

Intergovernmental Oceanographic Commission

Workshop Report No. 124

GLOBEC-IOC-SAHFOS-MBA Workshop on the Analysis of Time Series with Particular Reference to the Continuous Plankton Recorder Survey

Plymouth, United Kingdom
4-7 May 1993

Edited by J.C. Gamble⁺ and M. Edwards

TABLE OF CONTENTS

	Page
PREFACE(iii)
INTRODUCTION(iv)
BACKGROUND TO WORKSHOP	(v)
TERMS OF REFERENCE	(vii)
SUMMARY REPORT	1
1. MISSING DATA IN TIME SERIES ANALYSIS	1
1.1 PRELIMINARY DATA ANALYSIS AND HANDLING MISSING DATA	1
1.2 IRREGULARLY SPACED DATA	1
1.3 MISSING DATA -V- SAMPLING INTENSITY	1
1.4 TECHNIQUES FOR HANDLING MISSING SAMPLES	2
1.5 SPATIAL-TEMPORAL REPRESENTATION OF CPR DATA	2
1.6 PERIODIC MISSING DATA AND GENERAL ADVICE	3
1.7 VALIDATION OF THE CPR DATA SET	3
2. UNIVARIATE TIME SERIES	3
2.1 SEASONAL PATTERNS AND TRENDS	3
2.2 TESTING FOR TRENDS.....	4
2.3 MODEL FITTING	4
2.4 VARIANCE	5
3. INTERCOMPARISONS BETWEEN TIME SERIES	5
3.1 PRETREATMENT OF DATA...	5
3.2 MODELLING TECHNIQUES..	6
4. MULTIVARIATE TIME SERIES..	7
4.1 DATA REDUCTION.....	7
4.2 USE OF THE ENTIRE ASSEMBLAGE	7
RECOMMENDATIONS	7
5.1 DATA PRESENTATION	7
5.2 MISSING DATA.....	8
5.3 IDENTIFICATION OF TRENDS,	8
5.4 TIME SERIES MODELS	8
5.5 INTERPRETATION	8
REFERENCES	9
ANNEXES	
I. List of Participants	
II. Papers submitted at Workshop	

PREFACE

At the first Annual General Meeting of the Sir Alister Hardy Foundation for Ocean Science (SAHFOS) on the 28th April 1992, a recommendation was made that the database of the Continuous Plankton Recorder (CPR) should be: "the subject of a workshop which will examine its value to current oceanographic research initiatives, will consider the best means by which data and information can be disseminated in the oceanographic community and make recommendations for new methods of interpretation, Sponsorship for this workshop should be sought from international agencies". As the first Director of the Foundation, Dr John Gamble energetically developed and implemented plans for this workshop which was hosted by SAHFOS in Plymouth in May 1993.

The workshop was a response to a recognised need for an improved statistical understanding of the temporal and spatial variability of the large database of plankton measurements compiled by the CPR survey since 1931 (more than 4 million miles of tow and 170,000 samples as of December 1995), John Gamble was active within the Global Ocean Eco-system Dynamics International (GLOBEC INT) community in promoting retrospective analysis of long term data sets. Time Series and Retrospective Data Analysis form one of the main themes of the GLOBEC International science plan. The conclusions of the workshop can be seen as one of the first products that contributes to this plan.

This report provides a record of the presentations made at the Time Series Workshop. Due to the tragic death of Dr Gamble in August 1994 publication of the report has regretably been long delayed. The contributions made to the workshop are however, as relevant to the study of long time series today as they were in 1993, While the workshop specifically focused on the CPR data set, many of the points raised are relevant to the interpretation of other long term data sets. The workshop discussed for example, methods of intercomparison between different data sets and the optimal design for new monitoring systems. The papers presented have relevance to the development of new monitoring programmes that may be established under the Global Ocean Observation System (GOOS) and GLOBEC.

Four main themes were addressed by the workshop:

- (i) Temporal and spatial missing data;
- (ii) Analysis of trends in univariate time series;
- (iii) Intercomparisons between time series; and
- (iv) Analysis of multivariate time series.

A total of 19 recommendations were made which, in addition to points made specifically in relation to the terms of reference (see page 9), included some on the use of gee-statistical techniques and GIS systems. Sponsorship of the workshop by GLOBEC INT, the International Oceanographic Commission the SAHFOS funding consortia and the Marine Biological Association of the UK is gratefully acknowledged.

I wish to thank Martin Edwards for the effort he has put into tracking down and compiling the various contributions to this report.

Philip C. Reid
Director SAHFOS

INTRODUCTION

THE CONTINUOUS PLANKTON RECORDER SURVEY

This survey was established in 1930 by the British marine biologist Alister Hardy. He had developed a robust instrument for continuously recording the abundance of mesoscale plankton as it was towed by commercial 'ships of opportunity'. By operating several of these recorders simultaneously and at regular intervals, Hardy was able to establish a quasi-synoptic survey of the spatial and temporal patterns of the abundance of plankton populations in the upper mixed layers of the seas around the British Isles. Since 1946 the Continuous Plankton Recorder (CPR) Survey has extended to cover the North Atlantic, including coastal routes down the north-eastern seaboard of North America. The Survey reached its peak coverage in the period between 1965-1980, but has subsequently declined until 1991 when it was re-invigorated under the auspices of the Sir Alister Hardy Foundation for Ocean Science. An extensive route network extends around the European shelf centred on the British Isles,

The mechanics of the Survey are summarised by Hunt in this report but, briefly, the Survey routinely monitors the distribution and abundance of about 150 phytoplankton and 200 zooplankton taxa. In the years immediately after establishment of the Survey, the investigations were mostly centred upon distributional patterns of the plankton together with the elucidation of seasonal patterns of distribution. The distributional data culminated in the production of a distributional atlas of the plankton sampled by the CPR in the North Atlantic. Many important distributions were described including specific regional patterns by which North Atlantic plankton could be categorised, e.g. neritic, north-east oceanic, etc. (Colebrook, 1972), and the complementary distributions of congeneric copepods such as *C. finmarchicus* and *C. helgolandicus*, which ultimately helped to confirm their separation into two distinct species,

While the description of seasonal patterns has always played an important part in the evaluation of the CPR database, it was only after about two decades that systematic trends could be extracted from the data. These were often coupled with analyses of the seasonality of the data where clear demonstrations of seasonal succession of occurrence of species were made, and where long term changes in the timing of the onset of peak growth within specific taxonomic categories were described.

Studies of the changes in abundance of single species during the multidecadal time series of the Survey have been a consistent feature of CPR data analysis. Such investigations most frequently use the inherent mobility of plankton populations to characterise anomalous hydrographic conditions, i.e. to use individual species as temporary indicators of transient change. A recent example has been the increase in the abundance of several species of *Ceratium* in the northern North Sea in the late 1980s which has been explained by an unusual incursion of low salinity water from the Baltic Sea. However such events are usually short-term and it is the long-term changes in the plankton populations which are the results of greatest interest from the Survey.

The procedure used to extract this information from the CPR Survey is one based upon the use of PCA to summarise the data from selected groups of the commoner plankton species. This method has been used very extensively and successfully for many years but recently the application appears to have broken down for zooplankton, due possibly to a change in the relative structural composition of the populations during the course of the time series. Nevertheless this method is at present the only one available for summarising CPR data and is used in the routine incrementation of the annual fluctuations in the trends of abundance of grouped phytoplankton and zooplankton. In turn these long-term changes have been related to environmental and climatic factors where similarities in the patterns have been interpreted as an indication of the effect of long-term environmental change on planktonic ecosystems.

Much of the recent work has therefore been concerned with the interpretation of the long-term changes in these few, selected taxa. The seasonality has been examined, as has the significance of seasonal patterns on the overall control of long-term events.

The interannual persistence observed in many of the populations is believed to have its genesis in their overwintering strategies. This has, however, left the bulk of the species routinely counted by the CPR ignored, which presents in itself an opportunity for an entirely new approach to the long-term time series of the Survey. In addition, the apparent breakdown of the validity of the PCA analysis invites the use of other time series techniques which might be more appropriate to the idiosyncrasies of the CPR data set.

Reference:

Colebrook, J.M. 1972. Variability in the Distribution and Abundance of the Plankton. *ICNAF Spec. Publ. No. 8, 167-186*

BACKGROUND TO WORKSHOP

In the past decade there has been a reappraisal of the importance of long-term time-series of ecological data. From being regarded as mere monitoring exercises necessary for environmental management and decision making, they have become recognised as an essential component in the description and understanding of ecological processes. It is now realised that the environment is subject to change on a global scale as the increasing production of radiatively active gases affects atmospheric processes. In this context therefore it is even more necessary to maintain time-series which can monitor change, provide a baseline for interpretation of contemporaneous data and serve as a lead for future prediction,

Marine ecosystems of course are also subject to anthropogenic influences, particularly through fishing, mineral exploitation and pollution. The longest and most detailed marine time-series are fishery records but analyses of these are often compromised by the directed nature of their collection. However, the need to understand the causes of fluctuation in fisheries and particularly the factors controlling variation in recruitment have led to the establishment of some of the best known planktonic time-series including CalCOFI (Chelton et al. 1982) off the Californian coast, MARMAP (Sherman, 1980) off the New England seaboard and the Continuous Plankton Recorder Survey (CPR) in the North Atlantic (CPR Survey Team, 1992). These are wide spatial surveys carried out on a regular basis but other, more erratically acquired sets of data exist such as those made in the sub-Arctic Pacific (Brodeur & Ware, 1992). In addition to these spatial surveys many planktonic data sets also exist based on single site monitoring, often in the vicinity of marine institutes.

Examples of these in Europe are station EI in the English Channel off Plymouth, (Southwood, 1980) where the Russell Cycle was described, and the Helgoland Survey (Radach et al, 1990) where the planktonic data are well supported by physical and chemical information.

The main purpose of the Workshop was to address methods of analysing long term planktonic time-series with particular reference to the Continuous Plankton Recorder Survey. Unlike all other planktonic surveys, the CPR Survey is based upon the continuous collection of planktonic organisms made by the specially designed Continuous Plankton Recorder while towed on regular routes by commercial "ships of opportunity". The basic sampling unit is an arbitrarily determined subdivision (10 nautical miles) of each individual tow route. The CPR procedure thus differs radically from the vertically integrated net hauls carried out at predetermined regular grid positions as in the CalCOFI and MARMAP surveys. In broad terms this strategy of the CPR has the advantage of enabling, at acceptable operational cost, regular surveys on an oceanic, basin-wide scale, throughout the year. However, the CPR Survey has serious drawbacks and limitations, some of which were addressed at the Workshop.

The fundamental operational principle of the CPR is to sample plankton along the track of the towing vessel by trapping it on a continuously moving (approximately 15cm wide) band of silk mesh. The tractive power for the silk winding mechanism is provided by a propeller turned as the recorder is towed through the water. Adjustments are made to the pitch of the propeller blades to ensure that an approximate 10cm length of mesh band corresponds to 10 nautical miles of towing.

While this has proved to be an ingenious and reliable system, it has several disadvantages. Perhaps the most serious is the lack of replication at the point of sampling (although the meaning of replication in pelagic sampling is ill-defined). Secondly, the variability of the sampled organisms can be very great between successive sampling units and is dominated in many cases by a preponderance of zero values. Thirdly, while the Survey is highly successful in operational terms, individual tows are missed due to adverse weather, equipment failure or to uncontrollable changes to the scheduling of towing vessels. Finally, the analytical procedures are restricted by the need to maintain the integrity of the multidecadal database which means that practices such as the categorisation of the abundance of the individuals (as opposed to counting) must be accommodated. In essence we have to deal with a distorted dataset, where replication can only be made between points along a transect, or between temporally distinct measurements in the same locality, and where the basic numerical unit is on a geometrically-defined scale of abundance,

CPR analysts measure the abundance of almost 400 different taxonomic entities defined pragmatically to their optimal taxonomic level. The survey is thus a multivariate database of many co-varying measurements and a principal objective of the analysis of the data is the extraction of the best common signal to compare with other long-term environmental variables. In the past, multivariate analysis of the CPR time-series has almost exclusively made use of Principal Component Analysis. A primary role of the workshop was to identify other possible methods of analysing this unique dataset.

