The Contingency Periodogram: A Method of Identifying Rhythms in Series of Nonmetric Ecological Data

Louis Legendre; Marcel Frechette; Pierre Legendre

Journal of Ecology, Vol. 69, No. 3 (Nov., 1981), 965-979.

Stable URL:
http://links.jstor.org/sici?sici=0022-0477%28198111%2969%3A3%3C965%3ATCPAMO%3E2.0.CO%3B2-3

Journal of Ecology is currently published by British Ecological Society.

Journal of Ecology (1981), **69**, 965–979

THE CONTINGENCY PERIODOGRAM: A METHOD OF IDENTIFYING RHYTHMS IN SERIES OF NONMETRIC ECOLOGICAL DATA

LOUIS LEGENDRE,* MARCEL FRÉCHETTE* AND PIERRE LEGENDRE†

**Département de biologie, Université Laval, Québec, Québec G1K 7P4, Canada, and*
†CERSE (Centre de recherche en sciences de l'environment),
Université du Québec à Montréal, Montréal, Québec H3C 3P8, Canada

SUMMARY

(1) Analysis of series, in time or space is a subject of growing interest in ecology, but the methods hitherto available (correlogram, periodogram, spectral analysis) were restricted to the analysis of metric (quantitative) data. Many ecological series, however, are of (or contain) essential variables which are qualitative (categorical), and many quantitative variables are more efficiently sampled as rank-ordered (ordinal) variables. There are no numerical methods to analyse series of nonmetric data.

(2) In this paper, the periodogram of Whittaker & Robinson is generalized to qualitative data series, using information theoretic measures. Algorithms are also described to partition rank-ordered variables into classes, in order to analyse them using the contingency periodogram.

(3) The validity of the contingency periodogram is assessed by comparing its results with those of the periodogram of Schuster, using two series of metric data. Both methods identify the same periods for a series of artificial data, and also for serial measurements of the photosynthetic capacity of estuarine phytoplankton, even when these metric data are reduced to a small number of states prior to contingency periodogram analysis.

(4) The contingency periodogram is also used to analyse a multivariate (multi-species) phytoplankton series. The multivariate data are reduced to a single multi-state qualitative variate by clustering the samples. At least one of the periods revealed is surprising, but can be explained.

INTRODUCTION

The search for rhythms in ecological data series is a subject of interest to a growing number of ecologists, especially as automatic recording of ecological variables becomes practicable (e.g. meteorological, behavioural, and even biological data such as chlorophyll fluorescence in open waters). Such rhythms may be forced by external phenomena, or may be endogeneous and sometimes phased on external rhythms. Ecological series are recorded along either time or space axes, or along a run in both time and space, and various approaches are known to the problem of identifying rhythms in such series. These approaches are the correlogram and periodogram analyses, and more recently spectral analysis. All three are restricted to series of metric (quantitative) data: neither qualitative (categorical) nor rank-ordered (ordinal) data can be analysed by these methods. The present paper describes a periodogram for the analysis of qualitative data series, its use being further extended—with some restrictions—to series of rank-ordered data. This new periodogram is therefore not in competition with the usual methods, as these cannot be used to analyse series of nonmetric data. We now describe all these methods, in outline.

† Present address: Départment de sciences biologiques, Université de Montréal, Montréal, Québec H3C 3J7, Canada.

0022-0477/81/1100–0965 $02.00

Spectral analysis is the most powerful tool for the analysis of metric data series. The variance (or power) spectrum is a partition of the variance of the series as a function of the observed frequencies (period^{-1}) or wavenumbers (wavelength^{-1}). High spectral densities are characteristic of rhythms, but the general shape of the spectrum, rather than its individual values, is generally considered (Kendall & Stuart 1966). When two series are analysed together, coherency and phase spectra may be computed. Maximum entropy spectral analysis, originally proposed by Burg (1967, 1975) and still under development, does not have the constraints of classical spectral analysis as to the characteristics of the data series (Ables 1974). The use of this new spectral technique has still to be assessed in ecology, while the classical method is more familiar to ecologists (Platt & Denman 1975).

Correlogram analysis is a generalized parametric correlation analysis, which measures the linear correlation between the successive terms of the series. Rhythms in the series correspond to the lags for which characteristic autocorrelation values are computed. Similarly, cross or lag correlation may be computed between two different series, in order to identify the lags between their rhythms. Nonparametric correlations (Kendall 1948) are designed to handle rank-ordered data, and they also prove useful in analysing metric variables which are not linearly related. In all cases, the data must exhibit a monotonic relationship (continuously increasing or decreasing relative to each other). No correlogram is based on nonparametric correlations, as its theoretical bases have not been developed yet. On the other hand, Fortier & Legendre (1979) have computed cross-correlation between ecological rank-ordered data using Kendall's rank correlation (τ).

Another approach to series analysis is the periodogram method. Two types of periodograms are known. One is the periodogram of Schuster (1898) which results from an harmonic analysis of the data series into a Fourier series. It can be applied only to metric variables as it involves the least-squares fit of harmonic sines and cosines to the data. Amplitudes are computed from the regression coefficients for each harmonic, and they are plotted as a function of the corresponding harmonic periods (Figs 5(b) and 7(b)): the interpretation is directly in terms of periods. The computation of critical values is described by Anderson (1971, p. 110 *et seq.*).

The other periodogram is that of Whittaker & Robinson (1924). Its use in ecology was reviewed by Enright (1965). In this periodogram, for each period (T) of interest to the ecologist, the series is divided into subseries of length T. These subseries are then assembled in a Buys–Ballot table, where each row of the table is one of the subseries, and where values corresponding to the same point of the period are in the same column. If the period considered corresponds to a rhythm of the series both the peaks and the troughs are in the same column. This results in a large difference between the highest and the lowest column averages. This difference, computed for the various possible periods T (on a new Buys–Ballot table for each period), is plotted in a periodogram as a function of the periods. This type of periodogram was devised for metric data, and it can be applied only to series with a single stable period (Enright 1965).

