Fluctuations of a Fortnightly Abundance Index of the Ivoirian Coastal Pelagic Species and Associated Environmental Conditions

Roy Mendelssohn

Pacific Environmental Group, Southwest Fisheries Center, National Marine Fisheries Service, P.O. Box 831, Monterey, CA 93942, USA

and Philippe Cury

CRODT, B.P. 2241, Dakar, Senegal

Mendelssohn, R., and P. Cury. 1987. Fluctuations of a fortnightly abundance index of the Ivoirian coastal pelagic species and associated environmental conditions. Can. J. Fish. Aquat. Sci. 44: 408–421.

In this paper we analyze time series of catch per unit of effort (CPUE) from 1966 to 1982 of small pelagic species off the Ivory Coast using sea surface temperature (SST) collected by merchant ships. A fill-in model is used to estimate missing values of CPUE and SST in the areas in which the fishery operates. A multivariate time series model of the fortnightly data is able to explain 43% of the observed variance in CPUE from 1966 to 1982. A model estimated by using only the data from 1966 to 1980 produced reasonable forecasts of the fortnightly CPUE for 1981–82. A new approach for estimating optimal transformations of variables in the model is used to examine the form of the relationships between CPUE and its predictors. The biological interpretation of the estimated transformations is consistent with previous results on the dynamics of zooplankton in the same area.

Les auteurs analysent une série chronologique de prises par unité d'effort (PUE) de petites espèces pélagiques capturées de 1966 à 1982 au large de la Côte d'Ivoire en fonction des températures de l'eau en surface (TES) relevées par des navires marchands. Ils ont utilisé un modèle à blancs pour déterminer les valeurs manquantes des PUE et des TES pour les pêcheries exploitées. Un modèle des séries chronologiques à plusieurs variables aléatoires pour des données bimensuelles permet d'expliquer 43 % de la variance observée des PUE obtenues de 1966 à 1982. À l'aide d'un modèle généré par les données recueillies de 1966 à 1980, ils ont obtenu des prévisions acceptables des PUE bimensuelles pour 1981–82. Une nouvelle approche pour l'estimation des transformations optimales des variables du modèle est utilisée pour étudier la relation entre les PUE et leurs éléments extrapolés. L'interprétation biologique des transformations estimées est en accord avec les résultats antérieurs sur la dynamique du zooplancton dans la même région.

Received September 13, 1985
Accepted September 30, 1986
(J8413)

Reçu le 13 septembre 1985
Accepté le 30 septembre 1986

The purse-seine fishery for coastal pelagic species off the Ivory Coast annually lands around 20 000 metric tons; this makes the fishery of local socioeconomic importance. The fishery also is of scientific interest because coastal pelagic species are important worldwide and are believed to be significantly influenced by environmental factors (see for example Parrish et al. 1983). The fishery off the Ivory Coast occurs in a region with two upwelling seasons, a smaller one around January and a larger one later in the year, which are followed by periods of strong rainfall, which affects salinity levels in the environment of the fish. Understanding the fishery dynamics in this area and how the dynamics are influenced by the environment can increase our understanding of other pelagic fisheries of greater economic importance.

In this paper our primary goal is to model the relative impact of the environment on the availability of pelagic species off the Ivory Coast in Africa and the manner in which the environment affects the dynamics of catch per unit of effort (CPUE) for these species on time scales as short as 2 wk. Previous studies suggest that the availability and abundance of these stocks are strongly influenced by the environment. Marchal (1967) and FAO (1974) presented evidence that sea surface temperature (SST) and salinity are important influences on the intrayear fluctuations in abundance of these stocks. Annual fluctuations in the abundance indices for the main Ivoirian or Ivoiro–Ghanaian stocks also have been related to changes in upwelling and rainfall (ORSTOM 1976; Binet 1982; Cury and Roy 1985). However, short-term fluctuations of these stocks never have been considered before.

Our approach to this problem is to use multivariate autoregressive moving-average (ARMA) models as described in Tiao and Box (1981) to model and forecast an index of relative abundance for the pelagic species off the Ivory Coast. We also use a new technique developed by Breiman and Friedman (1985) to analyze more closely the form of the relationship between CPUE and the relevant environmental variables. This technique allows us to examine the nonlinearities and discontinuities in the data that a priori are believed to exist but whose

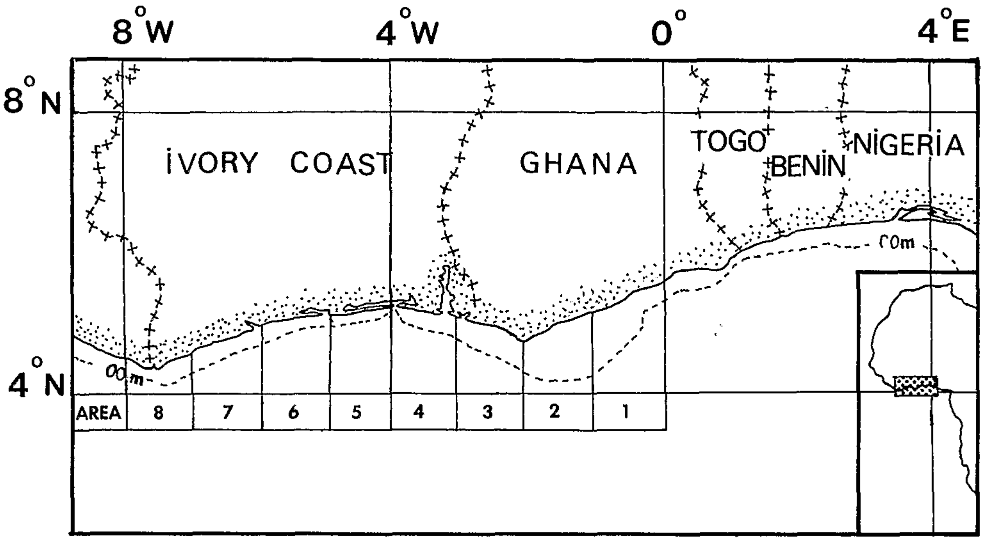

FIG. 1. Location of the areas studied off the Ivory Coast.

exact forms are not known. This should increase our theoretical understanding of how the environment affects the fish stocks as well as have important implications for modeling of fish abundance.

The ARMA class of models are stochastic rather than deterministic and have been of increasing interest in both applied and theoretical aspects of economics and engineering (among other areas). They emphasize the variation in the data and use empirically based methods to determine the proper form of the model. There have only been a few attempted applications of these techniques to fisheries problems (Boudreault et al. 1977; Saila et al. 1980; Mendelssohn 1981).

In contrast with our stochastic models, the majority of fishery management models developed over the last decades have been deterministic models concerned with estimating characteristics of the population at equilibrium (see for example Schaefer 1954; Beverton and Holt 1957; Ricker 1958; Pella and Tomlinson 1969; Fox 1970). Typically, these deterministic models operate on a yearly time scale, and in the best cases provide information related to optimal management of a fishery (such as maximum sustained yield (MSY)) and information necessary to determine policies to achieve this maximum, such as yield per recruit, age of first capture, and other characteristics of the exploited stocks.

Stochastic models of fisheries can be used for forecasting, even if only for short-term forecasts such as a fortnight or a month ahead, and remedy some of the deficiencies of deterministic models. They can be used to improve planning by fishermen and buyers, including both artisanal processors and industrial firms. Many of the exogenous changes that affect a fishery, such as price, or more to the point of this paper, the fluctuations in the physical environment, occur at time scales much shorter than a year. Such stochastic models can improve fishery management by defining the expected level of variation in a fishery and how much of the variation is due to environmental influences.

Biological Data

The fishery exploiting the pelagic species off the Ivory Coast consists mainly of purse seiners that are based in Abidjan. The fishery began in 1955 and expanded rapidly after that. Dependable statistics for the fishery have been available only since 1966 (FAO 1974). The fishery operates mainly over the rather wide continental shelf that exists off the Ivory Coast and Ghana. We have divided the area off the Ivory Coast and Ghana into eight zones, each 1° of longitude and extending from latitude 4°N to the coastline (Fig. 1). Areas 4−7 are the major fishing zones; catch in area 8 is typically less than 1% of the total annual catch, and the areas 1−3 off the coast of Ghana have only been exploited by the Ivoirian purse seiners at irregular intervals. The catch in these areas is dominated by *Sardinella aurita*, which is not the dominant species in the Ivoirian fishery.

During the period 1966−82, *Sardinella maderensis* was the dominant species caught in areas 4−7, comprising 54% of the total catch (Table 1; Fig. 2). *Brachydeuterus auritus* comprised 16.5% of the total catch during the same period, and various other species (Engraulidae, Scombridae, and Carangidae) comprised 14.6% of the catch. All of these species appear to be mainly Ivoirian stocks, as the catch of these species off the coast of Ghana appears to be coming from different stocks (FAO 1974). *Sardinella aurita*, which comprised 12% of the total catch, and *Scomber japonicus*, which comprised 2.4% of the total catch, appear to be shared stocks that are found off the coast of both countries (FAO 1974). *Sardinella aurita* and *S. japonicus* essentially disappeared from the catches in 1973, probably due to a combination of overfishing the previous year and due to anomalous climatic events in 1973. Binet (1982) has discussed this for *S. aurita*. The reasons for the disappearance of *S. japonicus* has not been conclusively determined. *Sardinella aurita* reappeared in 1976 and its relative abundance increased sharply off the Ivory Coast during 1981 and 1982.

