International Council for the Exploration of the Sea

CM 1980/M:18
Anadromous and Catadromous Fish Committee

THE MODELLING OF TIME SERIES OF CATCHES OF SALMON

by

J A POPE
DAFS Marine Laboratory, Aberdeen, Scotland, UK

ABSTRACT

A model proposed by Ottestad to explain the variation in annual Norwegian bag-net catches of salmon is discussed. Alternative models are considered.

RÉSUMÉ

On discute un modèle proposé par Ottestad pour expliquer la variation des prises annuelles de saumon faites par les Norvégiens à la bâche-trainante. On considère d'autres modèles possibles.

INTRODUCTION

A deterministic mathematical model, consisting essentially of a number of periodic terms, has been proposed by Ottestad (1979a) to explain fluctuations in the annual bag-net catches of salmon in Norway. Ottestad used data for the period 1900 to 1962 to estimate the parameters of his model and checked its plausibility by extrapolating it to 1976 and comparing the extrapolated values with actual statistics for these 14 years. The apparent success of his model in fitting the data led Ottestad to claim that the main cause of the observed variation in annual salmon yield is natural fluctuations in population size.

The use of models to describe time series has for long occupied the attention of many statisticians more particularly since the pioneering work of Yule (1927). Certain series have received considerable attention, notable among these being Wolfer's series of sunspot numbers and the annual trappings of the Canadian lynx (*Lynx Canadensis* Kerr).

The aims of time series model fitting are generally to:

(a) provide a satisfactory description of the mechanism by which values of a series are generated;

(b) forecast future values,

both these aims clearly being of practical importance.

Despite this, successes in model fitting have unfortunately not been commensurate with the great amount of effort that has, over the last fifty years, been put into developing and applying techniques.

TYPES OF MODELS

Time series models may be broadly classified into the following types:

I	Deterministic	$Y_t = \mu + f(t)$
II	Deterministic and error	$Y_t = f(t) + \varepsilon_t$
III	Non-deterministic	$Y_t = \mu + \Sigma\alpha_j \varepsilon_{t-j}$
IV	Mixed	$Y_t = \mu + f(t) + \Sigma\alpha_j \varepsilon_{t-j}$

where Y_t is the value of the series at time t, f(t) is an explicitly defined mathematical function of t, μ is the long term average value of the series, ε_t, ε_{t-j} etc are independent random variates with zero mean and α_o, α_1 etc are constants. Each of these types may be sub-divided into further types. Thus, within the non-deterministic types, there are autoregressive (AR), moving average (MA) and mixed autoregressive moving average (ARMA) models. These models are described in detail in numerous text books. For a simple introductory account see Kendall (1973). Other references are Anderson (1971), Box and Jenkins (1976) and Granger and Newbold (1977).

The model proposed by Ottestad for Norwegian bag-net catches of salmon belongs to Type II above. He chose for f(t) a function composed of the sum of a number of periodic (sinusoidal) terms of the type $\sin(2\pi\omega t)$ and $\cos(2\pi\omega t)$. Normally when adopting this formulation it is necessary to estimate both the number of terms to be included in f(t) and the values of ω, the frequencies. This may be done by carrying out a periodogram analysis, a procedure which takes the periods of the terms which are to be included in f(t) as integral divisors of the available number of terms of the series being modelled, n say. That is, only terms of the form

$$a_j \cos(2\pi\, jt/n) + b_j \sin(2\pi\, jt/n)$$

where j = 1, 2 ...., n/2-1 (n even) or j = 1, 2 ...., (n-1)/2 (n odd) are included. The attractions of this approach are that the maximum number of terms to be included is automatically fixed and, with the reasonably mild assumption that the error terms are normally and independently distributed with constant variance, the quantities $n(a_j^2 + b_j^2)/2\,\sigma^2$ are distributed as independent, central or non-central χ^2s each with 2 degrees of freedom. The statistical significance of these terms may, therefore, be assessed by means of a straightforward analysis of variance.