The questions investigated by the Survey relate to many aspects of wide-scale plankton ecology. Some of the main areas of interest are :-

- (i) Defining the basin scale patterns of plankton distribution.
- (ii) Describing the seasonal and interannual changes in distribution and abundance of plankton,
- (iii) Determining transient changes in distribution and abundance of plankton.
- (iv) Describing the structure of plankton populations,
- (v) Describing long-term trends in the abundance of planktonic organisms.
- (vi) Comparing trends in planktonic systems to trends in abiotic environmental parameters; ascribing and interpreting causes of change,
- (vii) Understanding the biological and physical processes which underlie the observed temporal trends and spatial patterns,
- (viii) Relating the spatial and temporal variability of plankton to parallel or complementary changes at other trophic levels,
- (ix) Isolating and identifying changes in the patterns of abundance of individual species.

Among the topics discussed were:

- detection of changes in periodicity within the time-series;
- determining annual periodicity, e.g. the timing of blooms, length of productive season;
- calculating annual abundance from spatio-temporally variable sampling in different years;
- estimating the error associated with annual abundance estimates and distinguishing accuracy from precision;
- consideration of the CPR sampling design in order to use manpower and recorder time to maximise the accuracy of annual abundance estimates;
- determination of long-term trends and anomalous deviation from trends.
- usage of multivariate methods to analyse multi-species datasets.

The CPR Survey is also concerned with spatial patterns, and preferred means of describing such patterns were also considered,

While the main focus of the Workshop was to advance the methods applied to the future analysis of the Continuous Plankton Recorder Survey, a second objective was to consider the analysis of long-term datasets in general. Existing datasets only cover small areas or are in localised sites and may not be extrapolated to global issues. New initiatives such as GLOBEC and the Global Ocean Observing System (GOOS) are developing new investigations and monitoring systems on a world-wide scale. In this context the Workshop discussed methods of inter-comparison between long-term datasets and also the optimal designs for new monitoring systems.

TERMS OF REFERENCE FOR THE WORKSHOP DISCUSSION

The Workshop addressed the following questions relating to the data from the Continuous Plankton Recorder Survey. These formed the terms of reference for the discussions and for the ensuing recommendations.

- (i) The methods for handling both temporal and spatial missing data. What are the preferred ways of filling the gaps?
- (ii) The analysis of trends in univariate time series. How do we measure periodicity, change and significant deviation? What are the best ways of condensing data and coping with the dominating seasonal periodicity?
- (iii) The means of making inter-comparisons between time series. How do we describe similarity and how do we measure changes in similarity?
- (iv) The analysis of multivariate time series. What means should be used to summarise multivariate data sets but retaining as much information as possible? How do we make use of all the data including rare as well as common taxa?

SUMMARY REPORT

1. MISSING DATA IN TIME SERIES ANALYSIS

1.1 PRELIMINARY DATA ANALYSIS AND HANDLING MISSING DATA

Information must be provided on the way data have been collected. This is essential for making initial decisions on the preferred means of handling anomalous values before entering into more complex analytical and modelling procedures. In particular, information should be provided on:

- (i) accuracy of monthly means, e.g. by quoting the number of 10-mile blocks used in averaging the data.
- (ii) the cruise tracks made during the period of investigation together with any changes in coverage of the area of concern.

1.2 IRREGULARLY SPACED DATA

It is useful to distinguish between missing data and irregularly spaced data. In time series analysis, missing data typically refer to occasional missing observations from a regularly spaced series. In contrast, irregularly spaced data typically refer to data in which there is no underlying sampling interval. For the most part, monthly samples from the CPR are regarded in the context of long-term time series analysis as being regularly spaced, although the precise time of month collected depends upon the availability of the appropriate ship of opportunity,

There are various categories of missing data and it is important to distinguish between data that are missing at random and data that are missing in a systematic way. In the CPR Survey missing data can arise from:

- (a) loss of one to several samples (blocks) along a particular route due, for instance, to malfunctions such as inadequate preservation, physical blockage of the sampler and adverse weather conditions,
- (b) loss of complete individual routes, again through unpredicted malfunction but also due to the rescheduling of tow vessels, and
- (c) loss of specific routes over a prolonged period through the unavailability of suitable ships of opportunity such as the withdrawal of ferry-routes during a w-inter closed season.

Quite clearly it is difficult to categorise the missing data precisely into random and systematic losses although it is probably correct to infer that instances in groups (a) and (b) above are random and (c) is more systematic.

1.3 MISSING DATA -V- SAMPLING INTENSITY

Techniques are outlined below, using mean monthly abundance values, "filling in" months for which there were no data. However, these different methods make no allowance of the fact that very different sampling intensities may have contributed to the mean monthly figures, i.e. a monthly mean calculated from 3 samples would be treated in exactly the same way as a monthly mean derived from 300 samples. There is, therefore, a need to investigate how important sampling intensity is in the context of calculating mean annual abundance figures, and how the variations in sampling intensity within the time series influence the confidence that can be placed in the calculated mean annual abundance figure.

One method of examining this problem is through "Monte Carlo" simulations. For example, in years in which many samples were collected, samples should be randomly dropped from the analysis, i.e. a reduction in sampling intensity is simulated, and the effect of this reduction in sample

size on the calculated mean monthly and mean annual abundance estimates could then be quantified. Such a procedure would help establish the relative importance of (i) sampling intensity, and (ii) "missing" data, on the calculated annual abundance figures, This is an important point as it will provide direction for future research into this problem. If, for example, the effect of the method of filling in missing data is trivial when compared to the effect of changes in sampling intensity, then establishing novel techniques for filling in missing data are clearly less important than methods for taking account of the changes in sampling intensity that occur throughout the CPR time-series.

1.4 TECHNIQUES FOR HANDLING MISSING SAMPLES

Missing samples are not a problem, provided the time-series can be modelled and the data are missing at random positions in the series. However, where the occurrence of missing data follows a seasonal pattern, the treatment of missing data may be more problematical. There is no strict recommendation for a particular technique but this will depend upon the distribution of plankton abundance or biomass within the year, In theory, given a suitable model, only one non-zero point is required for estimating the remainder of the year,

Where the missing data are occasional and random the preferred procedure will be to find estimates for the missing values and then apply standard time series methods. Traditionally much CPR analysis has been carried out upon annual means derived from averaging out the monthly mean values for specifically defined areas, Four different methods of estimating annual production (or mean annual abundance or biomass) of plankton in the presence of missing monthly values in month-year matrices were considered (see Appendix 2. "Treatment of Missing Data in CPR Data", Howell and Nicholson). The first method, developed by Colebrook (1975) for the CPR data, provides explicit estimates of the missing values by multiplying the long-term monthly mean corresponding to a missing month in a given year by the ratio of the sum of the non-missing values in that year to the sum of the corresponding long-term monthly means. The second and third methods assume that the production curve (i.e. monthly standing crop levels) follows a Normal probability density function. In the second method the duration of production is different in each year while in the third method the duration is fixed. The fourth method is based on a regression of the observations within a year on the estimated monthly effects,

The methods were compared on the basis of the mean square error of prediction (MSPE) for different patterns of missing data for *Calanus spp.* sampled by the CPR in the North Sea. With the exception of method two, which performed less well when information at the onset of plankton production is missing, the methods all gave MSPE's of approximately 50%. It should be noted that the *Calanus* production cycle tended to follow a Normal curve with consistent duration and similar timing each year, In this case the performance of the methods would be expected to be similar whereas application to other data sets may give different results,

The original Colebrook method might be improved by interpolating the value of a missing month not from values in the same year, but from values from the five months before and the five months after the missing one, An alternative approach, (Parzen,1963), uses the original data to estimate the auto-covariance function, and hence the spectral density function. Once the auto-covariance function is determined this will suggest a suitable time series model, and this can be fitted using the EM algorithm,

1.5 SPATIO-TEMPORAL REPRESENTATION OF CPR DATA

The conventional approach to handling missing data ignores the fact that the data are three-dimensional on a fine scale since samples are fixed in two spatial dimensions (latitude/longitude) and also in time. Moreover, samples that are spatially and temporally close are likely to be highly correlated, so that it should be possible to interpolate missing values from existing "nearby" (in space and time) values. Geostatistical methods that enable interpolation of missing points from observed spatial data can be extended to the 3-dimensional spatio-temporal situation. Such techniques should allow the creation of interpolated synoptic maps on a fine scale (e.g. every half-degree grid square) at any point in time and, given existing tow data, each interpolated point would be accompanied by an estimate of its reliability, i.e., its variance.

Values (and standard errors) on a standard area basis could be recreated by averaging (weighted with the reciprocal of the variance) all grid points that fall within a particular year, month and area.

With this technique, the notion of "missing value" becomes meaningless, as interpolation occurs throughout space and time at a scale considerable finer than that of year, month and standard area. Oliver and Webster (1991) give a simple overview of geostatistical analysis and provide more technical references,

1.6 PERIODIC MISSING DATA

Methods for estimating the spectral density function where the missing observations have a periodic structure are discussed by Parzen (1963) and Jones (1971). If the sampling interval changes, e.g. from monthly to bimonthly, then the approach described by Neave (1970) could be used. If the data occur at irregular sampling intervals then methods of fitting time series models to irregularly spaced data should be applied.

General problems of missing data are discussed by Bloomfield (1970) and by Little and Rubin (1987). The standard approach is through maximum likelihood, which does not require complete data. In many cases, however, the implementation of maximum likelihood is complicated by missing data and the EM algorithm is used (Parzen, 1984). The key point is that the spectrum of the point process of missing observations enters the calculations. The problem is most conveniently approached through state space representation using the Kalman Filter.

1.7 VALIDATION OF THE CPR DATA SET

Estimates and trends in the CPR data set need to be validated against those from other time series such as those from the Helgoland Reede Survey, although it is realised that specific inter-comparisons might be difficult because of the nature and form of the basic CPR data. It will therefore be necessary to convert the CPR data to some standardised values, such as biomass, which can then be compared more directly with other long-term data sets,

2. UNIVARIATE TIME SERIES

2.1 SEASONAL PATTERNS AND TRENDS

Seasonal patterns in time series data can be readily characterised using spectral analysis or other equivalent representations (e.g. the autocorrelation function). These methods will allow for quantification of periodicity in time series data. However any consideration of patterns in the CPR data must take account of fundamental aspects of the data such as diel variation in distribution, and it will be necessary to stratify the data by day and night periods prior to constructing the time series. It is also important to know the level of sampling intensity (for each monthly mean) in order to fit an appropriate time series model and also for constructing confidence intervals for estimates of mean annual abundance.

As it may be of interest to characterise interannual changes in seasonal periodicity or related factors, such as the start of the annual cycle or its duration, and it might also be possible to develop descriptors of these basic features using a model-based approach (see 2,3). For example, the seasonal cycle may be adequately described by a normal distribution, and estimation of the parameters for the normal model will provide estimates of the timing (mean) of the cycle peak. The variance will provide a measure of the spread of the seasonal period while the descriptors may be used for inter-annual and inter-regional comparisons.