The raw data consist of daily records for each boat and for each trip including the areas fished, the areas visited, and the catch. Search time was calculated for each area and fortnight using the method described in Fonteneau and Marchal (1970) and was used as a measure of fishing effort. CPUE was calculated by dividing the total catch of all species in an area during a fortnight by the corresponding search time for that area. When there was no fishing effort the CPUE was assumed to be a missing data point, as the lack of fishing does not necessarily imply the lack of fish. CPUE for the entire area was calculated as the mean of the CPUE of the four zones. CPUE thus is

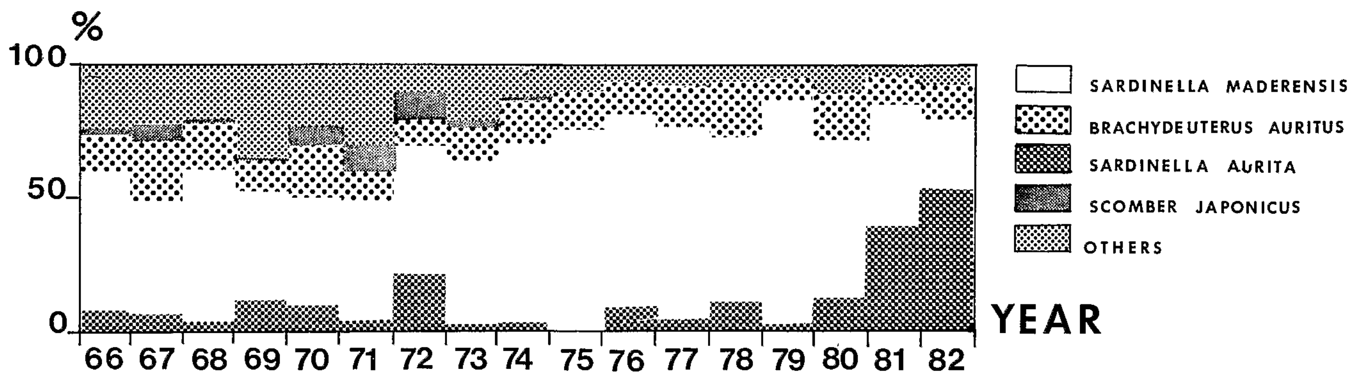

FIG. 2. Specific composition of the total catch of the Ivorian purse seiners in the main fishing areas (zones 4–7).

TABLE 1. Total catch (metric tons) per species for areas 4–7 from 1966 to 1982.

	S. maderensis	*B. auritus*	*S. aurita*	*S. japonicus*	Others	Total
1966	9 305	3662	1 783	262	4697	19 708
1967	8 944	6678	1 886	1405	5389	24 302
1968	12 150	5170	1 165	296	4611	23 391
1969	6 288	2745	2 301	100	5838	17 271
1970	5 794	3477	1 726	1195	3553	15 746
1971	8 629	2718	1 112	2131	5870	20 459
1972	10 742	3000	5 553	2485	2029	23 808
1973	5 368	1275	179	217	1775	8 815
1974	6 823	1850	268	61	1261	10 263
1975	7 805	1639	10	0	956	10 410
1976	10 974	2000	1 457	11	866	15 310
1977	14 898	3413	855	0	1398	20 563
1978	10 325	3568	1 892	6	923	16 716
1979	14 709	1208	439	0	912	17 268
1980	8 196	2572	1 789	0	1616	14 171
1981	9 790	2418	8 747	4	836	21 796
1982	5 108	2463	10 637	0	1608	19 815

defined in units of metric tons per day of fishing.

For CPUE to reflect the abundance of a fish that schools and whose schooling behavior changes rapidly, the measure used should reflect the number of schools present and the size and spatial distributions of both the fleet and the stock. Ignoring fishing, the boats can be viewed as taking a sample during each time period to provide an estimate of the abundance of the fish stock. From sampling theory we know that if there are too few boats out "sampling" (fishing) or if their spatial distribution is not appropriate to the target population, the estimates of abundance will be very inaccurate. Similarly, sampling needs to reflect the relative frequency of different school sizes if abundance estimates are to be accurate.

The measure of effort we use considers total search time, but does not consider either the number of boats searching or school size distribution. Thus, our measure will not always reflect abundance, particularly when there have been only a few boats out searching. More boats searching should also produce a less variable estimate of CPUE due to a more uniform coverage of the fishing area. This may explain the increased variability in the CPUE series after 1973, when the number of boats fishing decreased from 40 to fewer than 20.

Environmental Data

The environmental series considered in preliminary model identification consist of SST data collected by merchant ships passing through the area. The SST data were obtained from the National Climate Center (U.S.), and values were checked against a long-term atlas and further verified by C. Roy of ORSTOM. Fortnightly mean values then were calculated for the SST series by area, to agree with the fishery data, and for the two coastal environmental series.

The continental shelf off the Ivory Coast is influenced by two periods of upwelling. The main upwelling season occurs from June to September and a second, smaller upwelling occurs in January and February (Morliere 1970). The physical mechanisms underlying these upwellings are summarized and discussed in Picaut (1983, 1985). One of the reasons SST was selected as an environmental series is that it should be a reasonable surrogate variable for the occurrence of upwelling in the region.

Preliminary Data Analysis

Preliminary analysis of the data is necessary for two reasons. We consider areas that have no fishing as having missing data points and there are also areas with missing environmental data; most of our techniques, however, require complete data series. Therefore we must fill in the missing data in a manner that will not distort the underlying relationships in the data.

Also, the ARMA approach to model building is empirical, using properties of the covariance structure of the data to determine the appropriate form of the model to be estimated, rather than hypothesizing a model a prior. The form of the estimated lagged covariance matrices and generalized partial-correlation matrices are used to make initial estimates of the appropriate lags and variables to be included in the model.

The first step in the preliminary analysis is to fill in the missing data by area. Areas 4–7 have roughly 4% of the CPUE data missing. Also, during the years 1969 and 1982 there was a large gap in the SST data in all the areas. The CPUE data were filled in using an algorithm by Shumway and Stoffer (1982), used successfully in a related analysis by Mendelssohn and Roy (1986). The algorithm is described in the Appendix.

The basic idea of the algorithm is to iteratively fit a lagged model to the completed series and then to calculate minimum mean-square estimates of the missing data given the new parameter estimates. The algorithm has good statistical properties (see Shumway and Stoffer 1982), particularly in terms of preserving the covariance structure of the observed series. Other more obvious methods do not share this property. Also, the algorithm includes observation error in calculating both the parameter estimates and the smoothed estimates of the missing data.

The algorithm calculates smoothed values rather than predicted values of the missing data. Predicted values use only information previous (or perhaps contemporaneous) to the missing data point to estimate the fill-in. Smoothed values use all the information available consistent with the model. Thus, for an AR(2) model, the filled in value is calculated from a forward model using data from the last two periods, from a backward data using the next two periods of data, and from a regression-like model using contemporaneous innovation series.

The fill-in model for CPUE uses the observed data from each area and the model form to fill in any missing point. While this may cause some bias in our later analysis, it is a conservative procedure for several reasons. First, as one of our main interests is the role of the environment in the population dynamics, any bias introduced should be to lessen the importance of the environment as a predictor and to increase the importance of lagged values of CPUE as predictors. Also, we use average CPUE (across areas) in our final model. Thus most of the filled in values are averaged with real values from other areas, lessening the impact on the data of any of the filled in values. With only 4% of the data missing, a filled in data point would have to have significant leverage to greatly influence the final model.

The SST data were filled in using both the Shumway and Stoffer algorithm and the estimated seasonal cycle. The iterative algorithm produced data values for the missing blocks in 1969 and 1982 that did not appear reasonable; however, the spectra produced by this procedure were very consistent with those calculated from complete subseries of the SST data. Parameter estimates depend on the covariances of the data, so that the bad data values may not be important. Using the estimated seasonal cycle can produce its own distortions, as the missing year may have been very unlike the normal year. This can greatly emphasize the seasonal frequency in the data while lessening other frequencies that might be of importance in determining the appropriate model.

Fortunately, our results were almost identical for each method of filling in the SST data. Preliminary reviewers of our results felt more comfortable with using the estimated seasonal cycle; therefore, in what follows, the SST data have been completed using the estimated seasonal cycle.

The smoothed values for CPUE by area (Fig. 3) appear to be quite reasonable. None of the estimated values appear as particular outliers, which reinforces our belief that this should produce little bias in our final calculations.

We do not use effort at time t (i.e. no lag) as a predictor because this would introduce a large bias and a high degree of spurious correlation into the analysis. This point is shown mathematically in Eberhardt (1970); Mendelssohn (1981) gives other references that examine the degree of spurious correlation introduced.

An intuitive explanation of the source of this spurious correlation can be seen by considering log (CPUE) which equals log (catch) − log (effort). If this is regressed against a function of effort, then effort is clearly on both sides of the equation. If the coefficient of variation (CV) of catch is small compared with that of effort, then catch can be viewed as a constant, and we would be regressing a function of effort (the logarithm of effort) against another function of effort, so that some degree of fit is assured. Simulations in the references suggest that the r^2 value can be inflated by as much as 0.4–0.6 depending on the relative values of the CV of catch compared with the CV of effort.

