Appl. Statist. (1991)
40, *No.* 3, *pp.* 423–434

Nonparametric Comparison of Cumulative Periodograms

By PETER J. DIGGLE†

Lancaster University, UK

and NICHOLAS I. FISHER

CSIRO, Lindfield, Australia

[Received May 1988. Final revision April 1990]

SUMMARY

Motivated by a problem in the analysis of hormonal time series data, this paper proposes a simple graphical method for comparing two periodograms and describes a new nonparametric approach to testing the hypothesis that the two underlying spectra are the same. Simulation studies show that the new test has power characteristics that are competitive with existing procedures. The relative merits of nonparametric and semiparametric tests are discussed.

Keywords: Cumulative periodogram; Graphical method; Hormonal time series; Randomization test; Spectral analysis

1. Introduction

Coates and Diggle (1986) have considered several periodogram-based tests of the hypothesis that two independent time series $\{x_t: t=1, \ldots, n\}$ and $\{y_t: t=1, \ldots, n\}$ are generated from the same underlying stationary process. In particular, they proposed a semiparametric procedure based on the model that the underlying spectral densities $\lambda_x(\omega)$ and $\lambda_y(\omega)$ are related via the equation

$$\lambda_y(\omega) = \lambda_x(\omega) \exp(\alpha + \beta\omega + \gamma\omega^2). \qquad (1)$$

Note that the spectral density of a stationary process $\{X_t\}$ with autocovariance function $\gamma(u) = \text{cov}(X_t, X_{t-u})$ is

$$\lambda_x(\omega) = \sum_{u=-\infty}^{\infty} \gamma(u) \exp(-iu\omega).$$

Diggle (1985) set equation (1) in the context of the generalized linear model (McCullagh and Nelder, 1983), to give an immediate extension to the problem of comparing $k > 2$ periodograms.

Coates and Diggle (1986) included two nonparametric tests in their comparisons.

†*Address for correspondence*: Mathematics Department, School of Engineering, Computing and Mathematical Sciences, Lancaster University, Lancaster, LA1 4YF, UK.

 0035–9254/91/40423 $2.00

However, one test, based on the range of periodogram ratios, is extremely weak whereas the other, based on cumulative sums of transformed periodogram ratios, has an undesirable artefact: its result depends on the arbitrary labelling of the two series $\{x_t\}$ and $\{y_t\}$.

In this paper, we approach the problem of comparing the spectra of two time series via their cumulative periodograms. This leads naturally to an informative graphical procedure and to a general class of nonparametric tests.

Endocrinology provides a rich source of problems of this kind, one of which is illustrated in Fig. 1. This diagram shows four time series of luteinizing hormone (LH) concentrations measured in blood samples drawn at 10-min intervals from a healthy adult female. Each series is of length $n=44$; two series are taken in the early follicular phase of the subject's menstrual cycle and two in the late follicular phase. It is known that LH is secreted in an irregular pulsatile fashion, rather than in a steady stream (Dr A. Murdoch, personal communication). Of interest is whether the frequency characteristics of this pulsatile release pattern are the same in the two phases of the cycle.

In Section 2 we summarize the required distribution theory for the periodogram and introduce the plotting procedure and associated tests. In Section 3 we present the results of a simulation study of power, designed to be directly comparable with the study reported in Coates and Diggle (1986). In Section 4 we discuss some possible modifications to the test. Finally, in Section 5 we apply the method to the endocrine data of Fig. 1, to confirm that the pulsatile release patterns are different in the two phases of the cycle.

2. Basic Distribution Theory

Let $I_x(\omega)$ and $I_y(\omega)$ denote the periodogram ordinates of $\{x_t: t=1, \ldots, n\}$ and $\{y_t: t=1, \ldots, n\}$ respectively, each evaluated at frequencies ω of the form $\omega_j = 2\pi j/n$, where $j=1, \ldots, m$ and $m = [(n-1)/2]$, i.e.

$$I_x(\omega) = (2\pi n)^{-1} \left| \sum_{t=1}^{n} x_t \exp(-\mathrm{i}\omega t) \right|^2$$

with an analogous expression for $I_y(\omega)$. The corresponding normalized cumulative periodogram is

$$F_x(\omega_j) = \sum_{i=1}^{j} I_x(\omega_i) \Big/ \sum_{i=1}^{m} I_x(\omega_i)$$

and analogously for $F_y(\omega_j)$.

