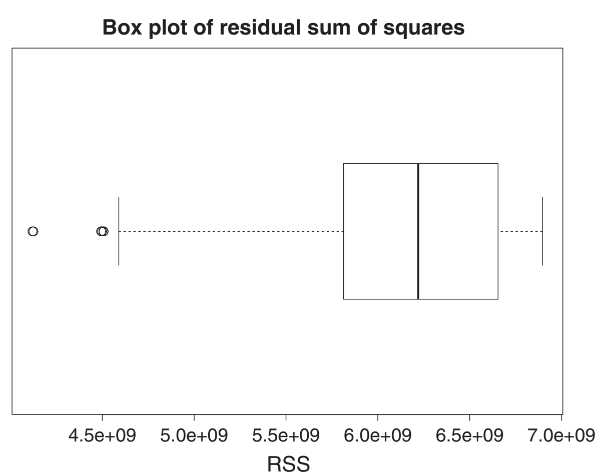


Figure 12. A box plot of the 720 total sums of squared residuals.

The ISI's Journal for the Rapid Dissemination of Statistics Research
(wileyonlinelibrary.com) DOI: 10.1002/sta4.13

Figure 9 shows the spectrogram for the cage C series. The horizontal line is at the known period of 20. One sees that intensity dieing away just before time point 200, as in the top panel of Figure 8. Both the input and the natural period may be seen in cage C's spectrogram. The fact that narrow horizontal line at period 10 extends entirely across the figure is to be noted. Nicholson does not appear to note the earlier presence of the 0.1 oscillation.

Of the cage C data, Nicholson writes:

> "It will be observed that the period of population oscillation in C is slightly longer than the natural period in the control (I) and that the oscillations up to about the 500th day correspond to alternate periods of food fluctuation. At about this time there is a sudden change after which the period of oscillation is halved. It illustrates the fact