The accuracy of the completed time series of CPUE and SST (Fig. 3 and 4) can be assessed in several ways: by comparing smoothed values with observed values when both exist; and by comparing smoothed values with observed values in other series where the data exist to check that the smoothed value is consistent. For the CPUE series, which are relatively complete, the smoothed values can be seen to be very close to the observed values, even at large peaks in CPUE, such as in area 4 during 1976 (Fig. 3a).

The parameter estimates produced by the AR(2) fill-in model for CPUE (Table 2) are of interest by themselves, as they appear to agree with and clarify previous work on migration in the stocks. The interpretation of the model for CPUE will be discussed in detail in a companion paper on the spatiotemporal dynamics of the fish. In what follows, we restrict our attention to CPUE and SST averaged over all of the areas.

Model Identification and Estimation

We model the temporal dynamics of CPUE and its possible relationships with SST by using for the most part the ARMA approach described in Tiao and Box (1981) and implemented in the computer program WMTS-1 (Tiao et al. 1980). At times, a method due to Akaike et al. (1979) for subset multivariate autoregression identification was also used to help select the appropriate lags and the parameters to be included in the model at those lags.

For several reasons, we analyze the natural logarithm of CPUE (actually of CPUE + 0.05 so that zero values are well defined) rather than analysing CPUE itself. First, CPUE is a ratio of two separate series, and ratios rarely are linear or gaussian in distribution. Taking logarithms of the data produces a new series that is the difference of the logarithms of the two original series, which should have better statistical properties. Second, inspection of the raw data (Fig. 4a) suggests that the variance of the series changes with the level of the CPUE series. A range–mean plot (Jenkins 1979, p. 96) of the CPUE data was calculated. This is found by dividing the series into a number of subseries (in this case years), calculating the range and mean of each subseries, and plotting the range versus the mean. This plot showed that a transformation such as a log

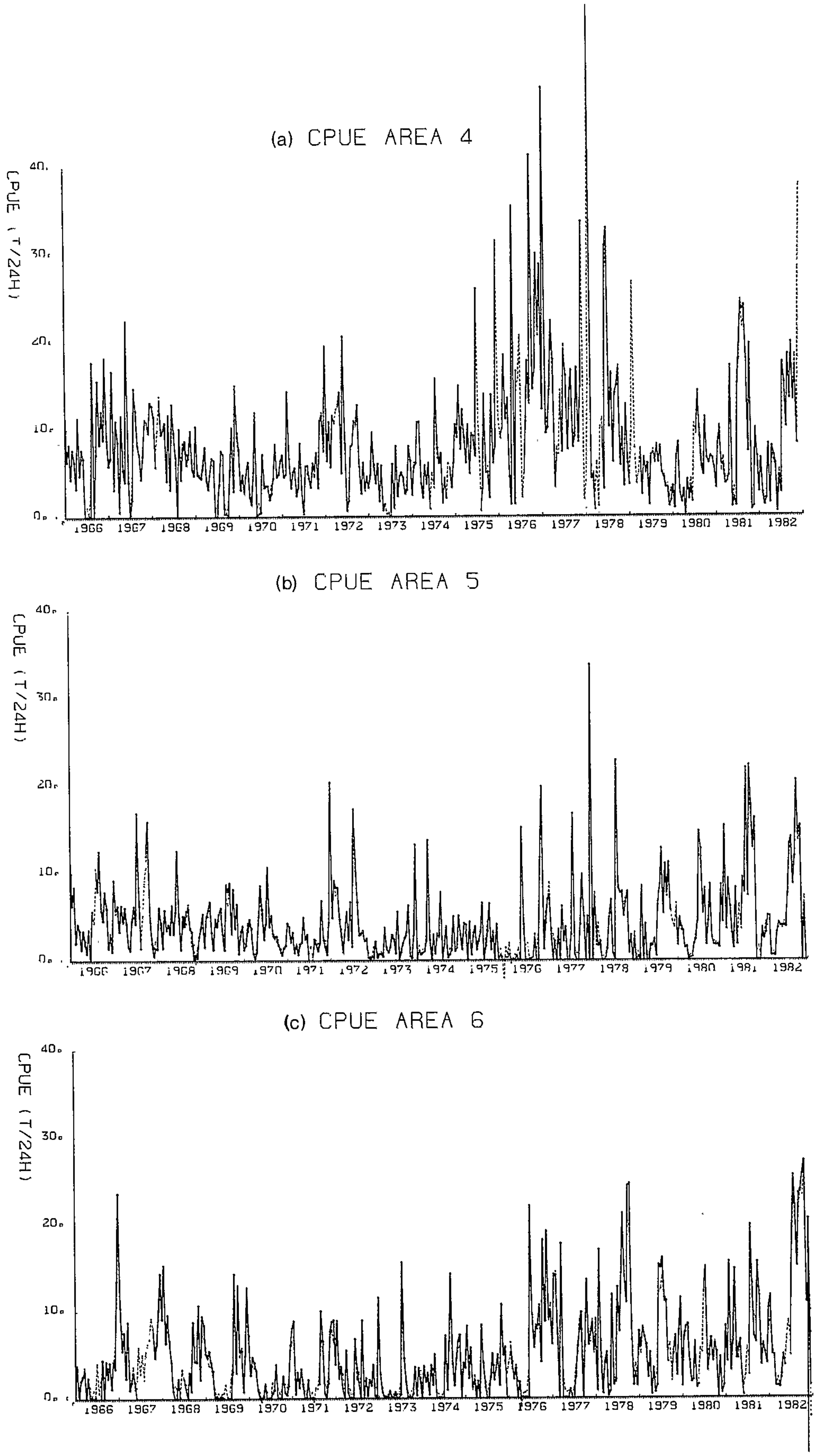

FIG. 3. Time series of CPUE (all species) for (a–d) areas 4–7 and (e) all areas combined. Solid lines = observed; broken lines = smoothed. *(Fig. 3 concluded next page)*

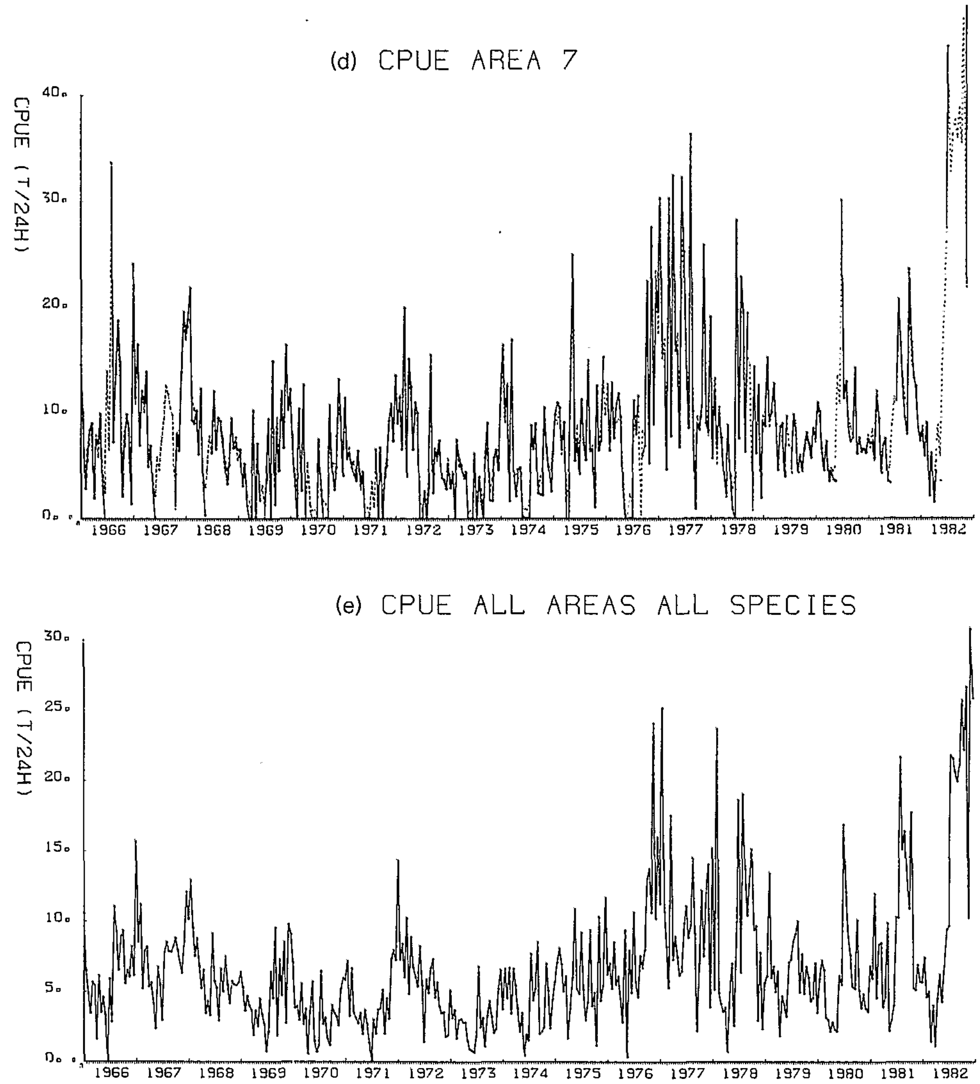

FIG. 3. (*Concluded*)

transform or a negative square root transformation was needed. Third, there are theoretical reasons for believing that a log scale is more appropriate for modeling the interaction of CPUE with SST. On the untransformed scale, a 1° change in SST would have to make the same absolute change in CPUE at very low levels of CPUE as with very high levels of CPUE. When ln (CPUE + 0.05) is used, a 1° change in SST produces a percentage change in the level of CPUE rather than an absolute change. Thus, at much smaller levels of CPUE, smaller absolute changes in the level will be observed. In the next section, we will examine in more detail the proper scales for each of the variables in our model and the form of the interactions between the different variables.