Bartlett (1954) proposed a plot of $F_x(\omega_j)$ against ω_j to assess departure from white noise, for which $\lambda_x(\omega)$ is constant. See also Jenkins and Watts (1968). It is natural to extend this idea to the comparison of two sample periodograms via a plot of $F_y(\omega_j)$ against $F_x(\omega_j)$. The use of *normalized* cumulative periodograms implies that we wish to detect only shape differences between the two underlying spectra. By analogy with the behaviour of *P–P* plots for comparing two samples (see, for example, Quade (1973)), we can expect the resulting cumulative periodogram (CP) plot to behave in characteristic manner when there is a shift in power in one periodogram compared

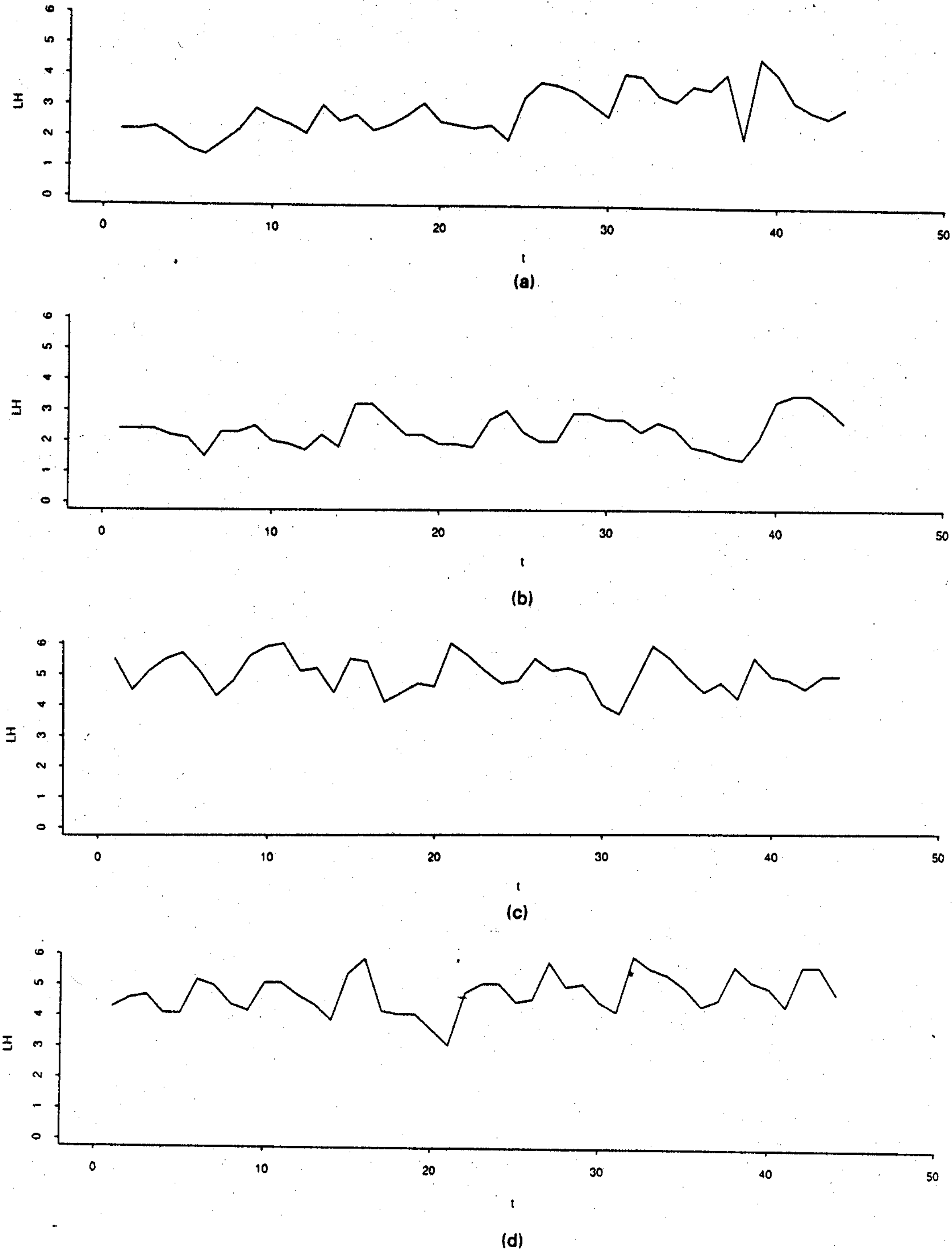

Fig. 1. Four time series of LH concentrations in blood samples: each series consists of $n = 44$ values derived from blood samples taken at 10-min intervals from a healthy adult female (data collected by Dr A. Murdoch): (a) early follicular phase, first cycle; (b) early follicular phase, second cycle; (c) late follicular phase, first cycle; (d) late follicular phase, second cycle