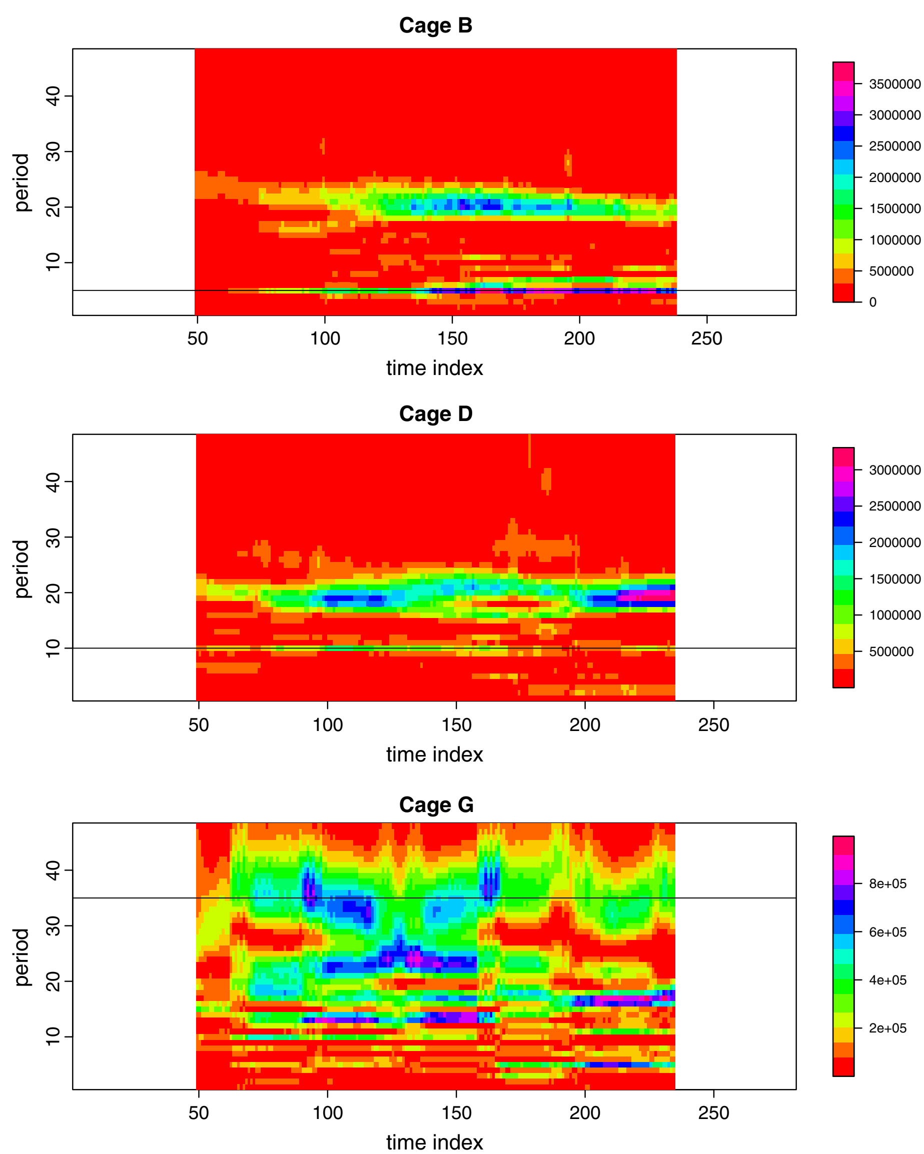


Figure 13. Spectrograms for cages B, D, and G. The horizontal lines are at levels 5, 10, and 35, respectively.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.13

The ISI's Journal for the Rapid
Dissemination of Statistics Research

that the period of population tends to be equal to, or to be either an exact multiple or simple fraction of, that of the fluctuation external influence."

This completes the discussion of the control experiment and also of a varying input experiment for which the period of liver input is known. Consideration now turns to the experiments for which the individual input periods are only known to be members of a set a set of six elements. Figures 10 and 11 give some clues about these periodicities.