The first step of the analysis is to identify the appropriate lags to be included in the model and whether an AR, MA, or ARMA model is more appropriate. This is done by examining the cross-correlation matrices and by calculating the general partial-correlation matrices (see Tiao and Box 1981 for details). The CPUE autocorrelation function (not shown) has significant lags up to 8 fortnights, and again around 24 fortnights (1 yr), although the pattern is not that of a seasonal cycle in the data. (A seasonal cycle is reflected in a sinusoidal autocorrelation function. The autocorrelation function of SST is sinusoidal, reflecting a seasonal cycle, while CPUE only has a peak at a lag of 24 fortnights. To call this a "cycle" would be equivalent to saying a peak at lag 1 shows a one-fortnight cycle.) The pattern in lags 1−8 suggests significant values at lags 1, 2, and 4 and possibly at lag 6. This is because autoregressive terms decay exponentially. Thus, if there is a significant autocorrelation at lag 1, then lag 2 will have an autocorrelation of that value squared, even if there is no lag 2 effect. Similarly, if there is a lag 2 term, then lag 4 will have that value squared even if there is no lag 4 effect. Thus, we look for the autocorrelations that stand out from the exponential decay of the previous significant lags.

The cross-correlation function (not shown) shows a significant peak and exponential decline with SST lagged 17 fortnights and significant cross-correlations with SST lagged until about 8−10 fortnights. The generalized partial-correlations (Table 3) exhibit "spiked" behavior consistent with an AR-type model at the lags suggested by the cross-correlation matrices.

An exact maximum likelihood algorithm was used to estimate the parameters of this model, which also included parameters modeling the SST series. This last step is important so that the final vector residual series are as close as possible to independent multivariate gaussian variables. Otherwise, the esti-

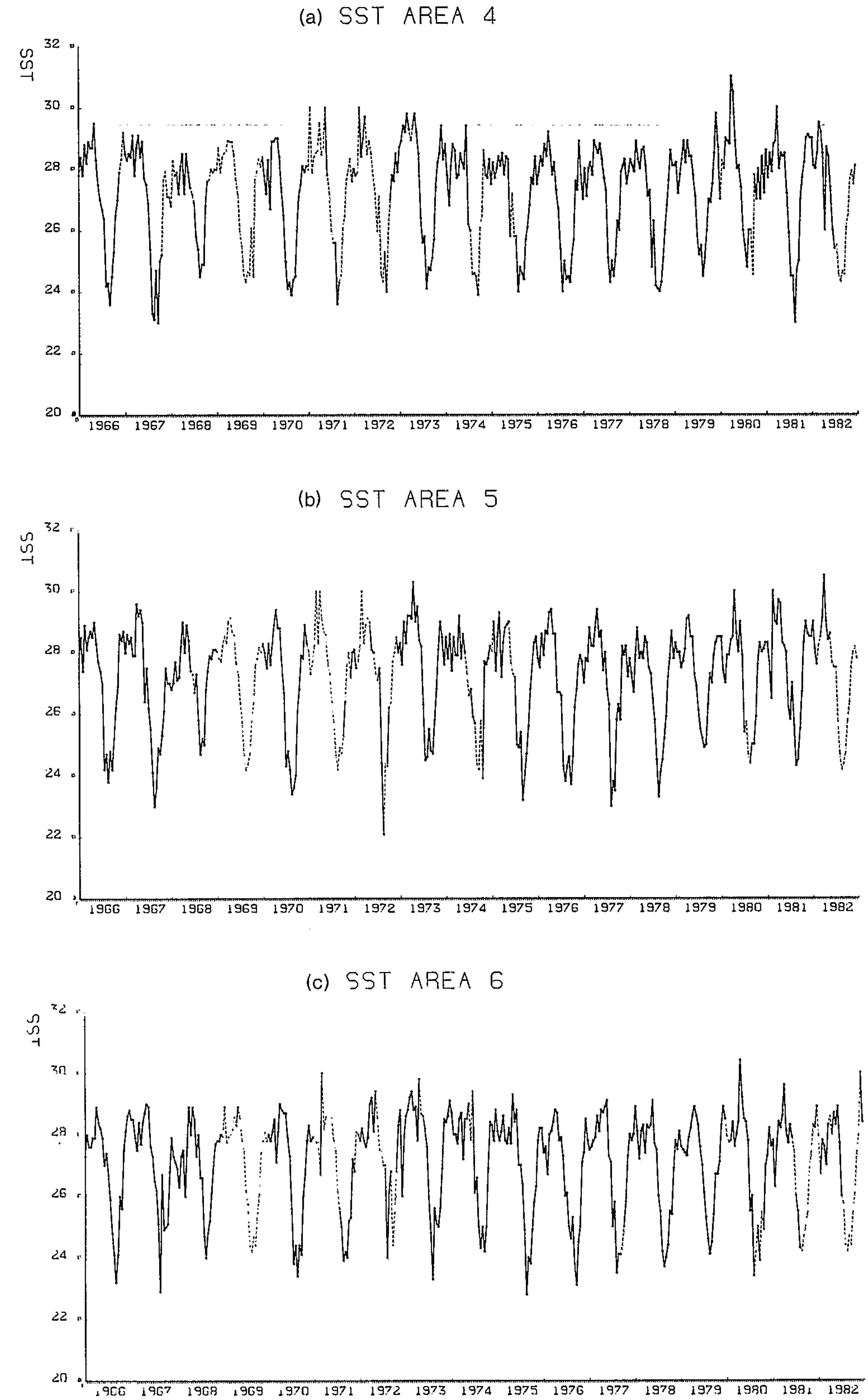

FIG. 4. Time series of SST for (a–d) areas 4–7 and (e) average of all areas. Solid lines = observed; broken lines = smoothed. (*Fig. 4 concluded next page*)

mates can be biased. The relationships between CPUE and SST at lags 6 and 7 were found to be nonsignificant. They were discarded from the model, and the remaining parameters were reestimated.

The final parameter estimates and their standard errors (Table 4a) were tested for adequacy by several methods. The residual cross-correlation matrices do not suggest a lack of fit (Table 5). The residual series (not shown) have a zero mean and otherwise do not suggest a lack of fit to the data. The normalized residuals for CPUE (not shown) are nearly gaus-

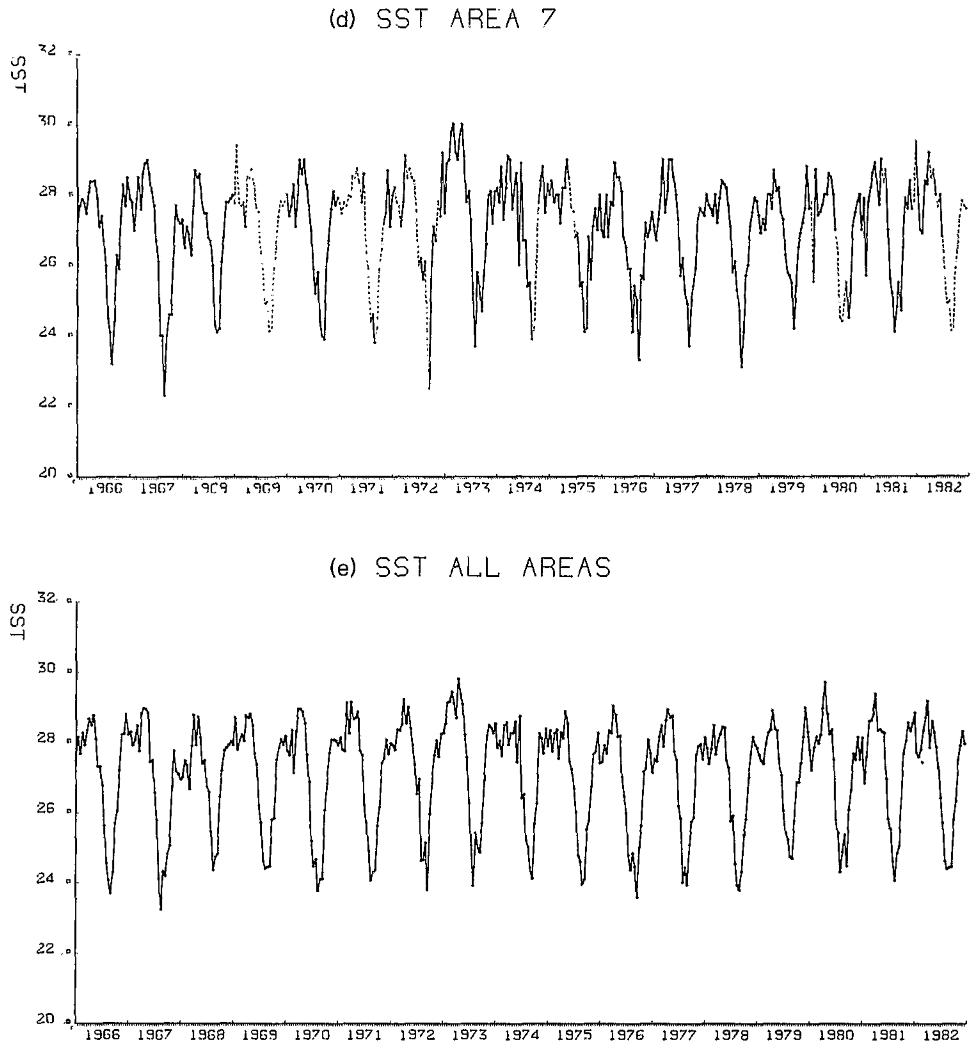

FIG. 4. (*Concluded*)

TABLE 2. Values of coefficients of the fill-in models for CPUE time series ($t-1$, $t-2$) in areas 4–7.