In ecological series, many essential variables are qualitative, and many quantitative variables are much more efficiently sampled as rank-ordered variables. Examples of qualitative ecological data which are often serially sampled are the dominant species, the biological association, the presence or absence of a species or of some substance above or below a critical threshold concentration, the type of substrate, and so on. Cyclical variables such as the direction of the wind or of a current, coded into nonordered states, are other examples of qualitative serial data encountered in ecology. Quantitative information, on the other hand, may often be sampled efficiently as importance or abundance scores; samples can also be rapidly enumerated in the laboratory as coded abundance scores (see for instance Frontier (1969, 1973)). Thus, the cost of data collection may be reduced, or the intensity of sampling may be increased for a given cost. The contingency periodogram described in the present paper is a generalization of the periodogram of Whittaker & Robinson (1924). It uses information theoretic measures for the analysis of qualitative—and eventually rank-ordered—ecological data series.

THE CONTINGENCY PERIODOGRAM

The contingency periodogram was first proposed by Legendre & Legendre (1979). It is based on a contingency table, introduced to ecology by Colwell (1974), where the columns are as in the Buys–Ballot table described above, and the rows are the various states of the variable under study (Table 2). Qualitative (non-ordered, categorical) variables are already divided into non-ordered states. On the other hand, rank-ordered (ordinal) variables must be divided into states before using them in contingency table analysis, and solutions to this problem are given in the next section. Values in the contingency table are the number of observations in a given state of the variable (row) made at one point of the period considered (column). When computing a periodogram, such a table must be constructed for each possible period.

A fictitious example of a 3-states non-ordered variable is presented in Table 1. The length of the series is $N = 16$ and the observational window, that is the range between the smallest and the largest detectable period (T) of the periodogram, is $2 \leqslant T \leqslant 16/2$.

For all periods between 2 and 8, data are recorded in a Buys–Ballot contingency table as described above. The table of the fictitious series, for $T = 5$, is given in Table 2. The value in row 2 and column 3 is the number of observations with state 2 recorded at the 3rd point of the period $T = 5$. These points are 3, 5 + 3 = 8 and 5 + 8 = 13. In this case, as can be seen from Table 1, all these observations belong to state 2, so the value in row 2 and column 3 of Table 2 is 3.

TABLE 1 Fictitious example: a series of sixteen observations of a variable which may be in one of three states. The particular state is shown by X.

States	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Total
3				X	X				X					X	X		5
2			X			X		X		X			X				5
1	X	X					X				X	X				X	6

TABLE 2. Contingency table of observations for the fictitious example in Table 1, for the period $T = 5$.

	Observation axis (X)					
States	1	2	3	4	5	Total
3	0	0	0	3	2	5
2	1	0	3	0	1	5
1	3	3	0	0	0	6
Total	4	3	3	3	3	16

The amount of uncertainty (H: entropy) as to the states (S), which is accounted for by a given period T, is the common entropy, which is defined as the information of the intersection* between S and X (the observation axis):

$$H(\mathrm{S} \cap \mathrm{X}) = H(\mathrm{S}) + H(\mathrm{X}) - H(\mathrm{SX})$$

where $H(\mathrm{S})$ is the uncertainty of the states, $H(\mathrm{X})$ is the uncertainty of the observation axis, and $H(\mathrm{SX})$ is the entropy of the whole contingency table. H values are computed following Shannon (1948):

$$H(\mathrm{S}) = -\sum_{i=1}^{s} \frac{N_i}{N} \log \frac{N_i}{N}$$

$$H(\mathrm{X}) = -\sum_{j=1}^{T} \frac{N_j}{N} \log \frac{N_j}{N}$$

$$H(\mathrm{SX}) = -\sum_{i=1}^{s} \sum_{j=1}^{T} \frac{N_{ij}}{N} \log \frac{N_{ij}}{N}$$

where N_{ij} are values of the contingency table in row i and column j, N_i (or N_j) are the totals of rows i (or columns j), and s is the number of states. If logarithms are base 2, H is measured in *bits*, while with natural logarithms the unit is the *nat* (as in Figs 1, 5, and 7).

According to Kullback (1959, p. 158), if the row (states) and column (X) classifications are independent, the quantity $2N[H(\mathrm{S}) + H(\mathrm{X}) - H(\mathrm{SX})]$ is asymptotically distributed as χ^2 with $(s-1)(T-1)$ degrees of freedom, when natural logarithms have been used in the computation of H. Therefore the hypothesis $H(\mathrm{S} \cap \mathrm{X}) = 0$ cannot be rejected when:

$$H(\mathrm{S}) + H(\mathrm{X}) - H(\mathrm{SX}) \leqslant \chi^2/2N$$

where χ^2 is the value of χ^2 at the preselected probability level with $(s-1)(T-1)$ degrees of freedom. However, it is not possible to test the significance of each single value computed in the periodogram because these tests would not be independent, the chance of one or more periodogram values exceeding their critical value being then greater than the probability level. Therefore this test can only be used to reject those values of $H(\mathrm{S} \cap \mathrm{X})$ which are definitely too small to be significantly different from zero. On the other hand, periodogram values exceeding their critical value are not necessarily significantly different from zero, but their ecological implications may be considered.

Finally, if it is sought to plot the contingency periodogram on a 0–1 scale, the ordinate axis may be divided by $H(\mathrm{S})$, which is a constant through all the periods of a given periodogram. The value thus plotted—$H(\mathrm{S} \cap \mathrm{X})/H(\mathrm{S})$, which is the fraction of the uncertainty as to the states being accounted for by X—is computed in the SPSS computer package (Nie *et al.* 1975) as the 'asymmetrical uncertainty coefficient'.