If f(t) is composed of terms

$$a(\lambda)\ \cos(2\pi t/\lambda) + b(\lambda)\ \sin(2\pi t/\lambda)$$

whose periods λ may lie anywhere in the interval $(2,\infty)$ and not merely at the discrete points n/j (j = 1, 2 ....) the number of terms is arbitrary and the advantages of the statistical independence of the χ^2s is also lost. Ottestad dealt with the first problem by postulating that the periods of the constituent terms of f(t) should be integral divisors of a fundamental period of length 4048 years, ie should be a subset of the values 4048/j (j = 1, 2 ..... 2024). Since the number of terms used in the series being modelled was 63, only values for which 4048/j < 63, ie for which j > 64, could be included. Ottestad chose a set of 13 values for j, namely j = 71, 96, 120, 176, 231, 296, 344, 358, 363, 378, 405, 426 and 476 for consideration. Ottestad's choice of 4 048 years for the fundamental period is based on his observation that the lengths of cycles found by him in an extensive study of 'biospheric' time series could be regarded as integral divisors of this value. The chosen set of 13 values for j is conjectural. His final model in fact only includes terms corresponding to 6 of these 13 values, the 6 chosen being selected by carrying out a step-wise regression of yield on 26 regressors, each value of j giving rise to both sine and cosine terms. Ottestad included an additional (27th) regressor, namely fishing effort as measured by the number of nets in use, this regressor also being significant. The actual significance levels, and, more particularly, the amount of variation explained by each regressor in turn are not quoted by Ottestad.

Ottestad thus chose to consider a model in which fishing effort appears as an additive rather than a multiplicative variable. A multiplicative model, which implies that the number of animals caught in a given time interval is proportional to fishing effort as well as to absolute population size, that is, that $Y_t = qe_tN_t$ and hence that $Y_t/e_t = qN_t$, is more usual in fish population dynamics.

ALTERNATIVE MODELS FOR SALMON YIELDS

Prior to any model fitting Ottestad smoothed his salmon catch data by taking three-point equally weighted moving averages. That is, his basic series (Y_t) was transformed according to $Z_t = (Y_{t-1} + Y_t + Y_{t+1})/3$. No formal justification for this choice of transformation is possible although smoothing of this sort has often been employed by others. It does introduce effects which should be taken into account in subsequent analysis.

A linear regression, carried out by the present author, of these smoothed values on the annual number of bag-nets employed showed a significant association between catch and effort but only some 12% of the total variability in catches was explained by the associated variation in fishing effort and, in the analyses presented in this section, fishing effort is not included.

The results of a periodogram analysis carried out on the Z_t values for 1901 to 1961 (n = 61) are given in Table I. This indicates that the components of highest intensity correspond to the longer periods. That is, most of the variability ('energy') is confined to the low frequency end of the spectrum. The analysis of variance in Table I shows that the remainder mean square has begun to level off after the first six periods (ie the six frequencies j/n, j = 1, 2 .... 6) have been fitted. These periods are from 10 to 61 years in length, those 'detected' by Ottestad having periods from 9½ to 57 years in length.

A similar analysis carried out on the untransformed series, ie on the Y_t values, gave essentially the same results (Table II). Some 77% of the total variation of the Y_t series is explained by the five longest periods.

Thus the periodograms of both the Y_t and Z_t series show the presence of large energy at low frequencies and not much energy elsewhere. Although the periodogram of a model consisting of sinusoidal components of discrete frequencies must be discontinuous by definition, a series which has a continuous spectrum but which is only observed at discrete points in time can also only be estimated at discrete frequencies. The periodograms given in Tables I and II and which are shown graphically in Figures 1 and 2 may therefore be regarded as estimating the spectra of time series which are either discrete or continuous. Models of type III have continuous spectra as distinct from those of type II which are essentially discrete. (The presence of a random error term whose variance is small compared with that of the series as a whole produces a continuous low level of intensity at all frequencies.) Different non-deterministic models have their own characteristic spectra, that of a simple first order autoregressive model (AR(1)) having the shape shown in Figure 3. The similarity between this and that of the discretely estimated spectra of Figures 1 and 2 is obvious.