In some cases a descriptive measure of the time series over a seasonal cycle is required. This is most commonly accomplished in plankton data series by integrating under the seasonal production curve.

2.2 TESTING FOR TREND

Testing for trend is one of the most common problems in time series analysis. There are many different approaches, Brillinger (1989), for instance, has devised a test statistic for detecting a monotonic trend superimposed on a stationary time series. However there are difficulties in determining whether a trend really exists. For example a 'trend' may be due to low frequency variation. If the time series is too short it will not be possible to determine which of these two possibilities is correct. Additional factors that need to be considered include whether nonstationarity is treated as drift (i.e. to be removed by differencing) or trend (to be removed by estimation); whether a parametric form for the potential trend is known (e.g. polynomial in time or a parametric function of regressions) or not (in which case, estimation is by kernel smoothing); whether or not the distribution law of the time series is known; and whether or not the time series is serially independent.

As an example, consider the common model $Y_t = m_t + e_t$, where m_t is the trend and e_t is a zero mean normal error process with constant variance. Under the null hypothesis $H_0: m_t = m$, m is estimated by the average of the time series with residual sum of squares RSS_0 . Under the general alternative, m_t can be estimated by kernel smoothing with residual sum of squares RSS_1 . A natural statistic for testing the trend is $D = RSS_0 - RSS_1$. Under the assumption that the errors are independent, the distribution of D is asymptotically Chi-Squared with degrees of freedom that can be approximated from the specific form of the kernel. If the time series is short, the asymptotic result may not hold and the null distribution of D can be estimated by randomisation. Note that data-based bandwidth selection is not taken into account in the asymptotic test and is only taken into account in the randomisation test if it is repeated for each randomisation. The randomisation test is only valid if the error process is serially independent. There are a number of approaches that can be used when serial dependence is a possibility (Altmann, 1990). These typically involve fitting a model to e_t and simulating from the model.

The methods available to remove seasonal trend from the data range from fitting a periodic function to the data and reducing the "series to the residuals from the model, to more sophisticated methods designed to filter periodic components from time series data. Certain nonparametric techniques such as STL have been used in this context (Cleveland et al, 1990). Other approaches in a time series modelling context that are used to remove unseasonal data include n th order differencing. For example, for monthly data, it is often possible to remove a seasonal cycle using 12th differences. Further time series modelling is then based on the difference series.

Trend detection by conventional time series methods is often hampered by such problems as insufficient data, censored data, missing data and irregular sampling times. A variety of distribution-free tests have been devised for trend analysis that do not require a time-series decomposition (Berryman 1988).

The exact nature of the series will determine the appropriate test. In the case of CPR series, which are characterized by seasonality and persistence, an "intra-block" test is probably suitable (Hirsch and Slack, 1984). This and related tests are an active area of investigation in long-term water quality assessments (Hipel 1988). They are easy to implement and of immediate use in the analysis of CPR series.

2.3 MODEL FITTING

A well fitting model will provide an adequate summary of a time series and any specific details, such as estimates of long-term trends and seasonal parameters, can be extracted from it. These can also be derived by more direct filtering and smoothing procedures. However obtained, the trend should always be displayed with the original data to provide a visual measure of the inherent variation in the data throughout the period of observation. Because of the importance of recognizing changes in seasonal pattern from year to year, simpler plots displaying the yearly annual patterns may be more useful. A simple procedure would be to log the (positive) data, remove a smoothed trend and examine the residual seasonal pattern each year. If a model is used, changes in seasonal pattern and occurrence of more general anomalies can be detected by monitoring the forecast errors of the model. The individual components of the model can also be generated, (such as level, gradient of trend, etc.),

and can then be monitored separately. The start of an anomalous increase would be indicated by unexpected increases in the projected gradient of the trend.

Model estimation should be subject to the usual residual checking and detection of points of influence for possible reassessment. Regular, semi-annual or even monthly, revision of the model parameters should be made, as should regular, though less frequent reassessment, of the model itself.

No software is currently available commercially for developing the kinds of models which the specific nature of the CPR data demands. It could be developed either entirely *de novo* or by using available software and skilled manpower.

Commercially available packages could be used to build more traditional models such as regression-type but these would not fit accurately although they could yield very useful information and are easy to use. Extracting trends and seasonal estimates by filtering and smoothing can be achieved easily with a large number of readily available packages.

Ideally it would be desirable to appoint someone with the responsibility to oversee model development, fitting and subsequent performance monitoring,

2.4 VARIANCE

Attention should be given to the variance of the data from the CPR Survey. Previous analyses of these data and the above considerations treat the data as point estimates with no associated variance. The estimation of the variance in this case is complicated by the probable spatial auto-correlation along a transect. Accordingly, it may be necessary to use geostatistical methods (see 1.3)

3. INTERCOMPARISONS BETWEEN TIME SERIES

3.1 PRETREATMENT OF DATA

An understanding of the structural relationships between time series variables collected in the CPR survey is extremely important and several approaches to this problem were considered. Cross-spectral analysis can provide a valuable description of the interrelationships between variables. The development of multivariable time series models is also important in this context. One approach is to develop Transfer Function models relating one or more input or 'explanatory' series to an output or 'response' series. Several aspects of this problem were considered, including the question of whether to filter or prewhiten the series during the model identification phase of examining the cross-correlations between the time series. For the case of one input and one output series, prewhitening involves developing a time series model for the input series, applying this model to both series and reducing each to the residuals from this model. The residuals are then used in a cross-correlation analysis to identify the model structure.

Prewhitening can be a conservative procedure and some information may in fact be lost, and it is not necessarily always desirable to prewhiten series. If the series to be compared are known to be causally related, it may be desirable not to use prewhitening in the identification phase. Alternatively, if there is no *a priori* knowledge of the structural relationships between the variables, prewhitening will help guard against identifying possibly spurious relationships between the series.

Variability in a time series may be due to two or more independent mechanisms. In a monthly series of total phytoplankton biomass, for example, a spring and fall bloom may arise from forces that are more or less unrelated and vary independently from one year to the next. In such cases, a certain type of principal component analysis (CA), in which monthly values are treated as variables and years as observations, can identify the underlying "modes" of variability. An example and relevant references are summarised by A. Jassby (Appendix 2).

One result of this CA is an annual time series for each significant mode, showing the intensity of that mode from year to year. Each of these "amplitude time series" can then be compared with other, possibly "causal" time series (see below, Section 3.2). Because the amplitude time series arise from a smaller number of mechanisms than the original series, the identification of causal relationships is more likely. A preliminary decomposition by this CA technique should therefore always be considered, particularly when a series represents more than one taxon (e.g. community biomass) and possibly multiple mechanisms of variability,

Even when the implicit assumptions of the method are not satisfied (such as when year to year variability is due solely to a shift in timing of the bloom), it is possible to make important deductions about the nature of the variability by examining diagnostic plots. The method must be applied with care as it has many options that can significantly affect the outcome, such as the ways in which the data matrix is first filtered and transformed, the significant components are identified, and the components are rotated,

3.2 MODELLING TECHNIQUES

Traditional time series methods for comparing two series may be overly restrictive, such as by assuming linearity. For many processes that govern phyto- and zooplankton variability, the physical and other mechanisms underlying variability can be distinctly nonlinear. There are a variety of alternative techniques that impose various levels of restriction. The most unrestricted approaches, for example, would be to explore relationships by means of nonparametric regression using thin-plate splines, neutral nets, or kernel regression models. However, the amount of data available for many of the CPR time series may be insufficient for reliable use of these algorithms,

As an intermediate step, and in addition to traditional time-series approaches, we suggest the exploration of mutual relationships between series by means of Generalized Additive Modelling (GAM). GAM imposes a restriction of additivity on predictor or explanatory variables, but allows for flexibility in the choice of response transform and variance model. In this sense, GAM can be seen as an extension of Generalized Linear Modelling (GLM). In addition to these features of GLM, however, GAM permits transformations of the predictor variables using either loess or spline smoothers. The partial dependence of the response on each predictor is not chosen by the user, but determined independently by the algorithm on the basis of the data characteristics. The user must specify only the "equivalent degrees of freedom" which determines the smoothness of the partial dependence. GAM is a recent development, but there are many examples of its successful use in biomedical applications. Hastie and Tibshirani (1990) provide a thorough introduction to GAM and related techniques. The algorithms are available most conveniently as part of the S-Plus software package, versions of which exist for Unix workstations and DOS-compatible microcomputers. A stand-alone package known as GAIM (Generalized Additive Interactive Modelling) is also available from R. Tibshirani at the University of Toronto.

An exploration with GAM will typically result in a collection of possible relationships between predictors and the response variable,

How can decisions be made among competing formulations? The inferences often used for more classical approaches may not be dependable in the context of GAM. In fact, they are often not suitable for the more classical models either. For example, comparisons that rest on the average-squared-residual do not take into account the discrepancy between this quantity and the prediction-squared-error (PSE) for future realisations of the variables (a discrepancy known as the "optimism").

Several approaches have been developed for estimating the PSE by means of bootstrap sampling. In one method, known as the "632 bootstrap" method (Efron, 1983), a bootstrap sample is selected from the original observations, the model in question is fitted to this sample, and the observations omitted from the bootstrap sample are used to estimate the PSE. About 200 replications are typically necessary,

4. MULTIVARIATE TIME SERIES

4.1 DATA REDUCTION

The multispecies time series of CPR data has traditionally been treated using Principal Component Analysis (CA) as a data reduction method. This technique is used to construct linear combinations of the original series with weightings based on variance and decreasing from the highest to lowest. The first principal component has been used to characterise the CPR time series and this component has typically been associated with trends in the series. However, CA has been shown to give misleading answers when used to summarise species-samples matrices due, in part, to the large number of zeroes in the original matrix and to the presence of non-linear relationships which are often found between different species. Hence it has become important to investigate other methods of summarizing such data.

An alternative approach is to use the autocorrelation in the series as the weighting factor as was described during the workshop by Andrew Solow (see Appendix 2). This method, Minimum/Maximum Autocorrelation Factor Analysis (MAFA), uses the lag-one autocorrelation (an index of smoothness) in the weightings. Here smoothness is associated with trend. It may be desirable to apply this technique to the CPR data. The method is similar in principle to CA but will give higher emphasis in smooth series with high autocorrelation. Accordingly, the results of such an analysis may differ substantially from those of CA where higher weight is given to species characterised by high variability. Both methods will give similar results if high variability is associated with smooth trend.

4.2 USE OF THE ENTIRE ASSEMBLAGE

The CPR analysis has tended to under-exploit the assemblage information on intermediate abundance and rare species, in its concentration on the time series of the first Principal Component or of single dominant species. There is some evidence from environmental impact studies of benthic communities (Gray 1982) that intermediate abundance species are the most informative in signaling change. Rare species are intrinsically interesting from the biodiversity perspective and the pattern of rare species can sometimes reflect physical anomalies, for species at the limits of their range.

Non-parametric multivariate methods could be beneficial in ascribing contributions of species, or higher taxonomic groups, to the overall multivariate pattern of "community" samples (for a particular subset of the samples taken in space, time or both). For example, whether the patterns of rare and common species are essentially "telling the same story" could be examined by comparing the corresponding non-metric Multi-Dimensional Scaling (MDS) plots, either by correlating the rank similarity matrices underlying each MDS (e.g. Clarke 1993) or, for a CA or Correspondence Analysis route to ordination, by a Procrustes approach (Krzanowski 1993). The same could be done with different subdivisions of the assemblage, most importantly by higher-level taxonomic or functional group.

5. RECOMMENDATIONS

5.1 DATA PRESENTATION

- Provide information on sample sizes (no. of 10-mile blocks) for summarised data to permit assessment of reliability of the data.
- Trends in the time series of CPR data should be clearly displayed graphically and made readily accessible. Smoothing methods can be employed to bring out trends.
- Visual representation of spatial information should be explored using new technologies such as Geographical Information Systems.