For the fictitious example, when $T = 5$ (Table 2), the various quantities computed are the following:

	$H(\mathrm{S})$	$H(\mathrm{X})$	$H(\mathrm{SX})$
$\log_e$	1·095	1·602	1·862
$\log_2$	1·579	2·311	2·686

Using natural logarithms, $H(\mathrm{S} \cap \mathrm{X}) = 0{\cdot}835$, for $T = 5$. The same quantity, computed for all the periods from $T = 2$ to $T = 8$, is plotted in the periodogram of Fig. 1. All the

* The intersection, $\cap$, of S and X consists of the co-occurrence of observations in both S and X (as in Table 2).

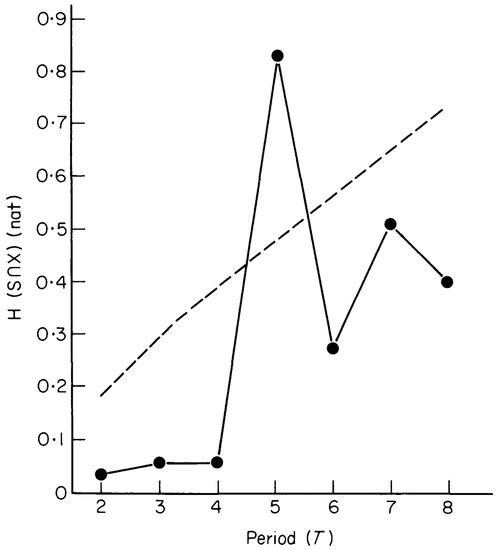

FIG. 1. Contingency periodogram of a fictitious data series (Tables 1 and 2). Broken line: critical values at the 0·05 probability level. The variable $H(\mathrm{S} \cap \mathrm{X})$, a measure of entropy, and the unit nat are explained in the text.

$H(\mathrm{S} \cap \mathrm{X})$ values except that for $T = 5$, are too small to be significantly different from zero at the 0·05 probability level: the series exhibits a single period $T = 5$.

THE USE OF RANK-ORDERED VARIABLES

Unlike qualitative (categorical) data, where the various states of the variable are not ordered, rank-ordered (ordinal) observations can be grouped into ordered states. Analysing series of rank-ordered data with the contingency periodogram, which does not take into account the ordering of the states, may therefore result in less certain identification of the rhythms than might be achieved. Kedem (1980) has proposed a method for reducing quantitative (and eventually rank-ordered) data to a binary form, and he has shown that if the series is stationary and of sufficient length (500–1000 observations), it is possible to estimate, from the binary series, the spectral density of the original data series. The method proposed hereafter is different in that the coding is very often in more than two states (limited by an optimality criterion) and in that it is applicable to much shorter series. The contingency periodogram may then prove useful to ecologists collecting information as abundance scores—with the restriction that rhythms present in the series may not emerge from this suboptimal analysis. In practice, examples (next section) show that the contingency periodogram is reliable and efficient in analysing series of rank-ordered data.

Rank-ordered variables must be divided into states in order to make them amenable to contingency periodogram analysis. This is no problem when there are biological or ecological criteria indicating where to partition the variable into states, and also how many states are required for a meaningful analysis of the variable. When this *a priori* knowledge is not available, the variable may be objectively partitioned into states according to some pre-established criteria. Such a method is now described. As in any series analysis, the data series must be stationary (the mean and variance must be stable) and detrended (long-term drift must be removed) prior to the numerical treatment. This subject has been reviewed at length by many authors, among them Kendall & Stuart (1966), Anderson (1971) and Legendre & Legendre (1979), and will not be discussed here.

It might be conceived that the best partitioning of a rank-ordered variable is when an equal number of data points falls into each state. This approach, however, does not take into account the structure of the data set in terms of tied ranks, and it does not give any indication as to the number of states. To obviate these problems, an optimum partitioning of a rank-ordered variable into states is here achieved (1) by minimizing the total within-states variability, (2) and maximizing the average information per state.

Criterion 1. The total within-states variability may be assessed by adding together the sums of squares of the differences from the mean rank for each state:

$$\sum_{i=1}^{s}\left[\sum_{\text{within state } i}(\text{rank}-\overline{\text{rank}})^2\right]$$

where $\overline{\text{rank}}$ is the mean rank within state i. This is equal to:

$$\sum_{k=1}^{N}\text{rank}_k^2-\sum_{i=1}^{s}\frac{1}{N_i}\left[\sum_{\text{within state } i}\text{rank}\right]^2.$$

(A similar measure of rank variability is used by Kendall (1948) in the generalized equation of correlation coefficients.) For all the possible partitions of the data set into states, the quantity $\sum_{k=1}^{N}\text{rank}_k^2$ is a constant. Thus, minimizing the total within-states variability is achieved by maximizing the sum of the s squared sums of ranks, each divided by its number of observations, therefore by maximizing the quantity

$$\sum_{i=1}^{s}\frac{1}{N_i}\left[\sum_{\text{within state } i}\text{rank}\right]^2,$$

abbreviated hereafter as $\sum(\sum^2 \text{ranks}/N_i)$.

Criterion 2. According to the previous section, the amount of information per state is given by:

$$H(\mathrm{S})/s=-\frac{1}{s}\sum_{i=1}^{s}\frac{N_i}{N}\log\frac{N_i}{N}$$

where N_i is the number of observations in state i, N is the total number of observations, and s is the number of states.

Ideally, the search for the partition optimizing the two criteria involves trying in turn all possible partitions of the data set in two states, then all the partitions in three states, and so forth until the information per state reaches a maximum and starts to decline. It will be seen that, as the number of ranks of the variable increases, this procedure rapidly becomes too tedious even for modern computers, so that the following approximate stepwise procedure may be worth using instead (see also the numerical example below):

(1) The first partition, in two states, is found by trying in turn all the possible partitions of the rank-ordered variable, and choosing the one which minimizes the total within-states variability (criterion 1 above).