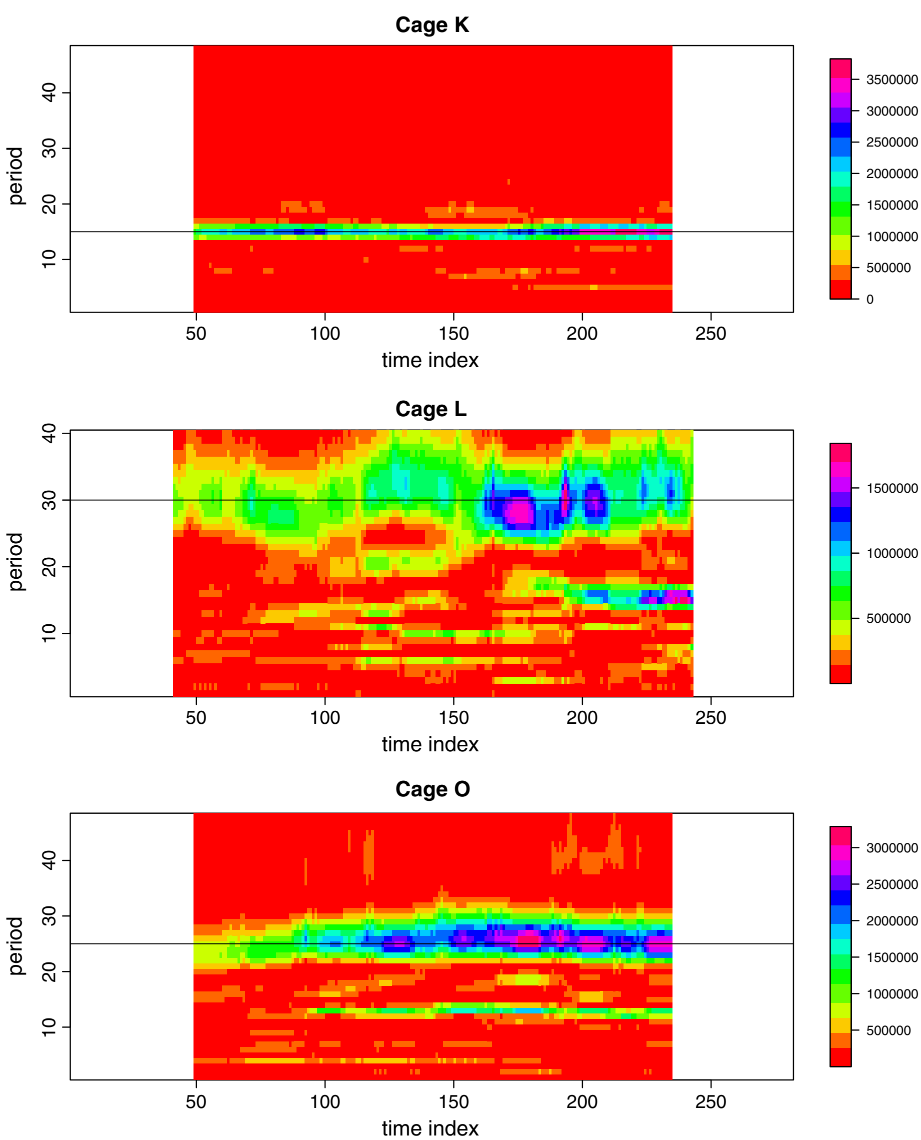


Figure 14. Spectrograms for cages K, L, and O. The horizontal lines are at levels 15, 30 and 25, respectively.

The ISI's Journal for the Rapid Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.13

5.3. Inferring the treatment assignment to cages

In this section the method of inferring the permutation described in Section 4 is implemented. In this approach the motivating assumptions are: (1) the time series for the different cages are statistically independent, and (2) approximating each series as stationary and mixing is reasonable.

For a given permutation the residual-squares are summed across all six of the cages. A box plot of the 720 values is provided in Figure 12. One notes the outliers on the left. The permutation corresponding to the leftmost outlier is taken as the estimate.

The estimated permutation is

$$\hat{B} = 5; \hat{D} = 10; \hat{G} = 35; \hat{L} = 15; \hat{K} = 30; \hat{O} = 25. \tag{2}$$

This estimate can be shown to be consistent as T, the length of the series, goes to infinity. There is discussion in Choirat & Seri (2012) of the theoretical results it is possible to develop for estimates of discrete parameters.

5.4. Further assessment of results

To seek some further confirmation of the estimated permutation obtained, spectrograms are provided in Figures 13 and 14. The horizontal lines are at the estimated periods given in expression (2).

In the B, D, G figures one finds the added horizontal lines lie within the bands in the figure. A notable case is the band around 20, the natural frequency. The figure for G shows a lot of activity, as did the periodogram in Figure 10.

Turning to the cage K result in Figure 14, the horizontal line corresponding to period 15 is dramatically apparent, as was the peak in Figure 11. The peak was supported by the harmonic at period 7.5. The line for period 30 in cage L runs along a broad ridge. Again, the clear peak in Figure 11 is seen in Figure 14. Lastly the line at 25 is evident for cage O in Figure 14, as is a harmonic.

6 Discussion

Interest in Nicholson's work has remained high. One reason perhaps may be that blowflies in the laboratory provide a controlled system that is both nonlinear and has feedback and which, when the input is periodic, exhibits the development of harmonics and subharmonics.

A general remark is that a form of multivariate analysis, specifically fitting a permutation rather than looking at the individual series separately, has added strength to the inference.

The primary statistical question in this study was how to estimate a permutation of a set of 6 periods. A peripheral question was whether the input could be seen in the output? The situation is of further interest because the system is nonlinear and has feedback.

The periodogram and spectrogram proved useful tools because they provide a decomposition of the series into additive frequency components. A seasonal adjustment procedure was necessary to calculate the criterion which was minimized in order to estimate the unknown permutation.

For cages C and H the period was indicated in Nicholson. Here those series are used to study the inference method in a situation where the answer is known.

The work presented here is preparation for future projects, including fitting a finite dimensional parameter dynamical system model with input. The work in Brillinger (2012) is related to the present paper; in particular, it provides further detail concerning the recovery of the data. It also studies the data from cages C, H, and I.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.13

The ISI's Journal for the Rapid
Dissemination of Statistics Research

Appendix A

Nicholson employed a sawtooth wave input to the blowfly system in an attempt to mimic some aspects of nature. This has the disadvantage that the frequency support contains only the frequencies $n\pi/P$ where P is the period; see the expression for $y(t)$ below.

Mathworld (2012) gives the following trigonometric expansion for a triangle on the interval (.5,1.5), with base at level 0, peak at level 1, and period P:

$$y(t) = \left(8/\pi^2\right) \sum_{n=1}^{\infty} n^{-2}(-1)^{(n-1)/2} \sin(n\pi t/P).$$

This will lead to harmonics $n\pi/P$ in the output of a linear system; however, it does not generate subharmonics as may occur in a nonlinear system.