5.2 MISSING DATA

- Information on the spatial location of samples can be used to choose suitable proxy series to fill in missing values.
- Current methods of filling missing values used by the CPR appear to perform adequately. Use of the Nicholson and Howell method 3 and the centred Colebrook method should also be explored.
- Methods of analysis of irregularly spaced data can be useful and should be considered. The methods may involve the use of the EM algorithm or the Kalman Filters.
- Geostatistical methods that incorporate the spatial and temporal correlations within the CPR data could be useful for interpolating missing values and constructing fine-scale synoptic maps, with variance estimates for the interpolated points.

5.3 IDENTIFICATION OF TRENDS

- A variety of distribution-free methods for trend detection have been discussed. These methods can be easily applied to the CPR series.
- Assessing the statistical significance of trends in the CPR series will require information on the variance of the estimates of abundances. Appropriate methods may include the use of spatial statistics.

5.4 TIME SERIES MODELS

- Characterisation of the seasonal cycle can be accomplished using spectral and autocorrelation analysis or through a distribution model (e.g. use of normal distribution or distribution mixtures) to describe the seasonal cycle within years.
- Careful evaluation of diagnostics from any model should be made including examination of residuals, influential points and outliers. Models should be revised frequently.
- Interrelationship between the CPR series and other time series of physical and biological data should be explored to assess possible causal linkages. Appropriate methods may include cross-spectral analysis and the development of multiple time series models.
- Alternative models can be composed with bootstrapped estimates of the prediction-squared error.

5.5 INTERPRETATION

- CA should be applied to univariate series, with months as “variables” and years as “observations”, in order to detect multiple underlying modes of variability.
- Where possible the CPR series should be cross-validated with other long-term time series of plankton data (e.g. Helgoland Roads).
- Interrelationship between many biological series may be highly non-linear. Suitable methods may include Generalised Additive Modelling which is well suited for series of the length of the CPR data.
- Alternative methods to CA of constructing summary measures on which to perform accurate time series fitting should be examined, for example the Minimum/Maximum Autocorrelation Factor Analysis (MAFA) technique.

- The extent of redundancy in the information content of different components of the biotic assemblage could usefully be studied, both for subdivisions into rarer/common species and taxonomic or functional groups. Non-parametric multivariate ordination and randomisation tests could be appropriate methodologies.
- There is a need to take account of rarer species on a routine basis, as they can be highly informative.

REFERENCES

- Altman, N (1990) Kernel smoothing with correlated errors. *Journal of the American Statistical Association*, 85, 749-759
- Berryman, D, Bobee, B, Cluis, D and Haemmerli, J. (1988) Nonparametric tests for trend detection in water quality time series. *Water Resources Bulletin*, 24, 545-556
- Brillinger, D.R. (1989) Consistent detection of a monotonic trend superimposed on a stationary time series. *Biometrika*, 76, 23-30
- Bloomfield, P (1970) Spectral analysis with randomly missing observations. *Journal of the Royal Statistical Society*, B32, 369-380.
- Brodeur, R. D., & Ware, D.M. (1992) Long-term variability in zooplankton biomass in the subarctic Pacific Ocean. *Fish Oceanography*, 1, 32-39.
- Chelton, D. B., Bernal, P. A., & McGowan, J.A. (1982) Large-scale interannual physical and chemical and biological interactions in the California Current. *Journal of Marine Research*, 40, 1095-1125.
- Clarke, K.R. (1993). Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* 18, 117-143
- Cleveland, R. B., Cleveland, W. S., McRae, J.E. and Terpenning, I., (1990). STL: a seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6, 3-73.
- Colebrook, J.M. 1975. Automatic data processing methods. *Bulletins of Marine Ecology*, 8, 123-142
- CPR Survey Team. 1992. Continuous Plankton records: The North Sea in the 1980s. *ICES mar. Sci. Symp.* 195:243-248.
- Efron, B. 1983. Estimating the error rate of a prediction rule: improvements on cross-validation. *Journal of the American Statistical Association* 78:316-331.
- Gray, J. S., Pearson, T.H. (1982), Objective selection of sensitive species indicative of pollution-induced change in benthic communities. 1. Comparative methodology. *Mar. Ecol. Prog. Ser.* 9:111-119
- Hastie, T, J., and R.J. Tibshirani. (1990). *Generalized Additive Models*. Chapman and Hall, London.
- Hipel, K.W. (ed) (1988). *Nonparametric approaches to environmental impact assessment*. American Water Resources Association, Bethesda, USA
- Hirsch, R.M. and Slack, J.R. (1984) A nonparametric trend test for seasonal data with serial dependence. *Water Resources Research*, 20, 727-732

IOC Workshop Report No. 124
page 10

Jones, R.H. (1971) Spectrum estimation with missing observations. *Annals of the Inst. Stat. Math.*, 23, 387-398.

Krzanowski, W.J. (1993). Attribute selection in correspondence analysis of incidence matrices. *Applied Statistics* 42, 529-541

Little, R.J.A. and Rubi, D.B. (1987) *Statistical Analysis with Missing Data*. Wiley and Sons, New York.

Neave, H.R. (1970) Spectral analysis of stationary time series using initially scarce data. *Biometrika*, 57,111-122

Oliver, M.A. & Webster, R., (1991). How geostatistics can help you. *Soil Use Manage.*7:206-217.

Parzen, E. (1963) On spectral analysis with missing observations and amplitude modulation. *Sankhya, Series a*, 25, 383-392.

Paczen, E ed. (1984) *Time series analysis of irregularly observed data*. Springer-Verlag, New York.

Radach, G., Berg, J., Hagmeier, E. (1990) Long-term changes of the annual cycles of meteorological, hydrographic, nutrient and phytoplankton time series at Helgoland and at LV ELBE 1 in the German Bight. *Continental Shelf Research*, 10, 305-328.

Sherman, K. (1980). MARMAP, a fisheries ecosystem study in the northwest Atlantic: fluctuations in ichthyoplankton-zooplankton components and their potential for impact on the system. In: *Advanced concepts in ocean measurements*, 9-37, F.P. Deimer (cd), South Carolina Press.

Southward, A.J. (1980) The western English Channel: an inconstant ecosystem. *Nature*, 285, 361-366.

ANNEX I

LIST OF PARTICIPANTS

CANADA

Dr D Sameoto
Biological Sciences Branch
Department of Fisheries and Oceans
Bedford Institute of Oceanography
P O Box 1006,
Dartmouth, Nova Scotia
Tel: (902) 426 3272
Fax: (902) 426 9388

Dr K. R. CLARKE
Plymouth Marine Laboratory
Prospect Place
The Hoe
Plymouth PL1 3DH
Tel: (44 1752) 633 100
Fax: (44 1752) 633 102

Dr J M COLEBROOK
16 Hartley Park Gardens
Plymouth PL3 5HU
Tel: (44 1752) 220 423

FRANCE

Dr F. IBANEZ
Station Zoologique
URA CNRS 716
Boite Postale 28
06230 Villefranche-sur-Mer
Tel: (16) 93 76 38 31
Fax: (16) 93 76 38 34

Dr J C GAMBLE (deseased in 1994)
Sir Alister Hardy Foundation for Ocean Science
c/o Plymouth Marine Laboratory
Prospect Place
The Hoe
Plymouth PL1 3DH
Tel: (44 1752) 633 100
Fax: (44 1752) 633 102

GERMANY

Dr W. GREVE
Biologische Anstalt Helgoland
Zentral Hamburg
Notkestrabe 31
D-2000 Hamburg 52
Tel: (49 40) 89 69 32 10
Fax: (49 40) 89 69 31 15

Dr G.C. HAYS
School of Ocean Sciences
University College of North Wales
Menai Bridge
Gwynedd L159 5EY
Tel: (44 1248) 382 603
Fax: (44 1248) 716 367

UNITED KINGDOM

Dr N. AEBISCHER
Game Conservancy Trust
Fordingbridge
Hampshire SP6 1EF
Tel: (44 1425) 652 381
Fax: (44 1425) 655 848

Dr L. HOWELL (now Dr L. POOLE)
MAFF Fisheries Laboratory,
Lowestoft,
Suffolk NR33 0HT.
Tel: (44 1502) 562 244
Fax: (44 1502) 513 865

Mr M R CARR
Plymouth Marine Laboratory
Prospect Place
The Hoe
Plymouth PL1 3DH
Tel: (44 1752) 633 100
Fax: (44 1752) 633 102

Mr H G HUNT
Sir Alister Hardy Foundation for Ocean Science
c/o Plymouth Marine Laboratory
Prospect Place
The Hoe
Plymouth PL1 3DH
Tel: (44 1752) 633 100
Fax: (44 1752) 633 102

IOC Workshop Report No. 124
Annex I - page 2

Dr E. McKENZIE
University of Strathclyde
Department of Statistics and Modelling Science
Livingstone Tower
26 Richmond Street
Glasgow G1 1XH
Tel: (44 141) 552 4400
Fax: (44 141) 552 2079

Dr M NICHOLSON
MAFF Fisheries Laboratory,
Lowestoft,
SUFFOLK. NR33 0HT,
Tel: (44 1502) 562 244
Fax: (44 1502) 513 865

Dr A TAYLOR
Plymouth Marine Laboratory
Prospect Place
The Hoe
Plymouth PL1 3DH
Tel: (44 1752) 633 100
Fax: (44 1752) 633 102

Dr W R TURRELL
SOAFD Marine Laboratory
PO BOX 101
Victoria Road
Aberdeen AB9 8DB
Tel: (44 1224) 876 544
Fax: (44 1224) 295 511

U.S.A

Dr M. J. FOGARTY
NOAA/NMFS
Northeast Fisheries Centre
166 Water Street
Woods Hole MA, 02543-1097
Tel: (1 508) 548 5123
Fax: (1 508) 457 5124

Dr A. JASSBY
University of California,
Division of Environmental Studies
Davis
California 95616
Tel: (1 510) 652 9287

Dr A SOLOW
Marine Policy Centre
Woods Hole Oceanographic Inst.
Woods Hole
Massachusetts 02543
Tel: (1 508) 457 2181
Fax: (1 508) 282 2535

Dr J R VANDE CASTLE
Long-term Ecological Research Network
College of Forest Resources, AR-10
University of Washington
Seattle, Washington 98195
Tel: (1 206) 543 6249
Fax: (1 206) 543 7295

ANNEX II

PAPERS SUBMITTED AT WORKSHOP

Aebischer, N.J. *The use of transfer-function models to relate temporal variation in a 'response' time series to variation in an 'explanatory' series.*

Castle, J.V. *Integration of spatial and temperal data using GIS.*

Clarke, K. R. *Analysing sparse and high-dimensionsal abundance data.*

Fogarty, M.J. *Some applications of time series analysis to ecological data.*

Greve, W. *The functional significance of specialised predation in planktonic communities.*

Hays, G.C. *Some problems with the CPR data.*

Howell, L., & Nicholson, M, *Treatment of missing data in CPR data.*

Hunt, H.G. *CPR Survey: Plankton analysis, data archiving and retrieval.*

Jassby, A. *Identification of interannual variability in multidimensional time series.*

McKenzie, E. *Non-Gaussian time series modelling*

Planque, B., & Ibanez, F. *Comments on mapping techniques*

Solow, A. *Detecting changes in the composition of a multi-species time series.*

Taylor, A. *Gulf stream position and the plankton of the European shelf.*

The use of transfer-function models to relate temporal variation in a “response” time series to variation in an “explanatory” series

N.J. Aebischer

The Game Conservancy, Fordingbridge, U.K.