(2) Keeping the first division fixed, a second one is found by the same trial method, in order to find the best partition in three states.

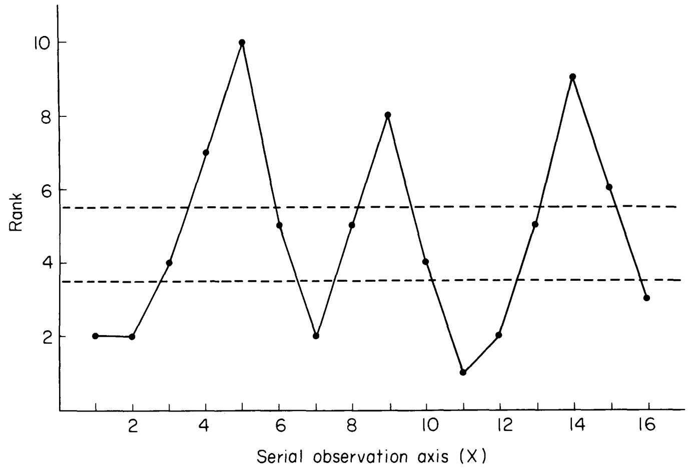

FIG. 2. Series of a fictitious rank-ordered variable (Tables 1 and 3). Dashed lines correspond to the chosen 3-state limits. See text for further explanation.

TABLE 3. Fictitious example: $N = 16$ ranked observations of an ordered variable.

Rank	1	2	3	4	5	6	7	8	9	10
No. of observations	1	4	1	2	3	1	1	1	1	1

(3) The following partitions are achieved in the same way, always keeping fixed the divisions found in the previous steps.

(4) At each step, the amount of information per state is computed for the best partition, and the partitioning procedure stops when $H(S)/s$ is maximum (criterion 2 above). Of course, this criterion may be waived if the ecologist prefers to choose himself the number of states s.

The usefulness of this stepwise procedure may be demonstrated by comparing the number of operations required by the two methods. With the relatively short series of 200 observations analysed at the beginning of the next section, the division into three states by the stepwise procedure involved the comparison of 594 different partitions. This included testing the best partition into four states against the chosen partition into three states. With the exact procedure, 1 313 599 partitions into two, three and four states would have to be compared, which would increase the amount of computing 2000-fold.

A fictitious rank-ordered data series, with $N = 16$, is shown in Fig. 2. The rank-ordered variable corresponding to the series is summarized in Table 3; it has ten rank-ordered states, with some tied observations. It will be used to illustrate the stepwise partitioning procedure. There are nine possible partitions of this rank-ordered variable into two states. For example, the partition between ranks 4 and 5 would result in:

Ranks	$\sum$ ranks	N_i	$\sum^2$ ranks$/N_i$
1–4	$1 + (4 \times 2) + 3 + (2 \times 4) = 20$	8	$20^2/8 = 50$
5–10	$(3 \times 5) + 6 + 7 + 8 + 9 + 10 = 55$	8	$55^2/8 = 378{\cdot}1$

and the total $\sum$ ($\sum^2$ ranks$/N_i$) would be $50 + 378{\cdot}1 = 428{\cdot}1$. The partition into two states, causing the largest reduction of the within-states variability, is that between 5 and 6, with

$\sum(\sum^2 \text{ranks}/N_i) = 431{\cdot}4$. The associated amount of entropy per state, if base 2 logarithms are used, is:

$$H(\text{S})/s = -\tfrac{1}{2}[\tfrac{11}{16}\log\tfrac{11}{16} + \tfrac{5}{16}\log\tfrac{5}{16}] = 0{\cdot}448 \text{ bit/state.}$$

Keeping this first division fixed, all the possible partitions in three states are tried. The best new division, with the previous division between ranks 5 and 6, falls between ranks 3 and 4. This partition results in:

Ranks	$\sum$ ranks	N_i	$\sum^2$ ranks/N_i
1–3	$1 + (4 \times 2) + 3 = 12$	6	$12^2/6 = 24$
4–5	$(2 \times 4) + (3 \times 5) = 23$	5	$23^2/5 = 105{\cdot}8$
6–10	$6 + 7 + 8 + 9 + 10 = 40$	5	$40^2/5 = 320$

$\sum(\sum^2 \text{ranks}/N_i) = 449{\cdot}8$ and its amount of entropy per state is $0{\cdot}526$ bit/state.

The best partition into four states, considering the two previous divisions, results from a new division between ranks 7 and 8, or between 8 and 9, with $\sum(\sum^2 \text{ranks}/N_i) = 457{\cdot}3$. The entropy per state is $0{\cdot}471$ bit/state, which is less than that for the best partition into three states. It is thus concluded that the optimum stepwise partition of this rank-ordered variable is into three states: state 1 goes from rank 1 to 3, state 2 from rank 4 to 5, and state 3 from rank 6 to 10. The exact procedure results in the same first division, but it positions the second division between ranks 6 and 7. When equally plausible partitions (such as into 4 states above) occur during the stepwise procedure, they can be tried in turn, the final choice being that partition which has the maximum entropy per state. On the other hand, when there are several equally plausible partitions at the end of the procedure, a predetermined decision rule may be used.

After partitioning the data set into three states of respectively 6, 5 and 5 observations, the series of transformed rank-ordered data may then be subjected to contingency periodogram analysis. The fictitious rank-ordered variable was defined to fit the data in Table 1; the corresponding possible series of Fig. 2 would then have the periodogram shown in Fig. 1.