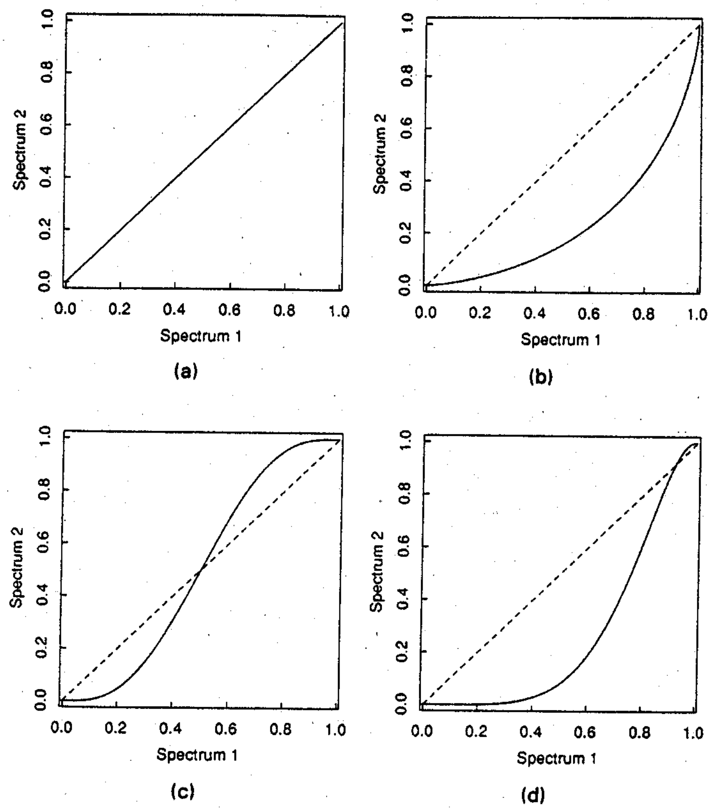

Fig. 2. Characteristic behaviour of CP plots: (a) identical spectra; (b) shift of power to higher frequencies; (c) greater dispersion of power; (d) shift in power with increased dispersion

with the other, and when one periodogram has more dispersed power than the other. Examples of such behaviour are shown in Fig. 2.

A further attractive feature of this graphical method is that it obviates the need for any subjective smoothing, such as is required when comparing spectral density estimates. Basing the plot on CPs automatically induces some smoothing.

It will often be desirable to supplement informal inference from a CP plot with a formal test, either of a general nature or with some specific alternative hypothesis in mind. For a single periodogram being compared with theoretical white noise, the statistical model under the null hypothesis is equivalent to that of testing a sample of independent and identically distributed measurements, for which Jenkins and Watts (1968) recommended use of the Kolmogorov–Smirnov statistic. A new feature occurs when comparing two periodograms: the CP ordinates no longer correspond to the order statistics from an independent and identically distributed sample.

Assume that $\{X_t\}$ and $\{Y_t\}$ can each be represented as a stationary general linear process (Priestley (1981), section 3.5.7) with independent, identically distributed Gaussian innovations. Then, asymptotically,

$$I_x(\omega_j) \sim \lambda_x(\omega_j)\chi_2^2/2$$

and

$$I_y(\omega_j) \sim \lambda_y(\omega_j)\chi_2^2/2.$$

Furthermore, for $i \neq j$, $I_x(\omega_i)$ and $I_x(\omega_j)$ are asymptotically independent (Priestley (1981), p. 466), as are $I_y(\omega_i)$ and $I_y(\omega_j)$. Since the two series $\{x_t\}$ and $\{y_t\}$ are independent realizations, each $I_x(\omega_i)$ must be independent of all $I_y(\omega_j)$ and vice versa. Thus, under the null hypothesis

$$H_0\colon \lambda_x(\omega) = \lambda_y(\omega) \qquad (0 \leqslant \omega \leqslant \pi)$$

$I_x(\omega_j)$ and $I_y(\omega_j)$ are independent and identically distributed, for each $j = 1, \ldots, m$, whereas $I_x(\omega_1), \ldots, I_x(\omega_m)$ are merely asymptotically independent. This suggests a *general* class of approximate randomization tests for H_0 for comparing the spectra of two stationary processes. Suppose that these asymptotic statements are exactly true for finite n, and let d be *any measure* of the distance between $F_x(\ )$ and $F_y(\ )$. Then, under H_0, the distribution of d will be invariant under all 2^m possible interchanges of $I_x(\omega_j)$ with $I_y(\omega_j)$, for $j = 1, \ldots, m$.