Standard regression assumes uncorrelated errors. In the context of time series, the problem is often how to adopt the regression approach while taking into account the correlated error structure; the answer is to use transfer functions (Box & Jenkins 1970). The method is illustrated using an example (Aebischer 1990) drawn from The Game Conservancy's Sussex Database of insects sampled since 1970 in cereal crops in the third week of June, from ca 100 fields per year scattered across 5 farms. The insect of interest was the sawfly (Hymenopteran: Symphyta), whose larvae were an important source of food for gamebird chicks. The amount of undersowing on each farm (where cereal is used as a nurse crop for grass), as well as summer temperature and rainfall, could all potentially explain the temporal and spatial variation in sawfly abundance, based on the insect's biology. Step 1 was to calculate the mean density of sawflies in each year on each farm, and to express it as $\log(n+1)$. Exploratory analysis using cross-correlations confirmed that sawflies showed relationships with all 3 variables, with a 1-year delay; the residuals from a multiple regression of the 3 variables against sawflies were themselves highly correlated with sawfly abundance in the following year, indicative of autoregression.

The transfer function for an input series x_t models the systematic component c_t of a time-series as $c_t = \delta_1 c_{t-1} + \dots + \delta_p c_{t-p} + a_0 x_{t-b} + a_1 x_{t-b-1} + \dots + a_q x_{t-b-q}$, where (in the absence of differencing) p =number of autoregressive parameters, q =number of moving-average-type parameters, b =delay time. In the sawfly example, the pattern of cross-correlations and the short time series (19 values on each of 5 farms) suggested an autoregressive model AR(1), i.e. $c_t = aBx_t / (-\delta B)$ ($b=1$, $p=1$, $q=0$), where $Bz_t = z_{t-1}$, for each of the 3 explanatory variables. As the sawfly series was likely to suffer from measurement error, the error term of the time-series model was modelled as ARMA(1,1), equivalent to an AR(1) error with measurement noise. A further complication was the replication of the series across 5 farms; this was modelled by considering the data as a simple series arranged in the order Farm 1 . . . Farm 5 within each year, and specifying a purely seasonal time-series model with a seasonality of 5; 4 dummy variables coding for the 5 farms were incorporated into the model as simple 1-parameter transfer models to allow for different mean abundances on each farm. Thus the initial model to be fitted was as in Fig. 1. The model was simplified by considering more restrictive nested models and testing for significant differences using maximum likelihood ratio tests. The final model to be chosen was $y_t = B(a_1 x_1 + a_2 x_2 + a_3 x_3) / (1 - \delta B) + a_t$, i.e. there were no significant farm effects, all inputs combined into a single autoregressive component, and noise was purely due to measurement error. All the coefficients a_1 , a_2 , a_3 , δ were significantly different from zero. The residuals from this model were uncorrelated and well behaved with respect to the input variables.

The model was used to assess the impact of large-scale use of a broad-spectrum insecticide on one farm. The observed value on the treated farm was outside the 95% confidence limits of the value predicted on the basis of undersowing and weather for that farm, but the observed value on the untreated area matched the prediction for that area (Fig. 2). By including an intervention variable to represent treatment, the model suggested a recovery time of 4-7 years for the sawflies on the treated area (Fig. 3); the prediction appears to be borne out by more recent data.

References

- Aebischer, N.J. (1990). Assessing pesticide effects on non-target invertebrates using long-term monitoring and time-series modelling. *Functional Ecology* 4:369-373.
- Box, G.E.P. & Jenkins, G.M. (1970). *Time-Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.

Figure 1. Initial time series model fitted to the sawfly data, including 3 lagged exploratory variables, 4 dummy variables coding for farm effects and an ARMA (1,1) error term.

y : annual mean number of sawflies, by farm (log)

x_1 : annual % of fields undersown, by farm

x_2 : annual mean summer temperature

x_3 : annual mean summer rainfall (log)

f_i : farm dummy variables ($i=1 \dots 4$)

Figure 2. Sawfly densities on a farm ²treated with abroad-spectrum insecticide (dimethoate) and on theremaining (untreated) farms: prediction (with 95% confidence limits) from the final time-series model, and observed values.

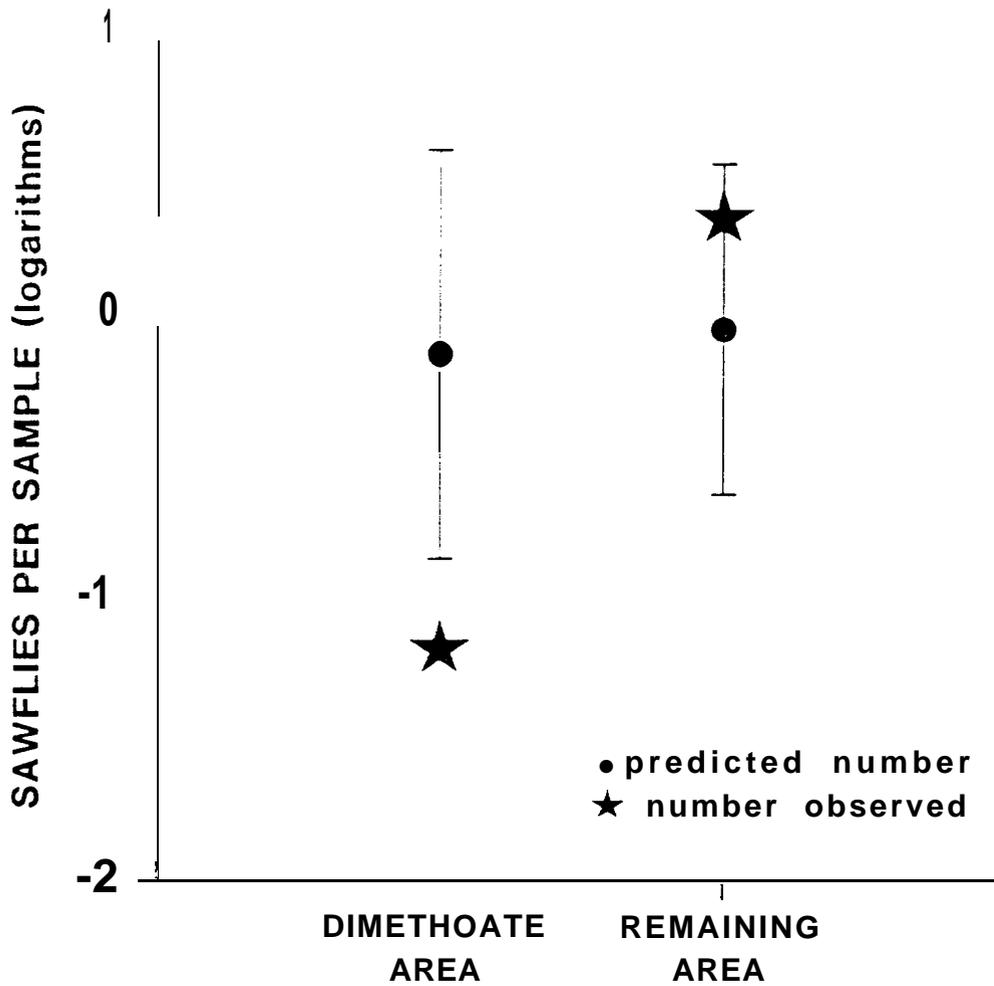
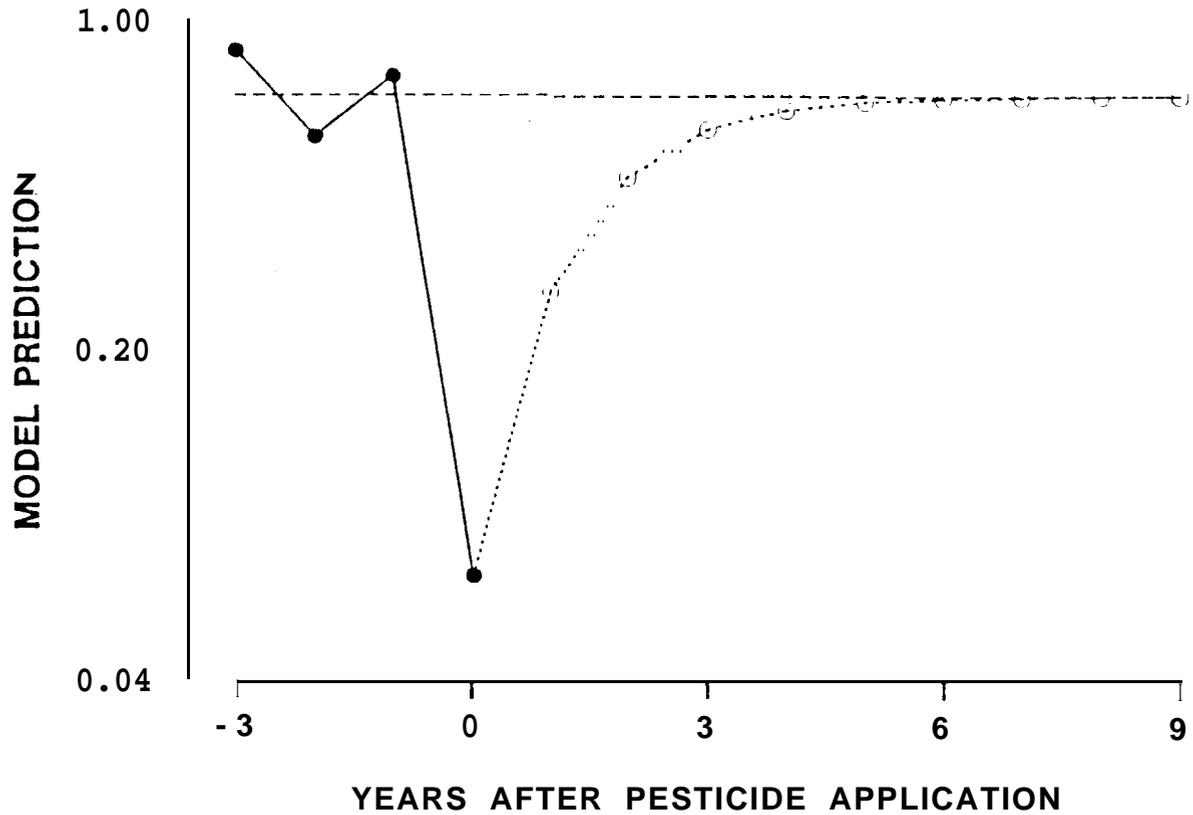


Figure3. Sawfly densities on a farm treated with a broad-spectrum insecticide: before pesticide application (solid line) and after application (dotted line) as predicted by the final time-series model. The horizontal dashed line represents the expected sawfly density on the farm under average climatic conditions and current lack of undersowing.



Integration of Spatial and Temporal Data Using GIS

J.R. Vande Castle

University of Washington, Seattle, U.S.A

The format of CPR data represents one of the few long-term spatial datasets ideally suited for analysis using Geographical Information System (GIS) technology. This technology represents a significant set of tools useful for integration of CPR data in a spatial and temporal format.

In simplified terms, a GIS manages data it contains as a "digital map" or GIS layer. The basic concept of any GIS is the ability to represent information as co-ordinates in a gee-referenced format. That is, each point of information is stored or accessed by its location in space. The actual data, or attributes of each co-ordinate is then stored in a database. In this way, a GIS can be viewed as a front-end tool for access, processing and display of information within a database. The format of the long-term CPR time series data is perfectly suited for GIS applications since the data are stored in raw format with temporal and location information.