In the next section, a comparison will be made between the contingency periodogram and the periodogram of Schuster (1898), the latter being valid only for metric variables. If a metric variable is to be analysed with the contingency periodogram then it must be partitioned into discrete states. The partitioning procedure already described may be used, because a metric variable is also rank-ordered. The variable is first plotted as a function of ranks, as in Fig. 3 which is equivalent to Table 4; computations are then the same as for rank-ordered variables. Also, the actual values of variable Y may be used instead of their rank, so that the total within-states variance is minimized when

$$\sum_{i=1}^{s} \frac{1}{N_i}\left[\sum_{\text{within state } i} Y\right]^2$$

is maximized. In order to prevent undue effects of the measurement scale on the values of within-state variance, the variable to be partitioned may be normalized or linearized relative to the ecological phenomenon of interest.

In the example of Table 4, the metric variable Y was so defined as to correspond to the rank-ordered variable above. The quantities computed during the partitioning of this variable into classes are shown in Table 5.

Thus this stepwise partitioning procedure makes it possible, for the ecologist who so desires, to analyse with the contingency periodogram all types of data, even those not

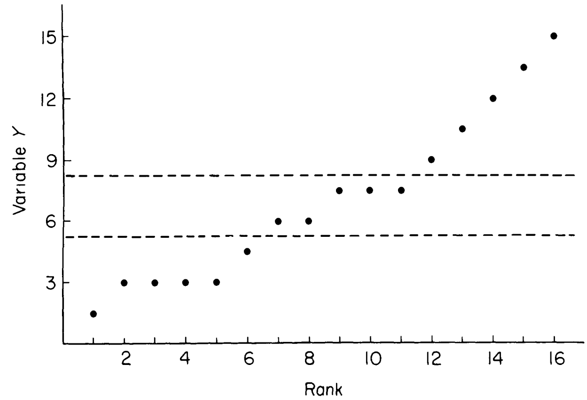

FIG. 3. Rank diagram of fictitious metric variable Y (Table 4). Broken lines correspond to the boundaries between states.

TABLE 4. Fictitious example: $N = 16$ ranked observations of a metric variable.

Value	1·5	3	4·5	6	7·5	9	10·5	12	13·5	15
No. of observations	1	4	1	2	3	1	1	1	1	1

TABLE 5. Stepwise partition of fictitious metric varible Y, from Table 4.

Number of states	Successive state limits	$\sum(\sum^2 Y/N_i)$	$H(S)/s$ (bit/state)
2	Between 7·5 and 9	970·6	0·448
3	Between 4·5 and 6	1012·1	0·526
4	Between 10·5 and 12 Between 12 and 13·5	1028·9	0·471

already divided into states and when there is no *a priori* knowledge as to the best places to make the partitions.

EXAMPLES AND DISCUSSION

In order to assess the value of the contingency periodogram, tests on three data sets are described. The first is an artificial set of quantitative data, of known properties. The second set is of well-understood quantitative field data. These two sets allow comparison of the contingency periodogram with Schuster's periodogram. The prime interest of the contingency periodogram is the analysis of qualitative and also eventually of rank-ordered variables, however, and the third data set is of this kind. It is of field data from a situation where the ecology is not well understood.

For the first test, an artificial series of $N = 200$ data was generated by adding together a sine of period $T = 9$, another sine of period $T = 15$ with an amplitude equal to half that of the previous one, and a noise component with an amplitude half that of the two sines. These three components and the resulting series are shown in Fig. 4. This artificial metric variable was partitioned into states, using the stepwise procedure described in the previous section: the resulting state limits are drawn on Fig. 4. The artificial series was then analysed into a Schuster periodogram (original data), and into a contingency periodogram (partitioned data). The results of both analyses are recorded in Fig. 5, where it is evident that periods $T = 9$ and $T = 15$ (and their harmonics on the contingency periodogram) are identified on both periodograms. Critical values should be used only to identify which

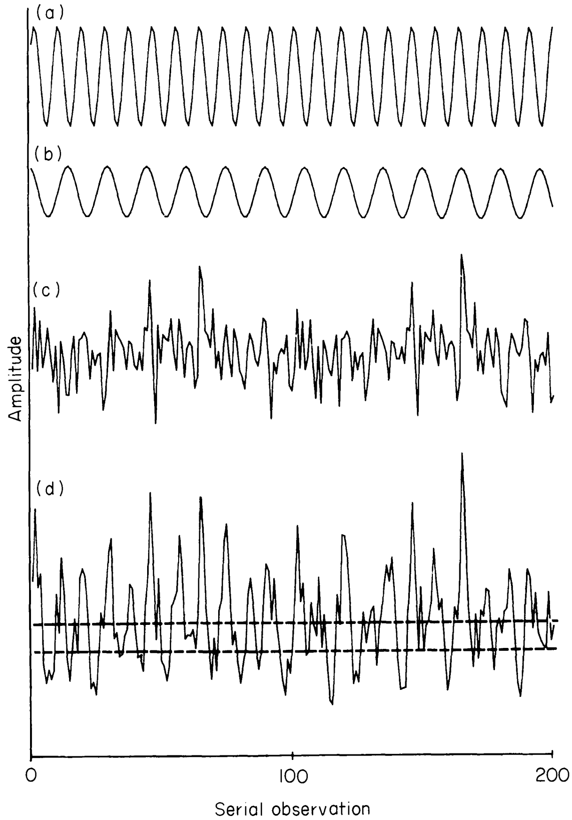

FIG. 4. Artificial data series generated by adding together (a) a sine of period $T = 9$, (b) another sine of period $T = 15$, and (c) a noise component. The resulting series (d) is shown with the boundaries between the three states (broken lines) computed from the stepwise partitioning procedure.

periodogram values are too small to be significantly different from zero, as the periodogram values exceeding their critical value are not necessarily significantly different from zero.