	Area 3		Area 4		Area 5		Area 6		Area 7		Area 8	
	$t-1$	$t-2$	$t-1$	$t-2$	$t-1$	$t-2$	$t-1$	$t-2$	$t-1$	$t-2$	$t-1$	$t-2$
Area 4	0.11	0.02	0.24	0.20	0.20	0.10	−0.51	0.34	0.22	0.06	0.11	−0.05
Area 5	0.01	−0.02	−0.07	0.19	0.40	0.09	0.02	0.07	−0.04	−0.14	−0.01	−0.11
Area 6	0.07	−0.03	0.08	−0.01	0.04	−0.01	−0.38	0.18	−0.02	0.10	0.17	−0.14
Area 7	−0.01	−0.16	0.08	0.09	0.09	−0.05	−0.04	0.44	0.31	0.53	−0.10	−0.13

TABLE 3. Significance (5%) of partial autoregression coefficients by lag for time series of CPUE and SST.

	Lag									
	24	17	1	2	3	4	5	6	7	8
LNCPUE	+		+	+		+				
SST	−	+	−	+				−	−	

sian, with a mean of zero and a median close to zero. The distribution of the normalized residuals is fairly symmetric also. There are two noticeable outliers, each with a normalized value greater than 4. These two points account for almost 10% of the residual variance. Both are from fortnights when there was only one boat searching, and from the data it is quite clear that in each instance the boat was just passing through the area to fish in another area. This is consistent with our previous discussion of problems with the measure of effort used.

The model was also estimated using the fill-ins for SST produced by the Shumway and Stoffer algorithm (Table 4b). There is essentially no difference, giving us some confidence that our model is robust to the method of completing the data series. Both models were reestimated for the years 1966–82. In 1981 and 1982, there was a sharp increase in the level of CPUE, and also a change in the species composition, with an increasing amount of *S. aurita* in the catch. Even with these changes, there is essentially no change in the parameter values. Thus the estimated model appears to be robust to additional years of data, even to data that are seemingly different from the previous years.

The estimated model for CPUE from the years 1966–80 is

$$\begin{aligned}\text{LNCPUE}(t) = {} & 0.248\cdot\text{LNCPUE}(t-1)\\ & + 0.254\cdot\text{LNCPUE}(t-2)\\ & + 0.162\cdot\text{LNCPUE}(t-4)\\ & + 0.143\cdot\text{LNCPUE}(t-24)\\ & - 0.143\cdot\text{SST}(t-1)\\ & + 0.112\cdot\text{SST}(t-2)\\ & + 0.044\cdot\text{SST}(t-17)\end{aligned}$$

where $\text{LNCPUE}(t) = \text{LN}(\text{CPUE}(t) + 0.05)$ and t references the time period in fortnights. The model explains 43% of the observed variance in the data. Without the two outliers, the model explains almost 55% of the observed variance. Observed and one-step-ahead predicted values of CPUE transformed back to the original scale (Fig. 5) show that the predicted values are close to the observed values except at the large peaks in CPUE. Also, the predicted values are generally better before 1973 than after, which may be due to the problems in effort mentioned earlier. The model also frequently predicts the beginning and end of the fishing season (see for example 1968, 1974, 1979, and 1981).

A scatterplot of the observed CPUE versus the predicted CPUE (not shown) shows a trend to underpredict at high levels of CPUE, particularly at values greater than 12 tons per day. In the next section we will discuss some reasons for this observation.

TABLE 4. Final estimates and estimated standard error of the parameters from time series analysis from 1966 to 1982 and 1966 to 1980 using the (a) seasonal cycle and (b) iterative algorithm to fill-in SST series.

Parameter		1966–82		1966–80	
		Estimate	SE	Estimate	SE
(a) Seasonal cycle					
Lag 1	LNCPUE	+0.258	0.047	+0.224	0.051
	SST	−0.143	0.039	−0.122	0.043
Lag 2	LNCPUE	+0.254	0.051	+0.223	0.054
	SST	+0.112	0.038	+0.090	0.042
Lag 4	LNCPUE	+0.162	0.048	+0.171	0.051
	SST	—	—	—	—
Lag 17	LNCPUE	—	—	—	—
	SST	+0.029	0.013	+0.045	0.013
Lag 24	LNCPUE	+0.144	0.042	+0.131	0.044
	SST	—	—	—	—
(b) Iterative algorithum					
Lag 1	LNCPUE	+0.261	0.047	+0.234	0.051
	SST	−0.141	0.043	−0.117	0.046
Lag 2	LNCPUE	+0.257	0.051	+0.227	0.054
	SST	+0.114	0.042	+0.089	0.045
Lag 4	LNCPUE	+0.159	0.048	+0.169	0.051
	SST	—	—	—	—
Lag 17	LNCPUE	—	—	—	—
	SST	+0.038	0.013	+0.042	0.014
Lag 24	LNCPUE	+0.148	0.042	+0.130	0.045
	SST	—	—	—	—

Interpretation of the Model and Examination of the Form of the Relationship between CPUE and its Predictors

An empirically derived model is more satisfying if it has an interesting biological interpretation which is consistent with or combines previous observations on the fishery. The model estimated in the last section has such an interpretation. All else being equal, CPUE will show some persistence on its own; at low levels it will tend to remain low and higher levels will exponentially decline at a reasonable rate. If a "kick" comes that pushes CPUE up to a higher level, then the results of this "kick" will be felt for a few time periods to come. This may be due to schooling behavior. Conditions favorable for the formation of large schools tend to persist so that the schools themselves are available to the fishery for an extended period.

From the model estimates, what appears to provide the needed "kick" is a drop in SST from two fortnights ago to one fortnight ago. The sharper the drop in temperature and the longer the duration of the drop, the greater the effect on CPUE. If SST was relatively warm two fortnights ago and relatively warm one fortnight ago, there will not be much change in CPUE. Similarly, if SST was relatively cold two fortnights ago and relatively cold one fortnight ago, again there will not be much change. Once the increase is started, it will be maintained for a while by the persistence of CPUE noted above. Thus, it is not colder or warmer waters per se that are conducive to high levels of CPUE, but rather a long and sharp drop in temperature. This is of course the behavior of SST at the onset of the upwelling seasons. The model suggests that the beginnings and endings of fishing seasons should be related to upwelling and the success of the fishing season related to whether the season had strong or weak upwelling. Binet (1976) has presented results that show a correlation between a drop in SST and an increase in zooplankton biomass a fortnight later. Binet (1983) also has suggested that the colder waters tend to increase the aggregation of the zooplankton at the surface, and hence presumably to aggregate the fish as well. It is during the upwelling seasons with the onset of colder waters that we find higher levels of CPUE (Fig. 4e). As the fish species in our model feed on zooplankton (Dia 1972), this explanation of the physical mechanism underlying the model would appear to be consistent with independent biological observations.

E. Marchal (Antenne ORSTOM, Centre Oceanologique de Bretagne, Brest, pers. comm.) has suggested that CPUE lagged on itself at periods of 2, 4, 6, and 8 arises from the affect of the lunar cycle on these species noted in Marchal (1967). He also suggested that the residual correlation at lag 35 is approximately the correct lag at which the lunar and solar fortnights would be in phase. Thus, some of the long-term (four fortnights) persistence in CPUE may be due to this lunar effect.

So far we have restricted ourselves to linear models, with only a minor transformation in the CPUE data. There are a priori reasons to believe that the relationships between CPUE with itself lagged in time and between CPUE with SST at the various lags are nonlinear and that there may be possibly nonlinear transformations of the data that would improve the fit of the model while still maintaining an additive model which has better statistical properties. Plots of LNCPUE at time t versus SST at time t and $t - 1$ (not shown) suggest a discontinuity or threshold in the relationship around approximately 27°C. These plots do not correct for the effect of the other variables in the model.

TABLE 5. Residual cross-correlations from lag 1 to 36 ("+" denotes a significant value and "−" a nonsignificant value at 5%).

Lag	CPUE	SST
1	−	−
2	−	−
3	−	−
4	−	−
5	−	−
6	−	−
7	−	−
8	−	−
9	−	−
10	−	−
11	−	−
12	−	−
13	−	−
14	+	−
15	−	−
16	−	−
17	−	−
18	−	−
19	−	−
20	−	−
21	−	−
22	−	−
23	−	−
24	−	−
25	−	−
26	−	−
27	−	−
28	−	−
29	−	+
30	−	−
31	−	−
32	−	−
33	−	−
34	−	−
35	+	−
36	−	−

To find "good" transformations of the data, we used a technique developed by Breiman and Friedman (1985) that empirically calculates optimal transformation of the data (see Appendix for further descriptions). The results of the algorithm are a smoothed empirical transformation of each of the observed data points. The implied functional form can be found by plotting the transformed value of the variable versus the original value. The transformed values are in no particular units (unless a functional transformation can be discerned from the plot; then the transformed data would be in whatever units are implied by that particular functional transformation).