Integration of data from many diverse sources is accomplished in a GIS since all data are reference by their location. For instance, plankton counts can be easily displayed or plotted on a map base since information at any particular point is matched by spatial co-ordinates. Temperature information, for instance from satellite data can overlay these data as well, with each point of temperature information can be referenced by its geographic co-ordinates as well. This is an important consideration for integration with future sources of remote sensing information. This capability to integrate data from many sources, is particularly powerful for data integration and modelling.

An important point to remember regarding any GIS, is that many of the internal analysis tools that are used for processing data are no different from more conventional processing techniques. The modelling tools contained within a GIS are, for the most part, standard techniques for statistical description and processing, including such techniques for correlations, maximum likelihood classification, principle component, and spectral analysis.

A significant contribution that GIS technology can offer for CPR data is its capability for display, analysis and output. The display capability of a GIS is flexible to match requirements of many types of analysis. Information can be expressed graphically as a map-based product, or in a more conventional co-ordinate plot.

An important consideration for the CPR survey is implementation of the GIS technology itself. More advanced systems such as commercially available ARC/Info., ER-Mapper or ERDA S/Imagine require fairly extensive resources in computer technology and personnel support. Advanced public domain packages such as GRASS and Khoros are also powerful systems, but in general require more local technical expertise. These systems have advantages in powerful analytic capabilities and graphical output support as well as direct links to attribute databases of Oracle, Ingres and other DBMS packages. These systems are capable of running in current PC technology, but are really designed for use on more powerful workstations. This type of technology would be recommended for extensive processing, but much of the basic data of the CPR program can be processed in a GIS framework requiring less local expertise and computer hardware support. Programs such as IDRISI, and EPPL-7, designed for use on PC's contain many routines that would be useful for modelling and graphical (screen and hardcopy) output. An additional consideration is that routine processing can be automated for producing hard-copy products for data screening on a visual basis as well as for product distribution.

As a final point, many of the routines contained within GIS packages are useful in themselves for visualization purposes. Routines for integration of data, filtering and modelling can be incorporated into almost any data which can be analysed in a matrix format of rows and columns. This includes composite data of location versus time, annual data broken down into rows of months or days etc.

Analysing sparse and high-dimensional abundance data

K.R. Clarke

Plymouth Marine Laboratory, UK

Typical species-by-samples abundance matrices in assemblage surveys, such as the CPR, consist of highly non-normally distributed (and even non-multinomially distributed) variables, with the majority of readings being zero - even for the most abundant 30 to 40 species. Univariate summaries, such as a total abundance across species or a diversity index, can be insensitive to gradients of change in environmental variables (e.g. Warwick & Clarke 1991); multivariate representations are required but need to be tailored to the unorthodox statistical nature of the abundance data. One route finding increasing favour in the marine literature begins by defining, in an ecologically-relevant way, the similarity in assemblage between every pair of samples. For example, all species should be included in the analysis (and there may be several hundred of them) but the joint absence of particular species in two samples should not increase the assessed similarity between those samples (this and other biologically-motivated constraints are widely seen as desirable but not satisfied by traditional correlation-based approaches). The similarity matrix then forms the input to clustering or ordination algorithms, an example of the latter being some form of Multi-Dimensional Scaling (e.g. non-metric MDS, Kruskal & Wish 1978). This results in a 2-dimensional (say) sample plot, in which relative distance between pairs of samples reflects their relative dissimilarity in species composition.

An example for a spatial layout of samples at a fixed point in time is given in Fig. 1. Grab samples of soft sediment from 39 sites around the Ekofisk oilfield (Fig. 1a) were analysed for abundance of benthic macrofauna (209 species observed in total, Gray et al. 1990). For clarity in Fig. 1 the sites are arbitrarily grouped into 4 distance ranges from the oil-field centre. The Shannon diversity calculated for each of these 4 groups (Fig. 1 b) shows that sites less than 250m from the centre have significantly reduced diversity but other groups are not distinguishable with this univariate index. By contrast, the non-metric MDS of the 39 samples (Fig. 1 c), based on Bray-Curtis dissimilarity (Bray & Curtis 1957) from 4th-root transformed abundances, shows a clear pattern of change in the benthic community with distance from the oil-field, even out to a distance of 3km or so. Formal hypothesis tests, for example of differences in community composition of the four groups, can be accomplished by Mantel-type permutation/simulation tests (e.g. Clarke 1993), classical MANOVA tests (e.g. Mardia et al. 1979) being totally invalid because of the highly non-standard form of the data.

A second interesting feature of this data set is that, in spite of the relatively subtle effects which the multivariate approach is sensitive enough to detect, the change is still seen when taxonomic identification is at a higher level than species - for example the plot for family-level information is indistinguishable from that for species (Fig. 1 d), as always seems to be the case with benthic macrofauna (Warwick 1988). More surprisingly, some of the sensitivity remains even at phyletic level; raising questions about "redundancy" of information in the original species/samples matrix, i.e. to what extent there are identifiably different response patterns of species, or groups of species, to the complex of environmental gradients. One means of linking assemblage information to environmental variables, in the context of the similarity and ordination analyses carried out in this example, is given in Clarke & Ainsworth (1993). An alternative approach, starting from the stricter assumptions of multinomial frequencies, and thus employing chi-squared distance as a measure of similarity, leads to correspondence analysis as the ordination technique, and canonical correspondence as the means of linking to environmental variables, again with its own form of robust testing procedures (see the important work of C. ter Braak, e.g. in Jongman et al. 1987).

Clearly, the same multivariate ordination methods can be used to track community change through time rather than spatially. A temporal example is given in Fig. 2. This again concerns benthic macrofauna, though plankton data sets will be structurally similar and amenable to the same methodology (but note that they will generally have stronger seasonal signals). Twenty-one samples taken at approximately 3-monthly intervals from a single site in the Bay of Morlaix (Dauvin 1984) spanned the period of the Amoco-Cadiz oil-tanker wreck, some 50km distant from the sampling location. A similar non-metric MDS analysis summarises, in a graphically convenient way (Fig. 2), the scale of community change coincident with the spill, and the subsequent "recovery" to a more stable state. Note again that changes were relatively subtle, as indicated by univariate measures (species diversity tending, if anything, to increase after the spill).

There are strong limitations here, however, in attempting to meld this highly multi-variable and "ad-hoc" non-parametric approach with "classical" forms of time-series analysis, primarily designed for single variables (or perhaps just 2 or 3 variates), which can be cast into some form of parametric linear model with autoregressive and moving average components. Simple similarity-based permutation tests of, for example, whether changing species patterns are

IOC Workshop Report No. 124
Annex II - page 8

linked to changing environmental variables, could be very misleading in the presence of temporal autocorrelation. In effect, one is doing the equivalent in the univariate case of correlating two time-series without first "pre-whitening" either. It is unrealistic to expect to extend classical Box-Jenkins ARIMA models (Box & Jenkins 1970) to multivariate data matrices which have many more variables (species) than samples, and highly right-skewed observations, the majority of which are zero! Instead, for a data series like the CPR there could be some advantage in attempting to separate out a full time-series analysis of single summary statistics (such as some weighted measure of total zooplankton abundance) from questions concerning the potential "redundancies" in the complete species information. What are the relative contributions of particular species or groups to any inference about the overall pattern? How does the pattern of rare species compare with that for the more dominant ones? How important is the structural (taxonomic) level adopted and how do results compare with functional categorisations, etc? One of the methodological tools likely to be exploited will be the optimal matching of two multivariate descriptions. In the rank dissimilarity-based context underlying non-metric MDS, this problem is discussed in Clarke & Ainsworth (1993). Alternatively, Krzanowski (1993) describes a generalised Procrustes method which could be appropriate in the context of ordination by Principal Components or Correspondence Analysis.

Finally, it is worth pointing out that questions about taxonomic/functional level and the distinctiveness and scale of contributions of rarer species to the overall inference about patterns of change, and their relation to the environment, are very relevant to current debates on biodiversity and ecological complexity, and there is potentially scope here for further interesting analyses of the CPR data.

References

- Box, G.E.P. & Jenkins, G.M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco
- Bray, J.R. & Curtis, J.T. (1957). An ordination of the upland forest communities of Southern Wisconsin. *Ecol. Monogr.* 27:325-349
- Clarke, K.R. (1993). Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* 18:117-143
- Clarke, K.R. & Ainsworth, M. (1993). A method of linking community structure to environmental variables. *Mar. Ecol. Prog. Ser.* 92:205-219
- Dauvin, J.-C. (1984). *Dynamique d'écosystèmes macrobenthiques des fonds sédimentaires de la Baie de Morlaix et leur perturbation par les hydrocarbures de l'Amoco-Cadiz*. Ph. D. thesis, Univ. Pierre et Marie Curie
- Gray, J. S., Clarke, K. R., Warwick, R.M. & Hobbs, G. (1990). Detection of initial effects of pollution on marine benthos: an example from the Ekofisk and Eldfisk oilfields, North Sea. *Mar. Ecol. Prog. Ser.* 66:285-299
- Jongman, R. H. G., ter Braak, C.F.J. & van Tongeren, O.F.R. (1987). *Data analysis in community and landscape ecology*. Pudoc, Wageningen
- Kruskal, J.B. & Wish, M. (1978). *Multidimensional Scaling*. Sage Publications, Beverly Hills
- Krzanowski, W.J. (1993). Attribute selection in correspondence analysis of incidence matrices. *Appl. Statist.* 42:529-541
- Mardia, K. V., Kent, J.T. & Bibby, J.M. (1979). *Multivariate Analysis*. Academic Press, London
- Warwick, R.M. (1988). The level of taxonomic discrimination required to detect pollution effects on marine benthic communities. *Mar. Pollut. Bull.* 19:259-268
- Warwick, R.M. & Clarke, K.R. (1991). A comparison of methods for analysing changes in benthic community structure. *J. mar. biol. Ass. U.K.* 71:225-244

Figure 1. Ekofisk oil-field macrobenthos study (Gray et al. 1990). a) Map of the 39 sampling sites in relation to the centre of drilling activity. b) Shannon species diversity for the groups of sites in each of the 4 distance ranges (means and 95% confidence intervals, treating the sites in each range as "replicates" in a 1-way ANOVA). c) Non-metric MDS of the 39 sites, based on species-level abundances. d) MDS from family-level data.