AUTOCORRELATION ANALYSIS OF SALMON YIELDS

Autocorrelation analysis as a method of analysing time series was first introduced by Yule (1927) and subsequently developed by many others. The formal link between the autocorrelation and spectral functions of time series was first shown by Khintchine (1934). The basis of the method consists of calculating the correlation coefficients of the values of the series with themselves lagged by different amounts. Thus the kth order autocorrelation coefficient (r_k) is the ordinary correlation coefficient calculated from the (n-k) pairs of values (Y_{k+1}, Y_1), (Y_{k+2}, Y_2) ... (Y_n, Y_{n-k}). The correlogram is the graph if r_k against k. Different non-deterministic models exhibit different correlograms.

The correlograms up to lag 20 for the series Z_t and Y_t are given in Tables III and IV and both show good agreement with the theoretical autocorrelations of an AR(1) model, although neither correlogram dies out completely, both showing a tendency for slight oscillation at high lags. Hence, although both the periodogram analyses and the autocorrelation analyses strongly suggest that a suitable representation for the salmon yields, smoothed or unsmoothed, is given by an AR(1) process, it is perhaps wise at this stage not to regard this as the final model. The possibility that a model of type IV may be more appropriate is currently under investigation. For the present, however, an AR(1) model appears to give a very good approximation. Such a process has theoretical autocorrelations given by $\rho_k = \rho^k$, $-1 < \rho < 1$. The observed value of r_1 (which, for the Y_t series, is given by 0.7696) may be used as an estimator of ρ but, although a consistent estimator it is biased as shown by Marriott and Pope (1954). Correcting for bias gives an estimated value of ρ of about 0.80. The large sample standard error of this estimate is $\{(1-r^2)/n\}^{\frac{1}{2}} = 0.076$ and very approximate 95% confidence limits for ρ are 0.65-0.95. The theoretical correlogram of an AR(1) process with $\rho = 0.80$ is shown in Figure 4, along with the observed autocorrelations of the Y_t series. The estimated value of ρ is very close to unity and shows that the series is behaving rather like a random walk process where the value of the series at any time is equal to its previous value plus a random component.

FORECASTING SALMON YIELDS

Ottestad used his model to forecast salmon yields for each of the years 1963 to 1976. During this period the yields fell fairly steadily from a record high level of nearly 13000 tons to a value of less than 400 tons. Ottestad's model predicted this downward trend very successfully.

For an AR(1) process the predicted value h years ahead of a base year value Y_t is given by $Y_{t+h} = \mu + \rho^h (Y_t - \mu)$. Since $|\rho| < 1$, as h increases the predicted values steadily converge to the long term average of the series. For the unsmoothed series the average over the estimation period 1900 to 1962 is 850 tons. Clearly, therefore, predicted values many years beyond 1962 will greatly overestimate the actual yields, which, by 1975, had already fallen to roughly half the long term average. In reality, of course, one would normally forecast only one or two years ahead with an AR(1) process, continually moving the base year ahead as new data come to hand. Table V shows one year ahead predictions using both $\rho = 0.80$ and $\rho = 0.90$. Forecasts made using $\rho = 1.0$ would, of course, simply predict next year's value as being equal to this year's.

ACKNOWLEDGEMENT

I should like to express my very grateful thanks to Dr P Ottestad for so kindly providing me with the Norwegian salmon yield and bag-net statistics and also for drawing my attention to his most interesting paper on the analysis of biospheric time series (Ottestad, 1979b). Dr Ottestad has unfortunately not had time to comment on this paper prior to its submission.

The aim of this paper has not been to refute Dr Ottestad's most interesting theory but has been merely to show that alternative models may exist which equally successfully explain the behaviour of the salmon yield series. Further analyses of these data are clearly warranted and are at present in progress.

REFERENCES

Anderson, T.W. 1971 The Statistical Analysis of Time Series. J. Wiley, New York, xiv + 704 pp.

Box, G.E.P. and Jenkins, G.M. 1976 Time Series Analysis: Forecasting and Control. Holden Day, San Francisco, xxi + 575 pp.