The identification of both periods, despite the large noise component, demonstrates the robustness of the contingency periodogram to departures from smooth oscillation. Even the complete obliteration of a full period, for instance by allocating all the data points between 32 and 41 ($T = 9$) to state 2, changes only slightly the value of $H(S \cap X)$ (for period $T = 9$) in the periodogram (from 0·358 to 0·349 nat). These are indications of the potential usefulness of the method with perturbed series, as are sometimes encountered in field ecology.

For the second test we use a series of $N = 126$ data from samples taken in the St Lawrence Estuary in mid-June 1976 (Fig. 6). These data are measurements of the photosynthetic capacity of natural phytoplankton, sampled hourly at 1 m depth and incubated under a saturating light flux of 375 $\mu E\ m^{-2}\ s^{-1}$ (Fréchette & Legendre, in press). Two periods, $T = 11{\cdot}5$–$12{\cdot}0$ h and $T = 24$ h, are identified (Fig. 7) on both contingency (partitioned data) and Schuster's (original metric data) periodograms. These periods, when found at a sampling station located upstream in the Estuary, were interpreted by Demers & Legendre (1979) as a circadian endogenous rhythm in the photosynthetic activity, combined with a tidal pattern.

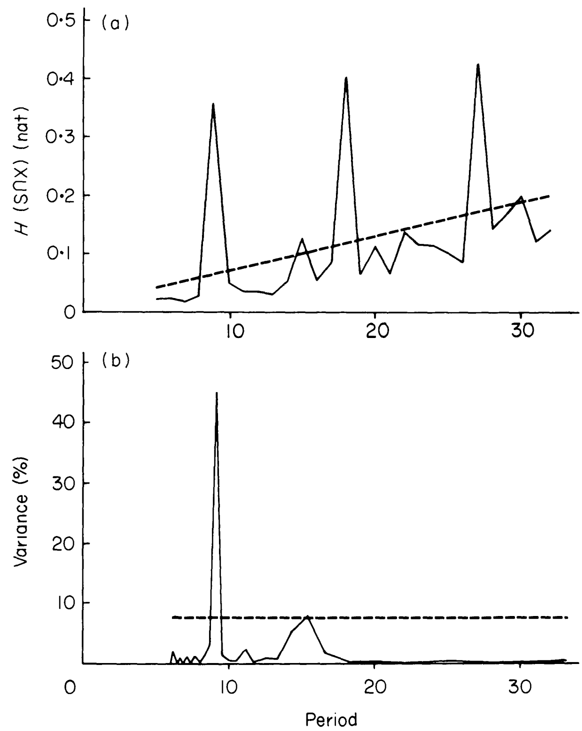

FIG. 5. (a) Contingency periodogram and (b) Schuster periodogram of the data series in Fig. 4(d). Dashed line shows the critical values at the 0·05 probability level. The variable $H(S \cap X)$, a measure of entropy, and the unit nat, are explained in the text.

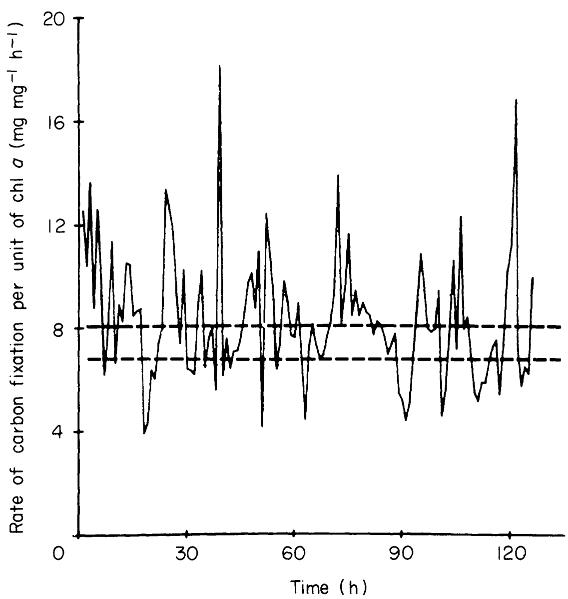

FIG. 6. Photosynthetic capacity of natural phytoplankton in the St Lawrence Estuary, N. America. The boundaries between the three states computed from the stepwise partitioning procedure are shown by broken lines.

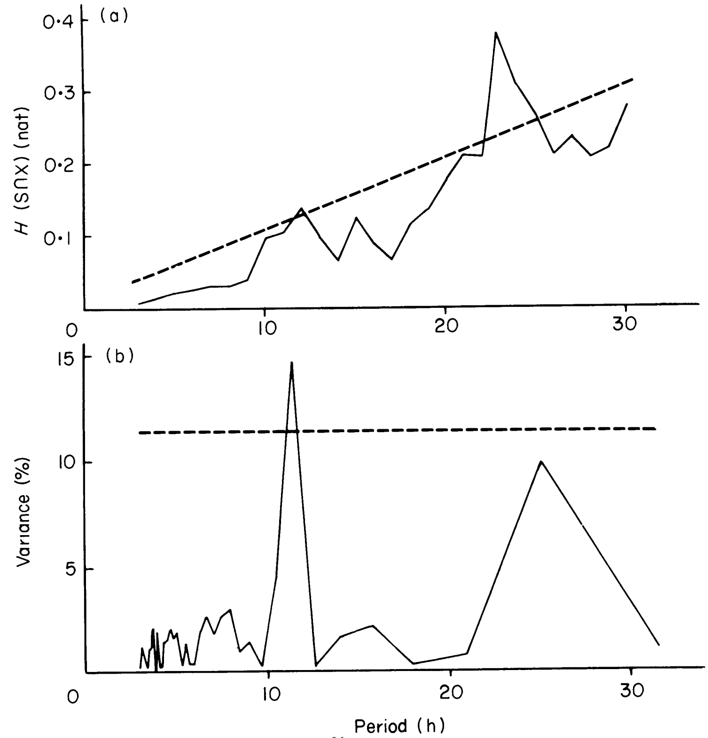

FIG. 7. (a) Contingency periodogram and (b) Schuster periodogram of the data series in Fig. 6. Broken line shows the critical values at the 0·05 probability level. The variable $H(S \cap X)$, a measure of entropy, and the unit nat, are explained in the text.