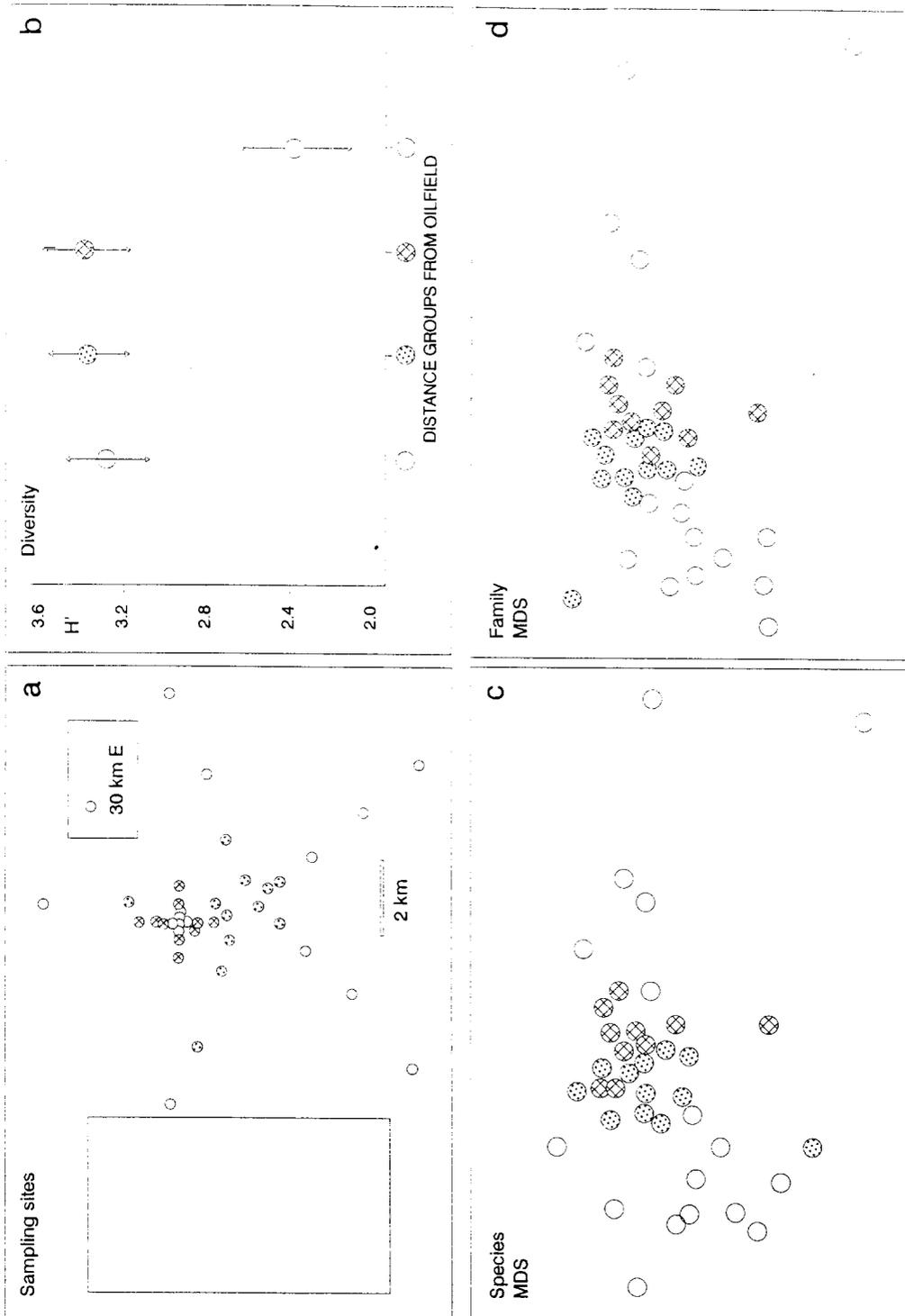
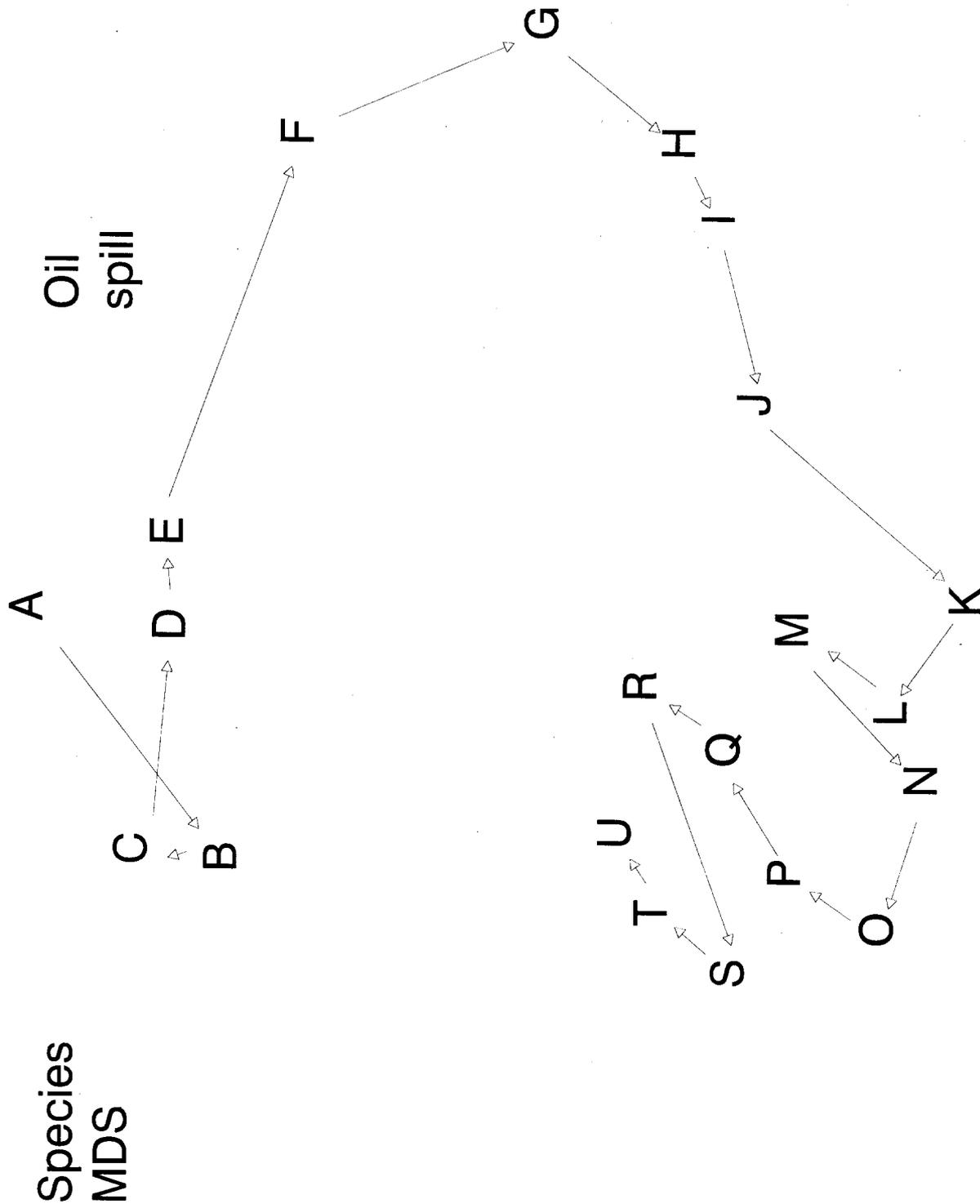


Figure 2. Macrobenthos communities (species-level data) at Station 'Pierre Noire' in the Bay of Morlaix (Dauvin 1984). Non-metric MDS for approximately quarterly sampling occasions (A to U) between April 1977 and Feb 1982, spanning the period of the wreck of the 'Amoco-Cadiz'.



Some applications of time series analysis to ecological data

M.J. Fogarty

NOAA, Woodshole, U.S.A.

Examples are provided of applications of time series models, including Autoregressive - Integrated Moving Average Models (ARIMA) for univariate time series and transfer function models for multivariable series. These models defined in the time domain generally require data series collected at equally spaced points in time (or summarised at equal intervals). The data must also be stationary in level and in variance. The residuals (errors) also the final model are required to be normally distributed with mean zero and constant variance.

The strategy of model building follows the Box-Jenkins approach for tentative process of (1) model identification based on diagnostic patterns in the auto correlations and partial auto correlations for univariate time series (ARIMA) models and the cross correlation function for multiple time series (transfer function models), (2) estimation of model parameters and (3) diagnostic checking in which the residuals of the model are tested for auto correlation, homogeneity of variance and normality. If the residuals do not meet the assumptions the identification and estimation phases are repeated. The most parsimonious model description consistent with the data and meeting all assumptions is selected.

Univariate models are derived for a monthly time series of landings for an American lobster population along the coast of Maine USA to illustrate the procedures used to identify an appropriate model structure, followed by estimation and diagnostic checking. Here, an autoregressive model was developed with coefficients for the effect of the previous months landings and the landings 12 months previous. One year of data was reserved and not used for fitting the model and subsequently used to test the model prediction. The 12 step ahead forecast from the model differed from the checking data set by a mean absolute deviation of less than 5%. The observed and estimated values for the series are illustrated in Figure 1.

An example of an application of univariate time series models to develop an abundance index for research vessel trawl survey data is also presented. A model corresponding to exponential smoothing is developed to provide an abundance index for two stocks of yellow tail flounder off the north eastern United States. An example for one stock is provided in Figure 2.

Two examples of transfer function models are presented next. The first describes a model relating annual landings of American lobster in Maine to annual mean sea water temperature for the period 1945-86. Following an identification phase which included filtering or pre-whitening of the landings series by the model for the temperature series, a transfer function model was defined which incorporated temperature at a lag of 6 years and in the current year (lag 0). The immediate temperature effect is interpreted as a reflection of a temperature mediated influence on activity levels, capability and growth. The lagged effect may represent the effect of temperatures during the early life stages resulting in increased survival. An additional transfer function was developed for stock-recruitment data for silver hake off the north eastern United States which included the role of spawning stock size or recruitment and random environmental effects represented through a moving average component in the model.