In our model, the response variable is CPUE at time t, and the predictor variables are CPUE at times $t-1$, $t-2$, $t-4$, and $t-24$ and SST at times $t-1$, $t-2$, and $t-17$. The resulting transformed model explains 59% of the observed variance in the original series. It is worth emphasizing that this is not necessarily the optimal transformed model. We used the lags and parameters identified in our linear model as the correct ones. It is possible that if the correct transformations had been known a priori, then other lags of either CPUE or SST would have been found to be more important as predictors of CPUE at time t. Also, slightly different results are obtained for CPUE depending on which set of fill-ins is used for SST.

The optimal transformations for CPUE and for CPUE lagged one fortnight (not shown) range between a linear to a log transformation; the most noticeable difference from a log transformation is that the curve is less flat at high values of CPUE (>12). The results suggest that a slightly less steep transformation might have been preferable.

The transformation for CPUE lagged four fortnights (Fig. 6a) shows a sharp discontinuity in the relationship. The transformation is almost piecewise linear with a point of discontinuity around 12 tons per 24 h of search. Beyond this point, the value increases sharply. It does not seem coincidental that this is the same value at which we find our model consistently underpredicting CPUE. Our best guess is that this result has to do with schooling behavior of the fish: first that such behavior may influence our measure of effort, and hence influence CPUE (discussed earlier in the paper), and second that once either large schools are formed or perhaps many schools are formed, the fish become much easier to find so that CPUE sharply increases. Why the sharpest discontinuity should be at a lag of four fortnights is not obvious, but is clearly a topic for further study.

We would expect that the relationship between CPUE and SST would be nonlinear, probably some type of threshold function. This is because there are temperatures too high for the *Sardinella*, so an increase in SST above some given temperature will not affect CPUE significantly, and similarly there is also probably a temperature too low for the species. The transformation for SST at a lag of one fortnight (Fig. 7a) decreases linearly to 28°C, increases to 28.5°C, and then again sharply decreases. Not surprisingly, the sharpest response is between 27 and 28°C. The reason for the change in sign at 28°C is unclear, but it is clear that the linear relationship of the original model will only roughly approximate this functional form. The transformation of SST at a lag of two fortnights (Fig. 7b) displays the expected threshold effect; it linearly increases to just short of 28°C and then levels off. Further increases in SST will not affect CPUE further. Marchal (1967) studied 1 yr of data and found 23−26°C to be the optimal temperature range and reported a sharp drop in CPUE between 28 and 29°C. Our transformation clearly supports this finding and shows that it is valid for more than the single year Marchal studied.

The transformation for SST lagged 17 fortnights (Fig. 7c) is not as easily interpreted. It is highly nonlinear, almost bimodal in form. E. Marchal (pers. comm.) has suggested that the relationship with SST 17 fortnights earlier possibly may be due to the effect on an earlier life stage of the fish.

The transformations show that our simple linear models will be adequate over some of the range of the observed data, but that there are sharp discontinuities that will cause systematic bias in the predictions. These results also show that we can find additive models with all the variables transformed to a proper scale that remove most of the difficulties. The nonlinear form of the transformations do not alter on the whole the physical interpretation of our model: indeed, except for the effect of CPUE lagged four fortnights, it would seem to reinforce it. The transformation of CPUE lagged four fortnights may reflect schooling behavior and changes in school size and distribution because of the sharp nonlinearity in the transformation for values greater than 12 tons per day of searching. More accurate forecasts could probably be developed using measures of CPUE that better reflect both the spatial and size distribution of the schools and the number of boats searching.

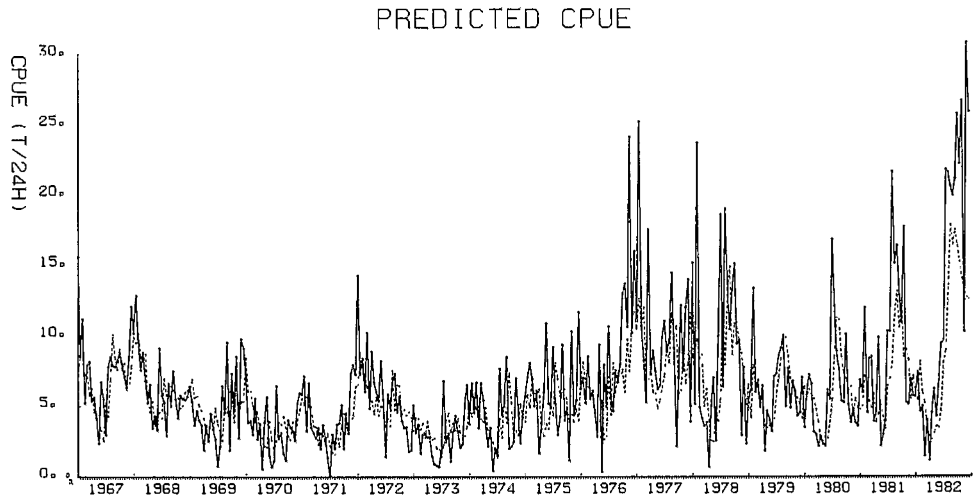

FIG. 5. Observed (solid line) versus predicted (broken line) CPUE from 1967 to 1982.

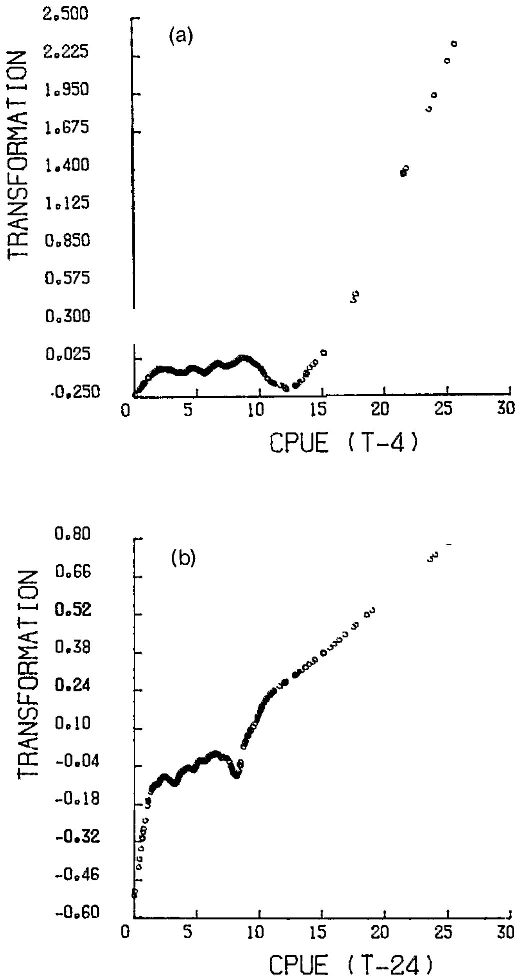

FIG. 6. Estimated optimal transformations of CPUE.

Forecasts

To test the ability of the model to forecast, the parameter values used were those estimated only from the data for the years 1966–80. The natural logarithm of CPUE was used in the forecasting model as discussed previously. The other transformations discussed in the previous section were not used in calculating the forecasts. Making use of these transformations to improve the forecasts is an area of future research. The estimated model then was used to produce one-step-ahead forecasts for the years 1981 and 1982 with no further reestimation of the parameters.This is similar to what might have been done at the end of 1980 if real-time data had been available, that is, if at the end of each fortnight we had been able to obtain the CPUE and SST data for that fortnight in order to produce the next period's forecast. With the present data collection system, real-time data are not available. However, real-time data could be made available, as SST is now collected through a satellite network, and catch and search data could be radioed in if the fishermen were willing to cooperate. Also, the real-time forecasts can be viewed as the best we can do; any delay in obtaining the data should only worsen the quality of the forecast.

As the model forecasts LNCPUE(t), we used the unbiased inverse transform (Granger and Newbold 1977, p. 307):

$$\mathrm{CPUE}(t) = \exp\{\mathrm{LNCPUE}(t) + 0.5\cdot(\mathrm{SE})\cdot 2\}$$

where SE is the standard error of the forecast on the logarithmic scale.

The forecasts for the years 1981 and 1982 give a reasonable fit to the observed data (Fig. 8). The amplitude and duration of the seasonal fluctuations are reasonably described. In 1981, when there were real SST data, the increase in CPUE at the beginning of the main cold season was anticipated by the model's forecast. In 1982, the forecast trailed by one period and was partly brought up by the increase in CPUE itself. As noted earlier, the estimated seasonal cycle for SST was used in place of the missing data in these periods, and it is likely that the seasonal cycle did not successfully capture the onset of the upwelling season that year. The actual fishing success responded to a real decrease in temperature, while the filled in data contained no change in SST that would predict a rise in CPUE. Based on the results for 1981 and the predictions for other years where real data were available, there are justifiable reasons to believe that with the proper SST data the model would have forecasted an increase in CPUE.