Granger, C.W.J. and Newbold, P. 1977 Forecasting Economic Time Series. Academic Press, New York, xii + 333 pp.

Kendall, M.G. 1973 Time Series. Griffin, London, ix + 197 pp.

Khintchine, A. 1934 Korrelationstheorie der Stationaren Stochastischen Prozesse. Mathem. Ann. 109, 604-615.

Marriott, F.H.C. and Pope, J.A. 1954 Bias in the estimation of autocorrelations. Biometrika 41, 390-402.

Ottestad, P	1979a	On the fluctuations in the yield of the Norwegian salmon fishery. ICES CM 1979/M:5.
Ottestad, P	1979b	The sunspot series and biospheric series regarded as results due to a common cause. Scientific Reports of the Agricultural University of Norway, 58(9).
Yule, G.U.	1927	On the method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot series. Phil. Trans. Roy. Soc., Series A, 226, 267-298.

TABLE I PERIODGRAM ANALYSIS OF SALMON YIELDS FOR 1900-1962 SMOOTHED BY 3-PT MOVING AVERAGE (n = 61)

	Period (yr)	Intensity ($x10^3$)		Period (yr)	Intensity ($x10^3$)
1	61.0	320.6	16	3.8	5.6
2	30.5	518.5	17	3.6	3.0
3	20.3	245.1	18	3.4	1.1
4	15.2	193.1	19	3.2	1.3
5	12.2	127.5	20	3.0	1.4
6	10.2	55.8	21	2.9	1.1
7	8.7	10.2	22	2.8	1.4
8	7.6	23.8	23	2.6	0.9
9	6.8	31.8	24	2.5	4.1
10	6.1	10.4	25	2.44	0.3
11	5.6	20.8	26	2.35	1.0
12	5.1	9.5	27	2.26	4.8
13	4.7	1.0	28	2.18	1.9
14	4.4	8.7	29	2.10	0.9
15	4.1	0.8	30	2.03	2.7

Analysis of Variance:

	df.	m.s. ($x10^3$)		df.	m.s. ($x10^3$)
Period 1	2	160	Remainder	58	22
Period 1-2	4	210	Remainder	56	14
Period 1-3	6	181	Remainder	54	10
Period 1-4	8	160	Remainder	52	6
Period 1-5	10	140	Remainder	56	4
Period 1-6	12	122	Remainder	48	3
Period 1-7	14	105	Remainder	46	3
Period 1-20	40	40	Remainder	20	1

TABLE II PERIODOGRAM ANALYSIS OF SALMON YIELDS FOR 1900-1962 UNSMOOTHED DATA

	Period (yr)	Intensity ($x10^3$)		Period (yr)	Intensity ($x10^3$)
1	63.0	423.7	17	3.7	1.6
2	31.5	619.0	18	3.5	5.9
3	21.0	254.9	19	3.3	3.6
4	15.8	160.2	20	3.2	4.9
5	12.6	159.5	21	3.0	5.0
6	10.5	48.1	22	2.9	6.9
7	9.0	45.6	23	2.7	11.8
8	7.9	46.8	24	2.6	8.5
9	7.0	39.3	25	2.5	34.9
10	6.3	2.8	26	2.4	6.9
11	5.7	51.8	27	2.3	1.3
12	5.2	23.2	28	2.25	24.2
13	4.8	15.5	29	2.17	7.5
14	4.5	27.7	30	2.10	1.2
15	4.2	17.5	31	2.03	13.1
16	3.9	28.0			

Analysis of variance:

	df.	m.s. ($x10^3$)		df.	m.s. ($x10^3$)
Period 1	2	212	Remainder	60	28
Period 1-2	4	261	Remainder	58	18
Period 1-3	6	216	Remainder	56	14
Period 1-4	8	182	Remainder	54	12
Period 1-5	10	162	Remainder	52	9
Period 1-6	12	139	Remainder	50	9
Period 1-7	14	122	Remainder	48	8
Period 1-20	40	49	Remainder	22	6