Ecological data series often consist of counts of many different species, which are usually analysed for rhythms by either pooling all the species to get a single variate (total counts), or alternatively by considering each species separately. Another approach, which we use as our third example, makes it possible to analyse the multivariate series as such. The multi-species data are first reduced to a single multi-state qualitative variate by clustering the samples (every data point of the series is then allocated to one of the clusters, each cluster being one of the states of the synthetic qualitative variate). Then the contingency periodogram is computed on the resulting multi-state qualitative data series.

A series of 175 phytoplankton samples were taken at 1-h intervals at a depth of 1 m, at an anchor station in the St Lawrence Estuary in July 1975. The density of each species was recorded and total density is analysed by Lafleur, Legendre & Cardinal (1979). They divided the series into two parts, corresponding respectively to the neap tide and the spring tide; only the neap tide (first 80 h) are considered here. The data (Fig. 8) were in the form of six functional taxonomic groups (1) *Thalassiosira pacifica* Gran & Angst, (2) *Cyclotella striata* (Kutz.) Grun., (3) other centric diatoms, (4) pennate diatoms, (5) Chlorophyceae, and (6) Chrysophyceae. The index of association (Whittaker 1952) was computed between the samples and the resulting matrix was clustered using Lance & Williams' (1966, 1967) flexible strategy with $\beta = -0{\cdot}25$. The allocation of the eighty samples to the five clusters (states) thus computed is shown in Fig. 8. A period $T = 3$ is identified on the contingency periodogram (Fig. 9), with 9 harmonics (up to $T = 29$). This is rather surprising, as the total cell density show a periodicity of about 10–12 h, while the

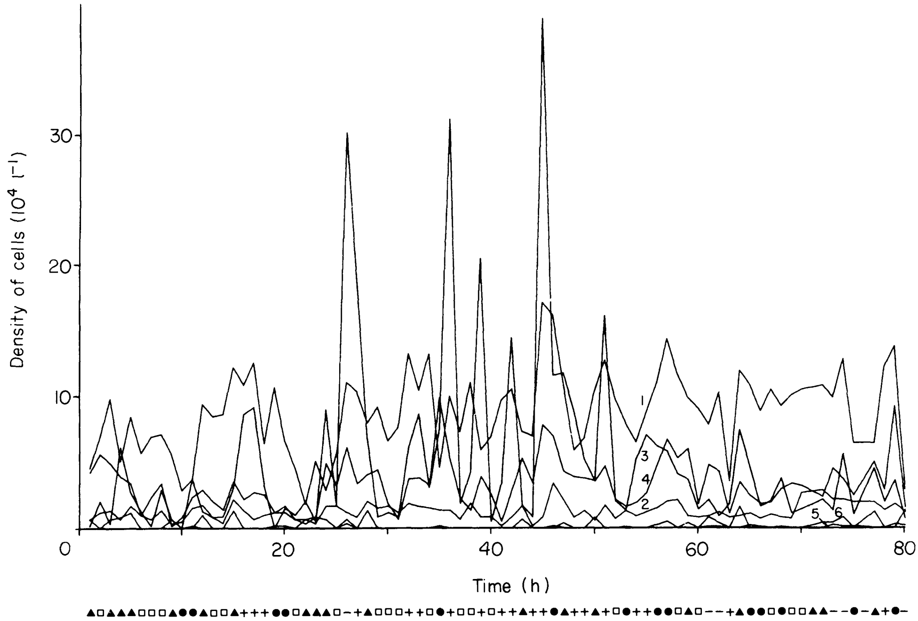

FIG. 8. Phytoplankton density (six taxonomic groups labelled 1–6, see text) from the St Lawrence Estuary, N. America. The allocation of each sample of the series to one of five clusters is indicated under the abscissa (five different symbols). Results of Lafleur, Legendre & Cardinal (1979).

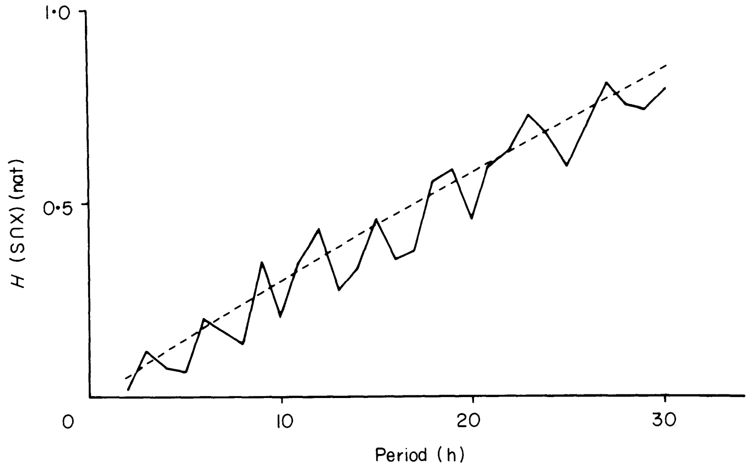

FIG. 9. Contingency periodogram of the data in Fig. 8. Broken line shows the critical values at the 0·05 probability level. The variable $H(S \cap X)$, a measure of entropy, and the unit nat, are explained in the text.

physical environment (water temperature) oscillates with a periodicity of about 6 h. The 6-h period is explained by Lafleur, Legendre & Cardinal (1979) in terms of an internal wave generated at a sub-surface sill close to the sampling station. At this same station, Fortier *et al.* (1978) found evidence of a 190 min oscillation in both the total cell density and physical variables during neap tide in August 1975. The observed 3-h periodicity in the relative abundance of phytoplankton taxa therefore appears as an indication of rapidly oscillating surface water. Along horizontal gradients, the relative abundance of the taxa changes probably more rapidly than does the total density of cells, which make the taxa a better indicator of subtle horizontal oscillations of the water masses.