The model did a reasonable job of predicting the very high peaks in CPUE in 1981 and especially in 1982. This is particularly surprising, since the catch in 1982 had a sharp increase in the percentage of *S. aurita* found in the catch. Although further testing would still be needed, it does appear that this

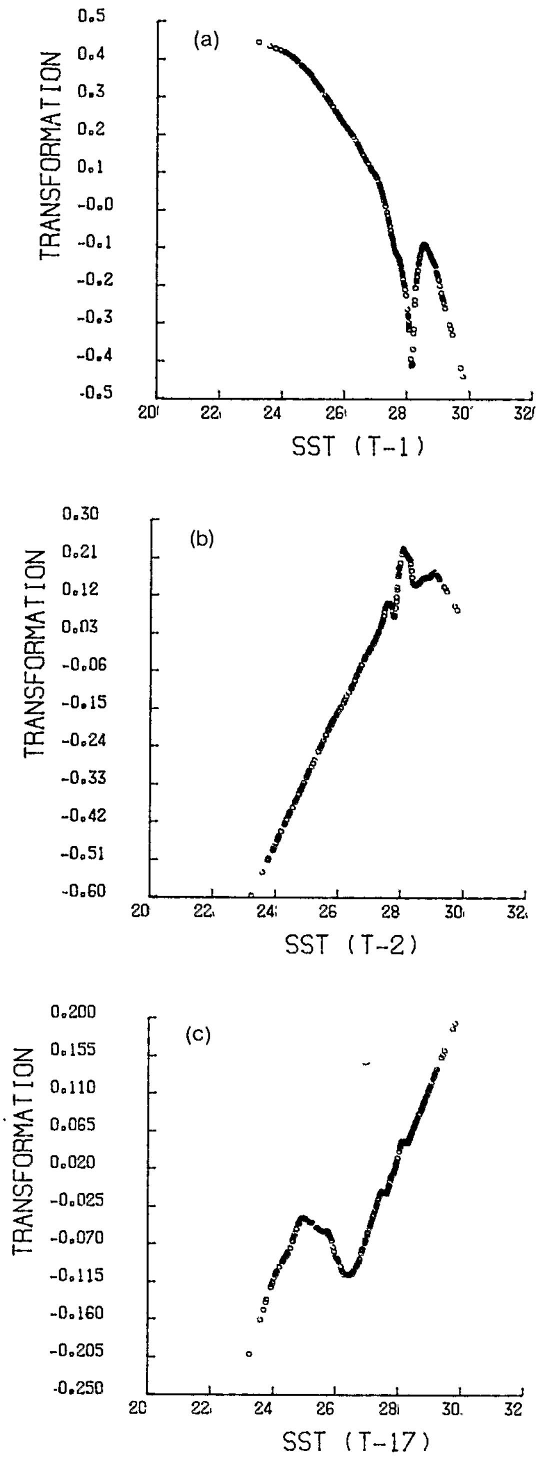

FIG. 7. Estimated optimal transformations of SST.

first-pass effort at modeling the dynamics could produce useful forecasts of the start and close of the fishing season and of the relative success in fishing that could be expected.

Discussion and Conclusions

The model estimated in this paper as well as the calculated transformations appear to provide a useful understanding of how the environment and past fishing success affect present fishing conditions. When good environmental data are available, the estimated model also provides a first-pass forecast that may be able to predict the start and close of fishing seasons and to predict catch rates during this period.

We have chosen to model the combined CPUE over all the species. There is a trade-off in this decision. With CPUE combined over all the species, it is not possible to interpret the results in terms of the dynamics of an individual species. Different species may react in different ways to environmental events so that the combined result may not reflect either species' behavior accurately.

On the other hand, CPUE combined over all species is more reflective of the situation as viewed by the fishermen. Also, it is not clear how to partition the effort between species, so that any species-specific measure of CPUE may have more problems than a CPUE combined over species. Further, it is likely that all the species feed on similar food sources and respond to an increase in food availability in a similar fashion. If the main influence of the environment is due to its affect on food availability, then CPUE combined over all species should be a very reasonable measure.

There are several features of the data that need to be checked. Some of the data were "filled in" using a smoothing algorithm, and it is not desirable to have these data points unduly influence the analysis. Also, in 1982 there was a sharp change in species composition, with *S. aurita*, a species associated with colder waters, increasing dramatically over the other species which apparently do not have as strong a temperature preference.

The filled in data appear to have had little influence on the model estimation, but they significantly affected the model fit and especially the model forecasts. The CPUE data have relatively few missing data points, and these points are for the most part randomly scattered throughout the data set. Thus, the filled in CPUE values probably affected the analysis little. SST data are missing almost the entire year of 1969 and the last half of 1982. However, an almost identical model is estimated whether the years 1981 and 1982 are included or not and for either method of filling in the missing SST values. So the estimated model appears to be robust to these fill-ins.

The predicted and forecasted values, particularly for the end of 1982, appear to be affected by the filled in data. This is not surprising, as the model looks for a sharp change in SST to produce a sharp change in CPUE; the seasonal cycle of SST may not accurately reflect the timing of the onset of upwelling in 1982.

At a fortnightly time scale, the measure of CPUE used probably reflects the availability of the pelagic stocks more than it reflects the year-to-year abundance of the stocks. As discussed earlier, the model predicts strong persistence in CPUE, so that once fish are present they tend to remain. The "kick" to increase appears to be associated with a drop in temperature from two fortnights ago to one fortnight ago. This appears to be consistent with upwelling periods. The transformations suggest a strong linear effect between 22 and 27°C and then a dropoff between 28 and 29°C. This is consistent with a study of Marchal (1967) on a single year's data where he found that 23−26°C was the optimum temperature range. Binet (1976) has shown that increases in zooplankton abundance occur one fortnight after a drop in temperature, and the increase in zooplankton is strongest following the onset of upwelling. This is consistent with the interpretation of our model. The pelagic species probably come to the surface more and school more when there is abundant zooplankton biomass around which to aggregate.

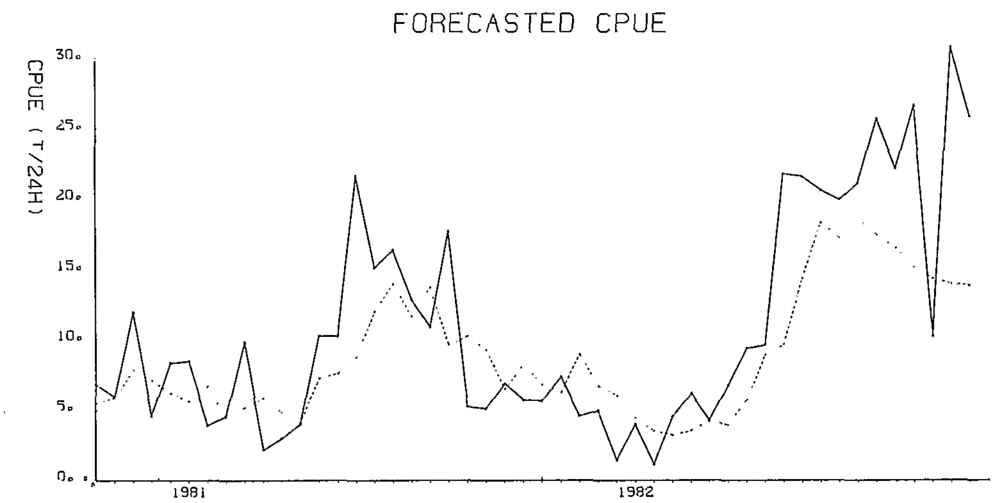

FIG. 8. Observed (solid line) and forecasted (broken line) CPUE in 1981 and 1982 per fortnight.

The CPUE of the coastal pelagic fish stocks off the Ivory Coast fluctuates greatly from one fortnight to another. Using mutivariate time series methods, a model has been constructed which explains 43% of the variance in the original series and provides reasonable forecasts of future data. The estimated model has a nice biological interpretation which is consistent with past studies on these species. By finding appropriate transformations of the original series, a model can be constructed which explains almost 60% of the total variance in the CPUE series. These transformations clarify the threshold and other nonlinear effects between CPUE and the environment.

The present results suggest that, by themselves, variables such as SST or salinity are not sufficient to explain the evolution of CPUE, particularly when measured at the same time period as the catch. Moreover, there are other environmental variables such as wind or subsurface data that were not available for our study but might also be important predictors of CPUE (Bakun and Parrish (1980) presented a list of likely environmental variables). Our results suggest that the measured environmental variables act as surrogate variables for oceanographic and biological processes that create favorable conditions for the targeted fish species. Therefore, it is the dynamics of these variables between periods that are important, not a static value at any one time period. A similar result was found by Mendelssohn and Roy (1986) for tuna in the Gulf of Guinea.

Finally, although our techniques are empirical, we have shown that they lend themselves to identifying the underlying processes involved. The modeling approach also avoids the problems of just examining cross-correlations, which are not independent of cross-correlations at other lags or of autocorrelations in either series. This can lead to spurious conclusions about relationships and the lags at which they occur. We have also used a new technique for identifying the appropriate scale or form of the relationships between the different variables under consideration.

Acknowledgements

We wish to express our gratitude to the Centre de Recherche Oceanographique d'Abidjan who made this study possible. Support for one of the authors (P. Cury) was provided through a cooperative agreement between NMFS and ORSTOM during a visit to the Pacific Fisheries Environmental Group, Monterey, California. We thank I. Barrett, A. Bakun, and J. R. Durand who made this visit possible. We are indebted to David Husby, Claude Roy, Remi Chuchla, and Denis Binet who helped us in different but complementary ways.

References

AKAIKE, H., G. KITAGAWA, E. ARAHATA, AND F. TADA. 1979. TIMSAC-78, Computer Science Monographs No. 11. The Institute of Statistical Mathematics, Tokyo. 279 p.