TABLE III OBSERVED AUTOCORRELATIONS OF SMOOTHED NORWEGIAN SALMON YIELDS, 1900-1962 (n = 61)

Lag (k)	r_k	Lag (k)	r_k	Lag (k)	r_k	Lag (k)	r_k
1	0.9017	6	0.2190	11	0.0992	16	0.0580
2	0.7509	7	0.1271	12	0.1150	17	0.0739
3	0.5860	8	0.0654	13	0.1032	18	0.1035
4	0.4545	9	0.0474	14	0.0728	19	0.1292
5	0.3375	10	0.0653	15	0.0534	20	0.1572

TABLE IV OBSERVED AUTOCORRELATIONS OF UNSMOOTHED NORWEGIAN SALMON YIELDS, 1900-1962 (n = 63)

Lag (k)	r_k	Lag (k)	r_k	Lag (k)	r_k	Lag (k)	r_k
1	0.7696	6	0.2088	11	0.0812	16	0.0257
2	0.6323	7	0.15100	12	0.0938	17	0.0036
3	0.4696	8	0.1020	13	0.0971	18	0.0571
4	0.3826	9	0.0573	14	0.0421	19	0.0352
5	0.3527	10	0.0760	15	0.0298	20	0.0974

TABLE V ONE-YEAR-AHEAD FORECASTS OF SALMON YIELDS FROM AR(1) MODEL USING $\rho = 0.80$ AND $\rho = 0.90$. BASE YEAR: 1962

Year	Actual (tons)	Forecast $\rho = 0.80$	Forecast $\rho = 0.90$	Year	Actual (tons)	Forecast $\rho = 0.80$	Forecast $\rho = 0.90$
1963	1270	1194	1237	1970	644	828	826
1964	1410	1186	1228	1971	538	635	665
1965	1221	1298	1354	1972	666	600	569
1966	1094	1146	1184	1973	626	703	684
1967	1095	1045	1070	1974	516	671	648
1968	857	1046	1070	1975	412	583	549
1969	823	856	856	1976	368	500	456