In practice, it will not be feasible to calculate the randomization distribution of d, but it can be approximated very adequately by calculating $d_2, \ldots, d_s$ for some large number $s-1$ of random interchanges of periodogram ordinates at each frequency, and by calculating the significance probability of the observed d-value, d_1 say, as the proportion of the values $d_1, \ldots, d_s$ at least as large as d_1.

Two obvious choices for d are the Kolmogorov–Smirnov and Cramér–von Mises statistics

$$D_m = \sup |F_x(\omega) - F_y(\omega)| \tag{2}$$

and

$$W_m = \int_0^{\pi} \{F_x(\omega) - F_y(\omega)\}^2 \, d\bar{F}(\omega) \tag{3}$$

where $\bar{F}(\omega) = 0.5\{F_x(\omega) + F_y(\omega)\}$, each of which yields an omnibus test in terms of detecting departures from H_0. In practice, we evaluate statistic (2) or (3) from the discrete sets of values $F_x(\omega_j)$ and $F_y(\omega_j)$. The next section gives the results of a comparative study of these statistics. As noted earlier, we are using the normalized CPs because our interest is in detecting shape rather than scale differences between the two underlying spectra. Thus, although the null hypothesis is strictly that $\lambda_x(\omega) = \lambda_y(\omega)$ for all ω, we deliberately use test statistics which are insensitive against alternatives of the form $\lambda_x(\omega) = c\,\lambda_y(\omega)$ for all ω, where $c \neq 1$. If we wished to detect purely scale differences between the two underlying processes, we could work with non-normalized CPs. However, the statistics D_m and W_m then seem less natural and a more direct attack, using for example the semiparametric approach of Coates and Diggle (1986) would seem to be preferable.

3. Power Study

3.1. *Outline of Study*

We have compared the power characteristics of the proposed tests based on the statistics D_m and W_m defined by equations (2) and (3) by applying them to simulations of pairs of autoregressive (AR(p)) processes,

$$X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + Z_t,$$

or moving average (MA(q)) processes,

$$X_t = Z_t + \sum_{i=1}^{q} \beta_i Z_{t-i},$$

where in each case $\{Z_t\}$ is a sequence of mutually independent $N(0, \sigma^2)$ random variables, i.e. Gaussian white noise. The study is somewhat more extensive than that reported in Coates and Diggle (1986) but includes cases considered by Coates and Diggle to enable direct comparisons with their results.

Each implementation of the Monte Carlo test used 99 randomizations, i.e. $s = 100$. In applications, a larger value such as $s = 1000$ would alleviate the slight loss of power due to the Monte Carlo implementation, often at negligible cost in terms of increased computing. For further comments on the value of s, see Marriott (1979).

The first phase of the study consisted of checks on the nominal significance levels of the tests, in view of the approximate nature of the underlying distribution theory. For this phase, we performed 1000 replicate simulations of each pair of processes to give reasonably precise estimates of actual significance levels. The second phase consisted of obtaining estimates of power when the two underlying processes are different. For this phase, we performed only 100 replicate simulations of each pair of processes since good coverage of a range of cases was more important than precise estimation of power in any particular case. Generally, we simulated pairs of processes with values of σ^2 adjusted so that $\{X_t\}$ and $\{Y_t\}$ have the same variance. The power of the test therefore derives from shape differences between the two underlying spectra. However, in a final phase of the study we simulated pairs of processes which differed only in their variances, in part to assess the robustness of the test to purely scale differences in the underlying spectra.

Pseudorandom numbers were generated using the algorithm of Wichmann and Hill (1982).

3.2. *Checks on Nominal Significance Levels*

The results for this phase of the study are summarized in Table 1. Generally, for underlying processes which are AR(1) or MA(1), the Cramér–von Mises statistic W_m gives a test which is conservative, at least for the larger nominal significance levels, but becoming less so as the strength of the serial correlation within each process increases. The Kolmogorov–Smirnov statistic D_m gives an approximately valid test except for the combination of small n and very strong autoregressive dependence, when the test appears to be slightly liberal. The standard error of each entry in Table 1 is $\sqrt{\{p(1-p)/1000\}} \approx 0.032\sqrt{p}$, where p is the actual significance level, i.e. 0.1, 0.05 and 0.01.

3.3. *Power Estimates*

In view of the results of Section 3.2, we present power estimates only for the test based on the Kolmogorov–Smirnov statistic D_m.