Appendix B

The estimate is consistent. This result does not seem to be a particular case of classic theorems. There is discussion of the situation in Choirat & Seri (2012).

A sketch of an argument runs as follows:

Empirical quantity: $Q^T(\theta)$, θ *in* Θ
Fixed quantity: $Q(\theta)$ minimized at θ_0 *in* Θ
Stochastic quantity: $Q^T(\theta)$ minimized at $\hat{\theta}$ *in* Θ
Convergence: $Q^T(\theta)$ tends to $Q(\theta)$ in probability/almost surely
Result: $\hat{\theta}$ tends to θ_0 in probability/almost surely

The model may involve a stationary mixing time series.

Acknowledgements

I thank: N. Fisher, J. Guckenheimer, P. Guttorp, A. Ipaktchi, G. Oster, R. Spence, and R. Lovett for helpful conversations. Partial support was provided by the NSF Grant DMS-1007553.

References

Brillinger, DR (2012), 'The Nicholson blowfly experiments: some history and EDA', *Journal of Time Series Analysis*, **33**, 718–723.

Brillinger, DR, Guckenheimer, J, Guttorp, PM & Oster, G (1980), 'Empirical modelling of time series data: the case of age and density dependent vital rates', *Lectures in Mathematics in the Life Sciences*, **13**, 65–90.

Choirat, C & Seri, R (2012), 'Estimation in discrete parameter models', *Statistical Science*, **27**, 278–293.

Cleveland, RB, Cleveland, WS, McRae, JE & Terpenning, I (1990), 'STL: a seasonal-trend decomposition procedure based on loess', *Journal of Official Statistics*, **6**, 3–73.

The ISI's Journal for the Rapid Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.13

Cushing, JM, Constantino, RE, Dennis, B, Desharnais, RA & Henson, SM (2003), *Chaos in Ecology*, *Academic*, Amsterdam.

Dahlhaus, R (1997), 'Fitting time series models to nonstationary processes', *Annals of Statistics*, **25**, 1–37.

Guttorp, P (1980),'Statistical, Modeling of Population Processes', Ph.D. Thesis, University of California, Berkeley.

Ionides, E (2011), 'Feature matching versus likelihood for dynamic systems: Nicholson's blowflies a case study', In *Michigan Statistical Learning Seminar,* Ann Arbor. http://www.stat.lsa.umich.edu/∼ionides/pubs/MSL.pdf.

Ipkaktchi, A (1979),'Dynamics of, Density Dependent Populations', Ph.D. Thesis, University of California, Berkeley.

Koenig, W, Dunn, HK & Lacy, LY (1946), 'The sound spectrograph', *Journal of the Acoustical Society of America*, **18**, 19–49.

May, RM (1976), 'Simple mathematical models with very complicated dynamics', *Nature*, **261**(5560), 459–467.

Mathworld (2012), Fourier series– triangle wave. http://mathworld.wolfram.com/FourierSeriesTriangleWave.html.

Nicholson, AJ (1954a), 'Compensatory reactions of populations to stresses, and their evolutionary significance', *Australian Journal of Zoology*, **2**, 1–8.

Nicholson, AJ (1954b), 'An outline of the dynamics of animal populations', *Australian Journal of Zoology*, **2**, 9–65.

Nicholson, AJ (1957), 'The self-adjustment of populations to change', *Cold Spring Harbor Symposium on Quantitative Biology*, **11**, 153–173.

Nicholson, AJ (1960), 'The role of population dynamics in natural selection'. in Tax, S (ed), Evolution After Darwin, University of Chicago Centennial, University of Chicago Press, Chicago, pp. 477–521.

Nisbett, RM & Gurney, WSC (1976), 'Population dynamics in a periodically varying environment', *Journal of Theoretical Biology*, **56**, 459–475.

Oster, G & Ipaktchi, A (1978), 'Population cycles', *Theoretical Chemistry: Periodicities in Chemistry and Biology*, **4**, 111–132.

Priestley, M (1966), 'Design relations for nonstationary processes', *Journal of the Royal Statistical Society*, **28**, 228–240.

Wood, SN (2010), 'Statistical inference for noisy nonlinear ecological dynamic systems', *Nature*, **466**, 1102–1104.

Xia, Y & Tong, H (2011), 'Feature matching in time series modelling', *Annals of Applied Statistics*, **26**, 21–46.