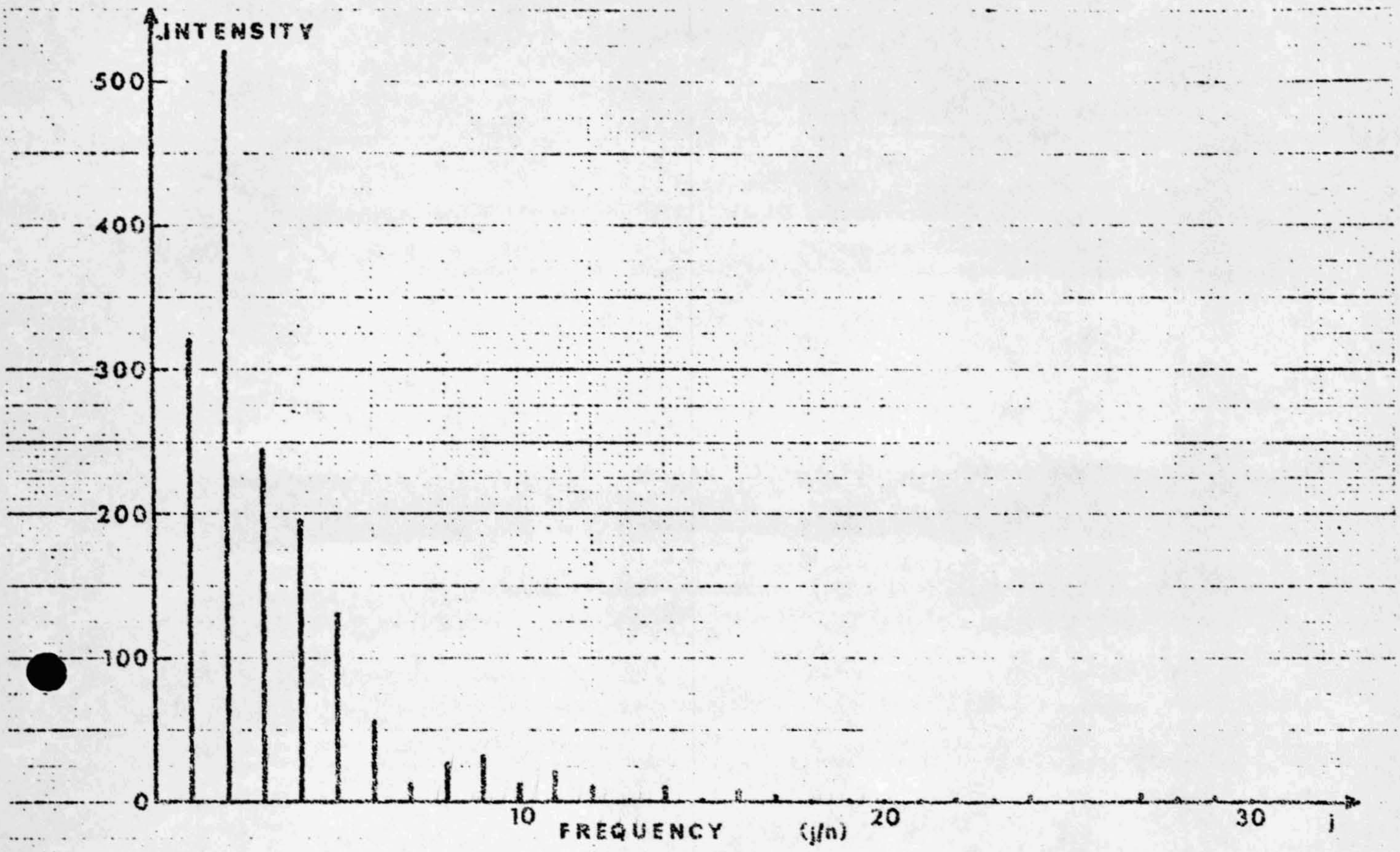

FIG. 1 Periodogram of Smoothed Salmon Series

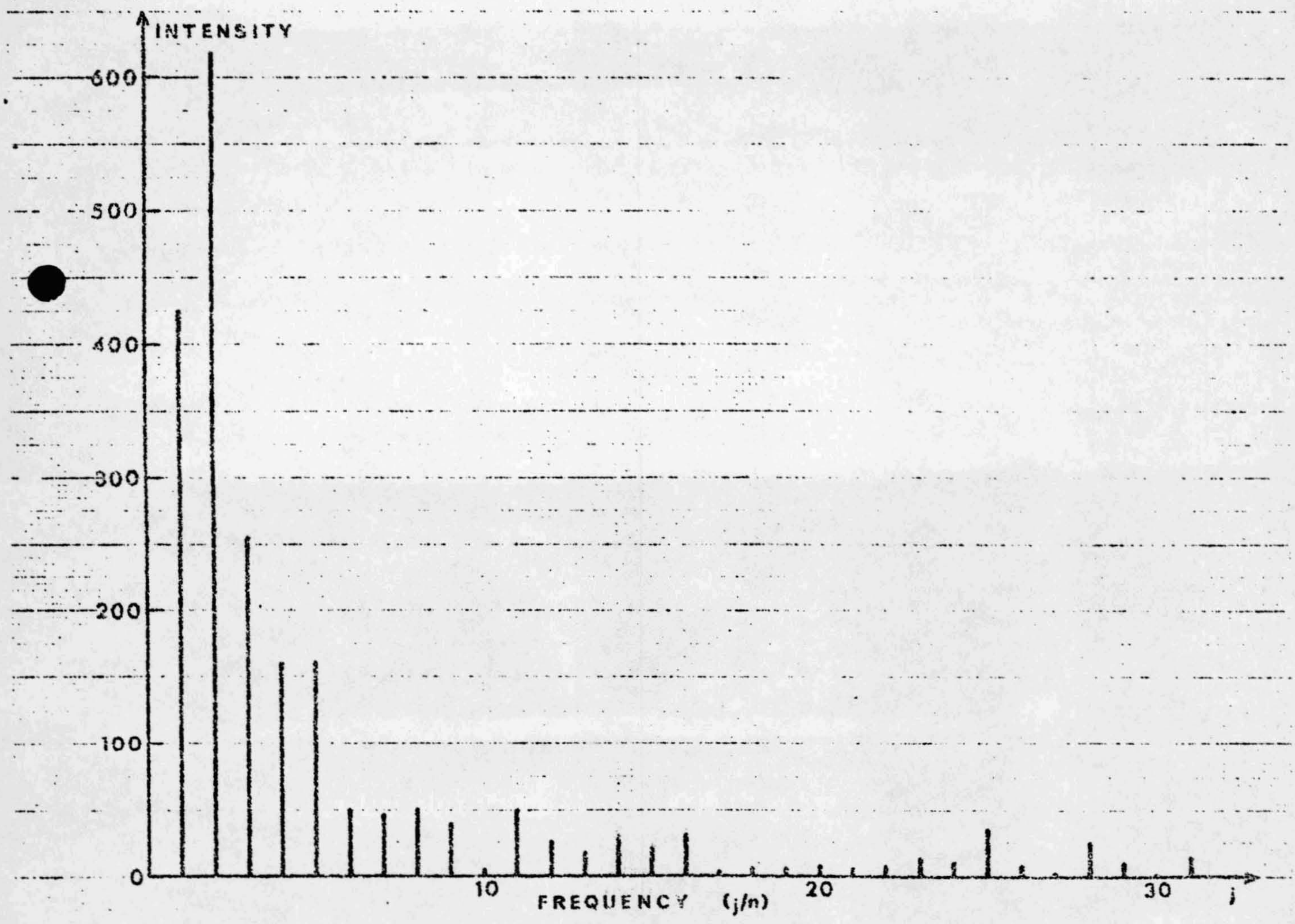