TABLE 1
Estimated sizes of tests of nominal sizes 0.1, 0.05 and 0.01

n	α_1 or β_1†	D_m *for the following sizes:*			W_m *for the following sizes:*		
		0.100	*0.050*	*0.010*	*0.100*	*0.050*	*0.010*
$\{X_t\}$ and $\{Y_t\}$ each AR(1)							
64	0.0	0.100	0.050	0.013	0.042	0.021	0.012
	0.1	0.106	0.047	0.013	0.040	0.023	0.011
	0.5	0.099	0.049	0.011	0.046	0.024	0.012
	0.9	0.124	0.068	0.020	0.064	0.032	0.020
256	0.0	0.093	0.043	0.011	0.038	0.021	0.014
	0.1	0.089	0.045	0.012	0.037	0.023	0.011
	0.5	0.089	0.045	0.011	0.039	0.020	0.006
	0.9	0.093	0.053	0.012	0.038	0.020	0.010
1024	0.0	0.099	0.051	0.008	0.040	0.024	0.014
	0.1	0.101	0.054	0.009	0.041	0.022	0.012
	0.5	0.100	0.055	0.014	0.045	0.023	0.011
	0.9	0.111	0.066	0.007	0.043	0.023	0.012
$\{X_t\}$ and $\{Y_t\}$ each MA(1)							
64	0.1	0.108	0.047	0.012	0.039	0.022	0.013
	0.5	0.097	0.051	0.007	0.039	0.023	0.011
	0.9	0.102	0.052	0.005	0.040	0.024	0.012
256	0.1	0.090	0.045	0.012	0.036	0.023	0.011
	0.5	0.088	0.040	0.010	0.036	0.021	0.011
	0.9	0.086	0.039	0.011	0.034	0.019	0.009
1024	0.1	0.101	0.050	0.009	0.042	0.022	0.012
	0.5	0.101	0.057	0.012	0.042	0.025	0.008
	0.9	0.102	0.061	0.014	0.042	0.021	0.011

†α_1 for the AR(1) process, β_1 for the MA(1) process.

Table 2 gives power estimates for series of length $n=64$ in each of the following cases:

(a) white noise *versus* AR(1), $\alpha_1 > 0$;
(b) white noise *versus* AR(2), $\alpha_1 = 0$, $\alpha_2 > 0$;
(c) white noise *versus* AR(3), $\alpha_2 = 0$, $\alpha_1 = -\alpha_3 > 0$;
(d) AR(1), $\alpha = 0.5$, *versus* AR(1), $\alpha_1 > 0$.

By comparison with results in Table 2 of Coates and Diggle (1986) for the quadratic likelihood ratio test which they recommend, the D_m-test has comparable power in case (a), slightly less power in case (b) but rather more power in cases (c) and (d). The form of the cumulative spectrum in case (b) suggests that a more powerful test would be one which weights heavily the tails of the distribution, e.g. the two-sample Anderson–Darling statistic (Pettitt, 1976). In fact, Coates and Diggle deliberately chose cases (b) and (c) to be favourable and unfavourable respectively to their recommended test. Also, Coates and Diggle included results for $\alpha_1 = 0.8$ in case (c), which we have excluded because it gives a non-stationary process.

Table 3 gives power estimates for series of length $n = 1024$ in the same four cases, but with narrower ranges of parameter values to reflect the generally higher power of the test. By comparison with results in Table 3 of Coates and Diggle (1986), the D_m-test has generally better power in cases (a), (c) and (d) but is still outperformed by the quadratic likelihood ratio test in case (b).

TABLE 2
Estimated power of D_m-tests, series length $n = 64$

α_1 or α_2	Results for the following sizes: 0.10	0.05	0.01
White noise versus AR(1), $\alpha_1 > 0$			
0.2	0.20	0.13	0.02
0.4	0.58	0.50	0.20
0.6	0.87	0.81	0.54
0.8	0.94	0.90	0.70
White noise versus AR(2), $\alpha_1 = 0$, $\alpha_2 > 0$			
0.2	0.11	0.07	0.00
0.4	0.15	0.11	0.02
0.6	0.34	0.20	0.06
0.8	0.56	0.39	0.21
White noise versus AR(3), $\alpha_2 = 0$, $\alpha_1 = -\alpha_3 > 0$			
0.2	0.28	0.19	0.04
0.4	0.66	0.56	0.31
0.6	0.85	0.72	0.45
AR(1), $\alpha = 0.5$ versus AR(1), $\alpha_1 > 0$			
0.1	0.65	0.52	0.21
0.3	0.30	0.19	0.07
0.5	0.10	0.04	0.00
0.7	0.45	0.36	0.17
0.9	0.60	0.53	0.34

Finally, Table 4 summarizes the performance of the D_m-test when both $\{X_t\}$ and $\{Y_t\}$ are white noise processes but with different standard deviations, $\sigma_y \neq \sigma_x$. Two versions of the test are given, based on normalized and non-normalized CPs respectively. The top part of Table 4 shows that in the former case the test is biased, becoming more so as the discrepancy between σ_y and σ_x increases. The bottom part of Table 4 confirms that we could obtain a powerful test against this alternative by using non-normalized CPs. However, for the reasons given at the end of Section 2 we do not recommend this.