BAKUN, A., AND R. H. PARRISH. 1980. Environmental inputs to fishery population models for eastern boundary current regions, p. 67–104. *In* G. D. Sharp [ed.] Workshop on the effects of environmental variation on the survival of larval pelagic fishes. IOC Workshop Rep. 28.

BEVERTON, R. J. H., AND S. J. HOLT. 1957. On the dynamics of exploited fish populations. U.K. Minist. Agric. Fish Food. Fish. Invest. (Ser. 2) 19: 533 p.

BINET, D. 1976. Biovolumes et poids secs zooplanctoniques en relation avec le milieu pelagique au dessus du plateau ivoirien. Cah. ORSTOM Ser. Oceanogr. 14: 301–326.

1982. Influence des variations climatiques sur la pecherie des *Sardinella aurita* ivoiro-ghaneennes: relation secheresse-surpeche. Oceanol. Acta 5: 443–452.

1983. Zooplancton des regions cotieres a upwellings saisonniers du golfe de Guinee. Cah. ORSTOM Ser. Oceanogr. 18: 357–380.

BOUDREAULT, F. R., J. NORMAND DUPONT, AND C. SYLVAIN. 1977. Modeles linearies de prediction des debarquements de homard aux iles de la Madeleine (Golfe du Saint-Laurent). J. Fish. Res. Board Can. 34: 379–383.

BREIMAN, L., AND J. H. FRIEDMAN. 1985. Estimating optimal transformations for multiple regression and correlation. J. Am. Stat. Assoc. 80: 580–619.

CURY, P., AND C. ROY. 1985. Small pelagic fish abundance and upwelling off the Ivory Coast. Presented at UNESCO Symposium on vertical motions in the equatorial upper ocean and its effect upon living resources and the atmosphere, Paris, May 6–10.

DEMPSTER, A. P., N. M. LAIRD, AND D. B. RUBIN. 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B 39: 1–38.

DIA, A. E. K. 1972. Etude de la nutrition de certains clupeides de Cote D'Ivoire. Projet de developpement de la peche pelagique cotiere, Ministere de la production animale. R/S 11/72: 33 p.

EBERHARDT, L. L. 1970. Correlation, regression, and density dependence. Ecology 51: 306–310.

FAO. 1974. Rapport terminal du projet FI: DP/IVC/66/506: Etude et mise en valeur des ressources en poissons pelagiques cotiers. Conclusions et recommandations. FAO, Rome. 72 p.

FONTENEAU, A., AND E. G. MARCHAL. 1970. Recolte, stockage et traitement des donnees statistiques relatives a la peche des sardiniers (filet tournant) en Cote D'Ivoire. Doc. Sci. C.R.O. Abidjan. 1: 21–30.

FOX, W. W. 1970. An exponential surplus yield model for optimizing exploited fish populations. Trans. Am. Fish. Soc. 99: 80–88.

GRANGER, C. W. J., AND P. NEWBOLD. 1977. Forecasting economic time series. Economic theory and mathematical economics series. Academic

Press, New York, NY. 333 p.
JENKINS, G. M. 1979. Practical experiences with modelling and forecasting time series. Time Series Library. Gwilym Jenkins & Partners (overseas) Ltd., Jersey, Channel Islands. 146 p.
MARCHAL, E. G. 1967. La peche des sardiniers ivoiriens en 1966. Doc. Sci. Prov. C.R.O. Abidjan. 19: 26 p.
MENDELSSOHN, R. 1981. Using Box–Jenkins models to forecast fishery dynamics: identification, estimation, and checking. Fish. Bull. 78: 887–896.
MENDELSSOHN, R., AND C. ROY. 1986. Environmental influences on the french, Ivory Coast, Senegalese and Moroccan tuna catches in the gulf of Guinea. *In* Special publication of the International Skipjack Year Program, ICCAT, Madrid.
MORLIERE, A. 1970. Les saisons marines devant Abidjan. Doc. Sci. C.R.O. Abidjan. 1: 1–15.
ORSTOM. 1976. Rapport du groupe de travail sur la sardinelle (*S. aurita*) des cotes ivoiro-ghaneennes. Fishery Research Unit Tema, Centre de Recherches Oceanographiques Abidjan, ORSTOM. 62 p.
PARRISH, R. H., A. BAKUN, D. M. HUSBY, AND C. S. NELSON. 1983. Comparative climatology of selected environmental processes in relation to eastern boundary current pelagic fish reproduction, p. 731–777. *In* G. D. Sharp and J. Csirke [ed.] Proceedings of the Expert Consultation to examine changes in abundance and species of neritic fish resources, Costa Rica, 18–29 April 1983. FAO Fish. Rep. 291.
PELLA, J. J., AND P. K. TOMLINSON. 1969. A generalized stock production model. Bull. IATTC 13: 419–496.
PICAUT, J. 1983. Propagation of the seasonal upwelling in the eastern equatorial Atlantic. J. Phys. Oceanog. 13: 18–37.
1985. Major dynamics affecting the eastern tropical Atlantic and Pacific oceans. Calif. Coop. Oceanic Fish. Invest. Rep. 26: 41–50.
RICKER, W. E. 1958. Handbook of computations for biological statistics of fish populations. Bull. Fish. Res. Board Can. 119: 300 p.
SAILA, S. B., M. WIGBOUT, AND R. J. LERMIT. 1980. Comparison of some time series models for the analysis of fisheries data. J. Cons. Int. Explor. Mer 39: 44–52.
SCHAEFER, M. B. 1954. Some aspects of the dynamics of population important to the management of the commercial marine fisheries. Bull. Int. Am. Trop. Tuna Comm. 1: 26–56.
SHUMWAY, R. H., AND D. S. STOFFER. 1982. An approach to time series smoothing and forecasting using the EM algorithm. J. Time Ser. Anal. 3: 253–264.
TIAO, G. C., AND G. E. P. BOX. 1981. Modeling multiple time series with applications. J. Am. Stat. Assoc. 76: 802–816.
TIAO, G. C., G. E. P. BOX, M. R. GRUPE, G. B. HUDAK, W. R. BELL, AND I. CHANG. 1980. The Wisconsin multiple time series (WMTS-1) program. A preliminary guide. Department of Statistics, University of Wisconsin, Madison, WI. 75 p.

Appendix

Shumway–Stoffer Missing Data Algorithm

The Shumway–Stoffer (1982) missing data algorithm calculates maximum likelihood estimates of the parameters of the following state–space model:

$$\begin{aligned} x(t) &= Ax(t-1) + e(t) \\ y(t) &= Hx(t) \qquad + w(t) \end{aligned}$$

where $x(t)$ is an underlying state of the system at time t, A is a matrix of parameters to be estimated relating the underlying state at time t to that at time $t-1$, $e(t)$ is the process error vector, $y(t)$ is the observed process at time t, H is a known matrix relating the underlying and observed processes, and $w(t)$ is a random vector, the observation error. The algorithm not only calculates maximum likelihood estimates of the parameters in the matrix A, but also calculates minimum mean-square error estimates of the underlying state $x(t)$, which is only observed with error in $y(t)$.

The algorithm has been extended to allow for missing data (i.e. missing components in the $y(t)$ vector), and the given reference shows how to put a multivariate autoregressive model into the required state–space form. The algorithm is a variant of the E–M algorithm of Dempster et al. (1977) and briefly proceeds as follows. For any estimate of the parameters, the algorithm uses a form of the Kalman filter to calculate the likelihood for that set of parameters. Minimum mean-square error estimates of the state vector $x(t)$ are then estimated using a slightly modified version of the Kalman fixed-interval smoothing algorithm. The new parameter estimates are then calculated by what is a regression-like equation. The next iteration then begins. The algorithm stops when the parameters have converged sufficiently.

For the applications in this paper, for example, the vector $y(t)$ is the observed CPUE in areas 4–7, the underlying state $x(t)$ is the "true" CPUE or density in areas 4–7 and the matrix A contains the parameters of a second-order autoregressive process. For the SST data, a second-order autoregressive process was estimated for each pair of bordering north–south areas.

Alternating Conditional Expectation (ACE) Algorithm

The ACE algorithm of Breiman and Friedman (1985) is an algorithm that estimates optimal transformations of variables in regression-like equations. The algorithm operates on the observation that for any given empirical transformations of the data, if we fix the values of all but one variable and solve the problem of what new transformation will minimize the normalized residual sum-of-squares, then the solution is a conditional expectation that can be empirically estimated using a smoothing algorithm.

The algorithm starts with the given variables and then alternately estimates the conditional expectations that would be derived from fixing all but one variable in turn. Hence, the name of the algorithm. The algorithm terminates when the changes in the estimated transformations are sufficiently small. The algorithm does not produce a given equation, but rather an empirical transformation of each data point for each variable. The shape of the transformation is found by plotting the transformed values of a variable versus the original values.

Reprinted from

Réimpression du

Canadian Journal of Fisheries and Aquatic Sciences

Journal canadien des sciences halieutiques et aquatiques

Fluctuations of a fortnightly abundance index of the Ivoirian coastal pelagic species and associated environmental conditions

R. MENDELSSOHN AND P. CURY

Volume 44 • Number 2 • 1987

Pages 408–421

Canada

Fisheries and Oceans
Pêches et Océans

Printed in Canada by K. G. Campbell Corporation