FIG.2 Periodogram of Unsmoothed Salmon Series

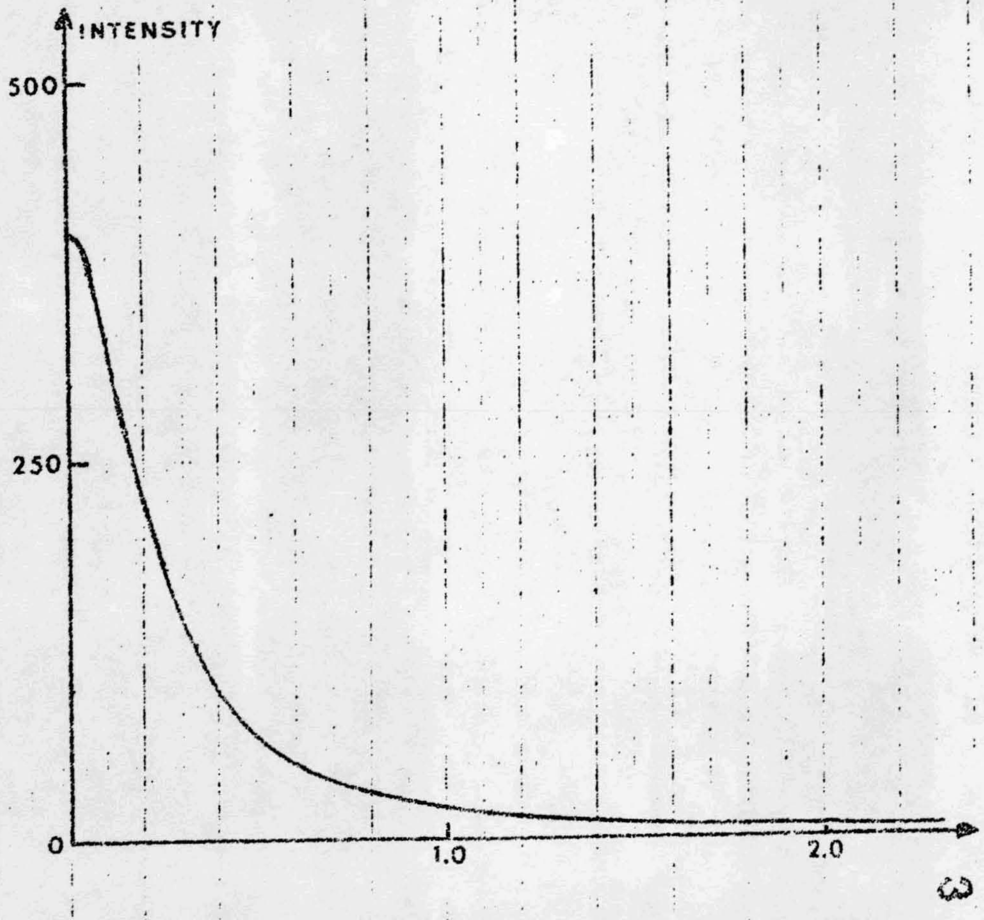

FIG.3 Theoretical Spectrum of AR(1) Process ($\rho=0.80, \sigma_\varepsilon^2=100$)

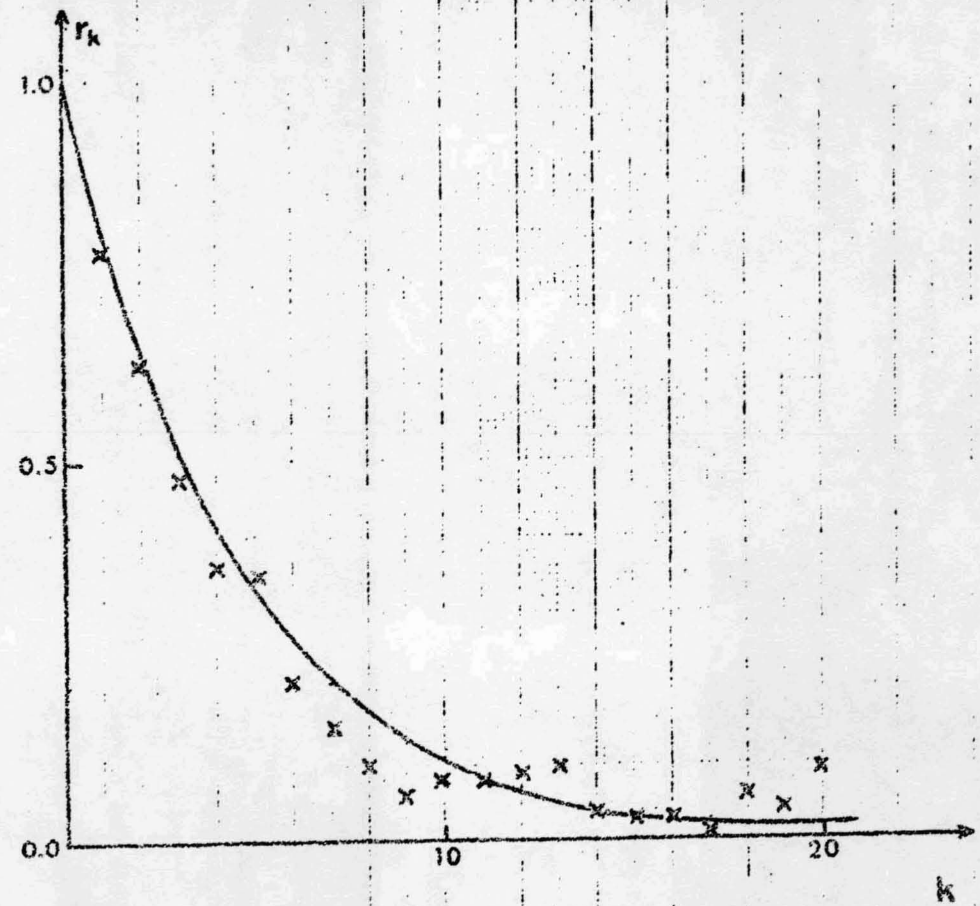

FIG.4 Observed and Theoretical Correlogram for Unsmoothed Salmon Series