4. Discussion

The simulation studies reported in Section 3 suggest that the new nonparametric test based on D_m is often competitive with existing semiparametric tests. In general, a semiparametric approach is likely to give a more powerful test when the parametric assumptions are approximately valid and vice versa. We believe that there is merit in using a nonparametric procedure at least in the exploratory phase of the data analysis.

It is possible to incorporate the test based on D_m into the CP plot by displaying a confidence band of chosen level $100(1-2\alpha)\%$. Let $D(\alpha)$ be the upper $100\alpha\%$ quantile of the values $d_1, \ldots, d_s$, corresponding to CPs $F_{x\alpha}(\omega)$ and $F_{y\alpha}(\omega)$ say. Then the curves

$$(u, F_{y\alpha}\{F_{x\alpha}^{-1}(u)\}) \qquad (0 \leqslant u \leqslant 1)$$

and

TABLE 3
Estimated power of D_m-tests, series length $n = 1024$

α_1 or α_2	Results for the following sizes: 0.10	0.05	0.01
White noise versus AR(1), $\alpha_1 > 0$			
0.05	0.32	0.18	0.06
0.10	0.70	0.56	0.32
0.15	0.95	0.89	0.64
0.20	0.99	0.99	0.90
0.25	1.00	1.00	0.98
White noise versus AR(2), $\alpha_1 = 0$, $\alpha_2 > 0$			
0.05	0.12	0.07	0.00
0.10	0.26	0.13	0.02
0.15	0.60	0.36	0.07
0.20	0.85	0.75	0.22
0.25	0.94	0.88	0.58
White noise versus AR(3), $\alpha_2 = 0$, $\alpha_1 = -\alpha_3 > 0$			
0.05	0.38	0.26	0.08
0.10	0.82	0.76	0.42
0.15	0.98	0.97	0.80
0.20	1.00	1.00	0.97
0.25	1.00	1.00	1.00
AR(1), $\alpha = 0.5$ versus AR(1), $\alpha_1 > 0$			
0.35	0.95	0.88	0.61
0.40	0.60	0.47	0.24
0.45	0.25	0.17	0.09
0.50	0.09	0.04	0.02
0.55	0.22	0.14	0.02
0.60	0.68	0.52	0.28
0.65	0.90	0.88	0.68

TABLE 4
Estimated power of D_m-tests when $\{X_t\}$ and $\{Y_t\}$ are each white noise, but with different variances

σ_y/σ_x	Results for the following sizes, $n = 64$: 0.10	0.05	0.01	Results for the following sizes, $n = 1024$: 0.10	0.05	0.01
Tests based on normalized CPs						
1.1	0.08	0.02	0.00	0.08	0.05	0.00
1.2	0.07	0.01	0.00	0.07	0.04	0.00
1.3	0.07	0.01	0.00	0.05	0.03	0.00
1.4	0.04	0.01	0.00	0.05	0.02	0.00
1.5	0.02	0.01	0.00	0.04	0.01	0.00
Tests based on non-normalized CPs						
1.1	0.22	0.10	0.03	0.89	0.79	0.64
1.2	0.40	0.31	0.10	1.00	1.00	0.99
1.3	0.60	0.48	0.23	1.00	1.00	1.00
1.4	0.76	0.64	0.37	1.00	1.00	1.00
1.5	0.86	0.79	0.51	1.00	1.00	1.00

$$(u, 2u - F_{y\alpha}\{F_{x\alpha}^{-1}(u)\}) \qquad (0 \leqslant u \leqslant 1)$$

provide bounds which the CP plot should not transgress if a test of level 2α is to be accepted. More generally, the methods of Doksum and Sievers (1976) and Switzer (1976) can be adapted as above to provide a simultaneous confidence band around the estimated shift difference, $F_y^{-1}\{F_x(\omega)\} - \omega$, between the two CPs.