Thus the contingency periodogram makes it possible to efficiently analyse rhythms within series of qualitative or rank-ordered data, or series of multivariate data reduced to qualitative multi-state data. By using the partitioning algorithm, rank-ordered data are reduced to a small number of states prior to contingency periodogram analysis: these same data may therefore be sampled directly with much lower precision, with consequent savings of work. Nonmetric variables (qualitative, and also rank-ordered) are commonly recorded in field ecology: using the contingency periodogram these previously neglected variables, as well as multi-species data, may now be analysed just as quantitative variables have been for so long.

A. Vaudor has written an efficient, transportable program, in the PASCAL language. This program first partitions the variable into classes (if needed) and then computes the contingency periodogram. Listings are available, free of charge, from the third author.

ACKNOWLEDGMENTS

We are grateful to Dr C. L. Trump and Y. Roy (Groupe Interuniversitaire de Recherches Océanographiques du Québec (GIROQ), Université Laval), and M. Etienne (Station Zoologique de Villefranche-sur-Mer), for help with computing. We also thank Dr J. Dodson and Dr J. Himmelman (GIROQ) for their suggestions, and I. Lamontagne, C. Onteniente and L. Roy for drafting the figures. The work was supported by grants from the Natural Sciences and Engineering Research Council of Canada and the Ministère de l'Education du Québec.

REFERENCES

Ables, J. G. (1974). Maximum entropy spectral analysis. *Astronomy and Astrophysics, Supplement Series*, **15,** 383–393.

Anderson, T. W. (1971). *The Statistical Analysis of Time Series.* Wiley & Sons, New York.

Burg, J. P. (1967). Maximum entropy spectral analysis. Paper presented at the 37th Annual Meeting of Exploration Geophysicists, October 31, Oklahoma City, Oklahoma.

Burg, J. P. (1975). *Maximum entropy spectral analysis.* Ph.D. thesis, Stanford University, California.

Colwell, R. K. (1974). Predictability, constancy, and contingency of periodic phenomena. *Ecology*, **55,** 1148–1153.

Demers, S. & Legendre, L. (1979). Effets des marées sur la variation circadienne de la capacité photosynthétique du phytoplancton de l'estuaire du Saint-Laurent. *Journal of Experimental Marine Biology and Ecology*, **39,** 87–99.

Enright, J. T. (1965). The search for rhythmicity in biological time-series. *Journal of Theoretical Biology*, **8,** 426–468.

Fortier, L. & Legendre, L. (1979). Le contrôle de la variabilité à court terme du phytoplancton estuarien: stabilité verticale et profondeur critique. *Journal of the Fisheries Research Board of Canada*, **36,** 1325–1335.

Fortier, L., Legendre, L., Cardinal, A. & Trump, C. L. (1978). Variabilité à court terme du phytoplancton de l'estuaire du Saint-Laurent. *Marine Biology*, **46,** 349–354.

Fréchette, M. & Legendre, L. (In press). Phytoplankton photosynthetic response to light in an internal tide dominated environment. *Estuaries.*

Frontier, S. (1969). Sur une méthode d'analyse faunistique rapide du zooplancton. *Journal of Experimental Marine Biology and Ecology*, **3,** 18–26.

Frontier, S. (1973). Evaluation de la quantité totale d'une catégorie d'organismes planctoniques dans un secteur néritique. *Journal of Experimental Marine Biology and Ecology*, **12,** 299–304.

Kedem, B. (1980). Binary time series. *Lecture Notes in Pure and Applied Mathematics.* Vol. 52. Marcel Dekker, New York.

Kendall, M. G. (1948). *Rank Correlation Methods.* Charles Griffin, London.

Kendall, M. G. & Stuart, A. (1966). *The Advanced Theory of Statistics.* Vol. 3. Hafner, New York.

Kullback, S. (1959). *Information Theory and Statistics.* Wiley and Sons, New York.

Lafleur, P. E., Legendre, L. & Cardinal, A. (1979). Dynamique d'une population estuarienne de Diatomées planctoniques: effet de l'alternance des marées de morte-eau et de vive-eau. *Oceanologica Acta,* **2,** 307–316.

Lance, G. N. & Williams, W. T. (1966). A generalized sorting strategy for computer classifications. *Nature, London,* **212,** 218.

Lance, G. N. & Williams, W. T. (1967). A general theory of classificatory sorting strategies. I. Hierarchical systems. *Computer Journal,* **9,** 373–380.

Legendre, L. & Legendre, P. (1979). *Ecologie Numérique. 2. La Structure des Données Écologiques.* Masson, Paris.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K. & Bent, D. H. (1975). *SPSS—Statistical Package for the Social Sciences,* 2nd edn. McGraw-Hill, New York.

Platt, T. & Denman, K. L. (1975). Spectral analysis in ecology. *Annual Review of Ecology and Systematics,* **6,** 189–210.

Schuster, A. (1898). On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Terrestrial Magnetism,* **3,** 13–41.

Shannon, C. E. (1948). A mathematical theory of communications. *Bell System Technical Journal,* **27,** 379–423, 623–656.

Whittaker, E. T. & Robinson, G. (1924). *The Calculus of Observations—A Treatise on Numerical Mathematics.* Blackie and Son, London.

Whittaker, R. H. (1952). A study of summer foliage insect communities in the Great Smoky Mountains. *Ecological Monographs,* **22,** 1–44.

(*Received* 24 *March* 1981)