We note two possible extensions to the basic test. Firstly, the randomization test could equally well be used to compare spectral estimates computed as the averages of r periodograms in each of two experimental groups. Note the requirement of equal replication in the two groups. If this were violated, we would need to revert to a semiparametric approach such as that in Coates and Diggle (1986). Secondly, the basic idea could be extended to a comparison between $k > 2$ periodograms, using for example the k-sample extension of the Kolmogorov–Smirnov statistic given by Kiefer (1959).

Dunstan and Harris (personal communication) have developed a different nonparametric approach to testing for equality of two estimated spectra. Their test statistics take the form

$$\sum [\hat{r}(\omega_j) + \{\hat{r}(\omega_j)\}^{-1}]$$

where $\hat{r}(\omega) = \hat{f}_x(\omega)/\hat{f}_y(\omega)$ and $\hat{f}_x(\omega)$ and $\hat{f}_y(\omega)$ are smoothed estimates of $f_x(\omega)$ and $f_y(\omega)$ respectively. Their simulations also indicate that with appropriate smoothing the test can be competitive with the semiparametric approach recommended by Coates and Diggle (1986).

5. Application

We now describe an analysis of the endocrine data shown in Fig. 1 and compare the result with the semiparametric analysis recommended by Coates and Diggle (1986). Recall that the data consist of four time series of LH concentrations in blood samples drawn at 10-min intervals from a healthy adult female. Each series is of length $n = 44$; two series are taken from the early follicular phase of the subject's menstrual cycle and two from the late follicular phase. Fig. 3 shows the CP plot for the average periodograms in the two phases. A comparison with Fig. 2 is informative: there is a strong implication that between the two phases of the menstrual cycle the power shifts from lower to higher frequencies with accompanying increased dispersion. The value of the D_m-statistic is 0.322. The formal test gives a p-value of 0.021, indicating a shape difference between the two underlying spectra. For comparison, the semiparametric procedure of Coates and Diggle (1986), assuming a quadratic form for $\log\{\lambda_x(\omega)/\lambda_y(\omega)\}$, gives a p-value of 0.020.

Fig. 4(a) shows an estimate of $\log\{\lambda_x(\omega)/\lambda_y(\omega)\}$ for these data, constructed as $\log\{\hat{\lambda}_x(\omega)/\hat{\lambda}_y(\omega)\}$ where $\hat{\lambda}_x(\omega)$ is a three-point moving average of the average periodogram in the early follicular phase and similarly for $\hat{\lambda}_y(\omega)$. Pointwise 90% confidence limits are based on the observation that the approximate sampling distribution of each $\hat{\lambda}_x(\omega)/\hat{\lambda}_y(\omega)$ is proportional to an F-variate with degrees of freedom 12 and 12, and constant of proportionality $\lambda_x(\omega)/\lambda_y(\omega)$.

For these data, $\log\{\lambda_x(\omega)/\lambda_y(\omega)\}$ apparently has a strong quadratic component. We would therefore expect the semiparametric procedure to perform well, and it is

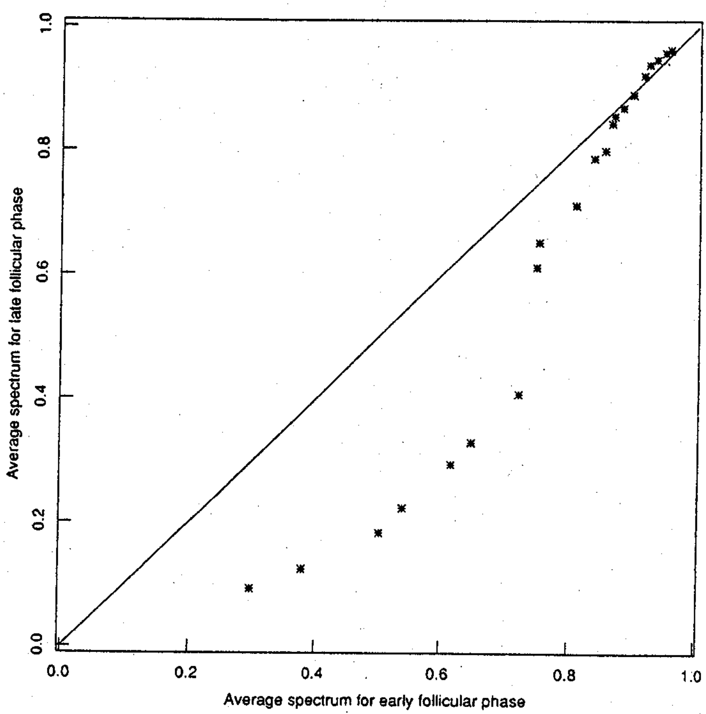

Fig. 3. CP plot for a comparison between early follicular and late follicular LH series

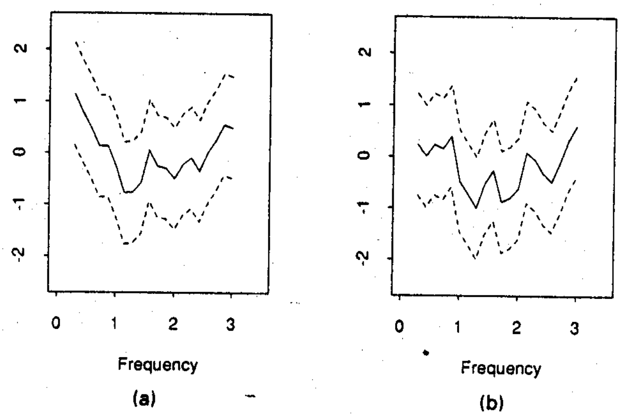

Fig. 4. Smooth estimate of $\log\{\lambda_x(\omega)/\lambda_y(\omega)\}$ for a comparison between early follicular and late follicular LH series, and approximate 90% pointwise confidence limits: (a) raw data; (b) detrended data

encouraging that the nonparametric approach gives essentially the same result. In other situations, such as cases (c) or (d) in the power study of Section 3, the extra generality of a completely nonparametric procedure can pay dividends.

In our analysis, we have not detrended the data before calculating the periodograms. This implies that any apparent trends in the data are ascribed to low frequency, but stationary, random variations in LH concentrations. An alternative interpretation might be that trends arise through natural circadian rhythms. Under this interpretation an appropriate form of detrending would be to subtract from each series of LH concentrations a fitted harmonic regression of the form

$$x_t = \alpha_1 \cos(t\omega) + \beta_1 \sin(t\omega) + \alpha_2 \cos(2t\omega) + \beta_2 \sin(2t\omega),$$

where the frequency ω corresponds to a 24-h cycle, and the parameters α_2 and β_2 are included to accommodate possibly non-sinusoidal circadian rhythms. The estimate of $\log\{\lambda_x(\omega)/\lambda_y(\omega)\}$ based on the detrended data is shown in Fig. 4(b). This is qualitatively similar to the estimate shown in Fig. 4(a), but the magnitude of the variation in $\log\{\lambda_x(\omega)/\lambda_y(\omega)\}$ over the frequency range is reduced, and the test based on D_m now gives a p-value of 0.144. We conclude that much of the evidence for apparent differences in spectral shape between the early and late follicular phases of the cycle derives from differences in patterns of circadian variation.

Acknowledgement

It is a pleasure to record the hospitality and facilities provided by JR and JTB during the initial work on this paper.

References

Bartlett, M. S. (1954) Problèmes de l'analyse spectrale des series temporelles stationnaires. *Publ. Inst. Statist. Univ. Paris*, **3**, 119–134.

Coates, D. S. and Diggle, P. J. (1986) Tests for comparing two estimated spectral densities. *J. Time Ser. Anal.*, **7**, 7–20.

Diggle, P. J. (1985) Comparing estimated spectral densities using GLIM. In *Generalized Linear Models: Proceedings Lancaster 1985* (eds R. Gilchrist, B. Francis and J. Whittaker), pp. 35–43. Berlin: Springer.

Doksum, K. A. and Sievers, G. L. (1976) Plotting with confidence: graphical comparison of two populations. *Biometrika*, **63**, 421–434.

Jenkins, G. M. and Watts, D. G. (1968) *Spectral Analysis and its Applications*. San Francisco: Holden-Day.

Kiefer, J. (1959) *K*-sample analogues of the Kolmogorov–Smirnov and Cramér–von Mises tests. *Ann. Math. Statist.*, **30**, 420–447.

Marriott, F. H. C. (1979) Barnard's Monte Carlo tests: how many simulations? *Appl. Statist.*, **28**, 75–77.

McCullagh, P. and Nelder, J. A. (1983) *Generalized Linear Models*. London: Chapman and Hall.

Pettitt, A. N. (1976) A two-sample Anderson–Darling rank statistic. *Biometrika*, **63**, 161–168.

Priestley, M. B. (1981) *Spectral Analysis and Time Series*. London: Academic Press.

Quade, D. (1973) The pair chart. *Statist. Neerland.*, **27**, 29–45.

Switzer, P. (1976) Confidence procedures for two-sample problems. *Biometrika*, **63**, 13–25.

Wichmann, B. A. and Hill, I. D. (1982) An efficient and portable pseudo-random number generator. *Appl. Statist.*, **31**, 188–190.