Figure 1. The observed and estimated values for the series

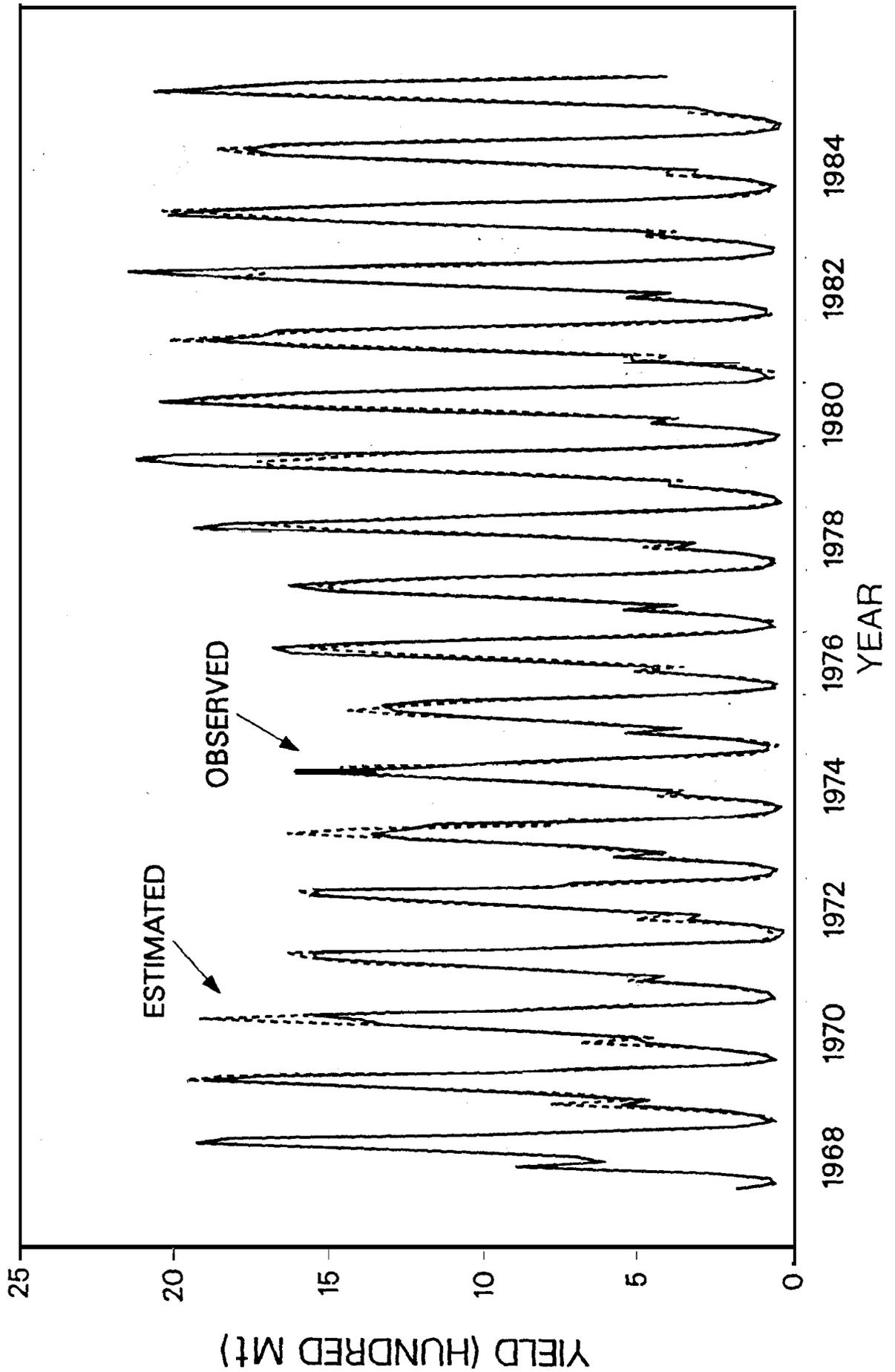
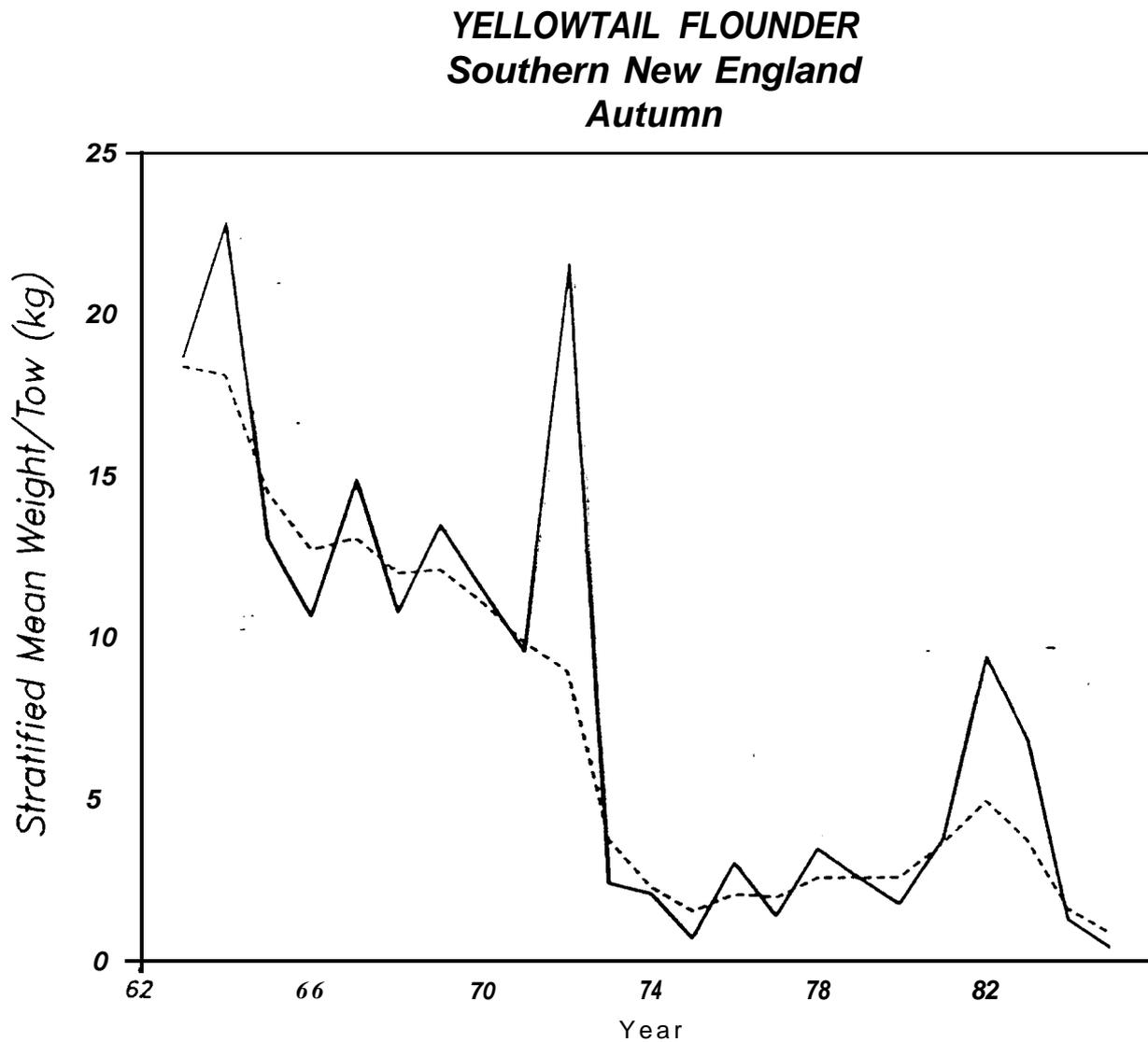


Figure 2. An example for Yellowtail Flounder



The functional significance of specialised predation in planktonic communities

W. Greve

Biologische Anstalt Helgoland, Hamburg, Germany

The "Helgoland Roads" zooplankton time-series was started in 1974 with the intention to study the systems ecology beyond the phytoplankton and nutrient measurements which have been obtained since 1962. While these are made daily, mesozooplankton (> 100µm) and macrozooplankton (>500 µm) were sampled every second workday. Since this year every fortnight a megazooplankton (> 1 cm) sample is 'taken to obtain information on scyphomedusae.

The almost complete coverage of the pelagic ecosystem permits the investigation of population processes from annual dynamics to multiannual comparisons to obtain trend-information. As to the investigation of the systems ecology the availability of the interacting populations is especially valuable in the light of recent findings (Huntley and Lopez, 1992, Sinclair 1988) and others who see the importance of predator control within the marine pelagic ecosystem to be more efficient than the food limitations in the population dynamics of copepods, at least. This result is supported during most of the year in the Helgoland data.

As to the significance of the specialized predation, examples are given of the year 1989 when the annual dynamics of small copepods was determined by a very early and low abundance of *Pleurobrachia pileus*, an undefined but probably efficient predation of *Alaurina composita*, an uninvestigated turbellarian which shows a distinct population growth in the last years, and the immigrating siphonophore *Muggiaea atlantica* that changed the traditional ecological equilibrium to a degree that could be traced from the predator via the copepods, the phytoplankton to the disturbance of the PO₄ remineralization disturbance.

The "Helgoland Roads" time-series is being used for a causal analysis of the trophic linkages within the ecosystem as a decisive element in the determination of the single population success.

References

- Greve, W., & Reiners. 1988. Plankton time-space dynamics in the German Bight-a systems approach. *Oecologia* 77: 487-496.
- Huntley & Lopez. 1992. Temperature dependent production of marine copepods: a global synthesis. *The American Naturalist*, 140,2: 201-242.
- Sinclair, M. 1988. Marine Populations. University of Washington Press. Seattle: 1-252.

Figure 1. Small copepod measurements 1975 to 1990

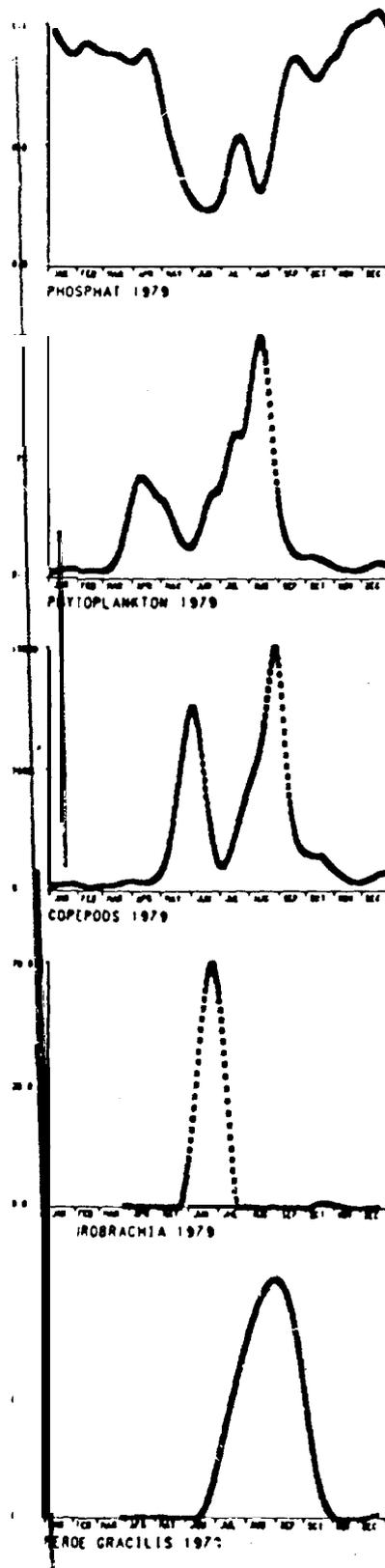
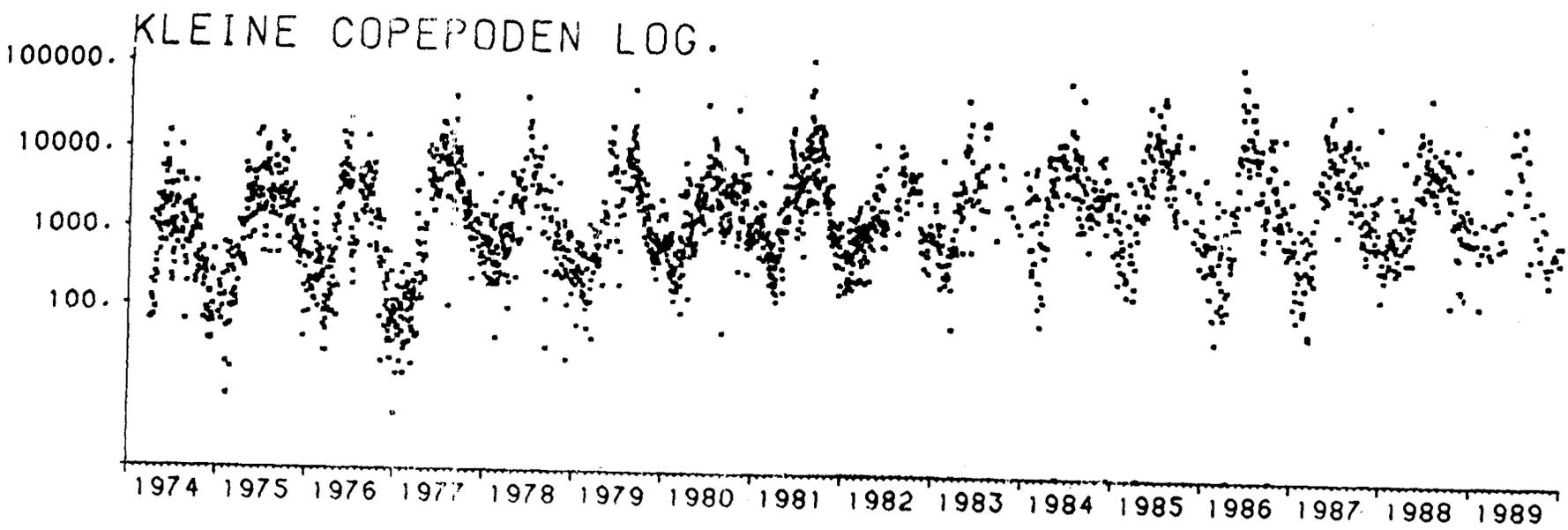


Figure 2. SPLINE representations of key components of the German Bight pelagic ecosystem in 1979



IOC Workshop Report No. 124
Annex II - page 18

Some problems with the CPR data

G. C. Hays

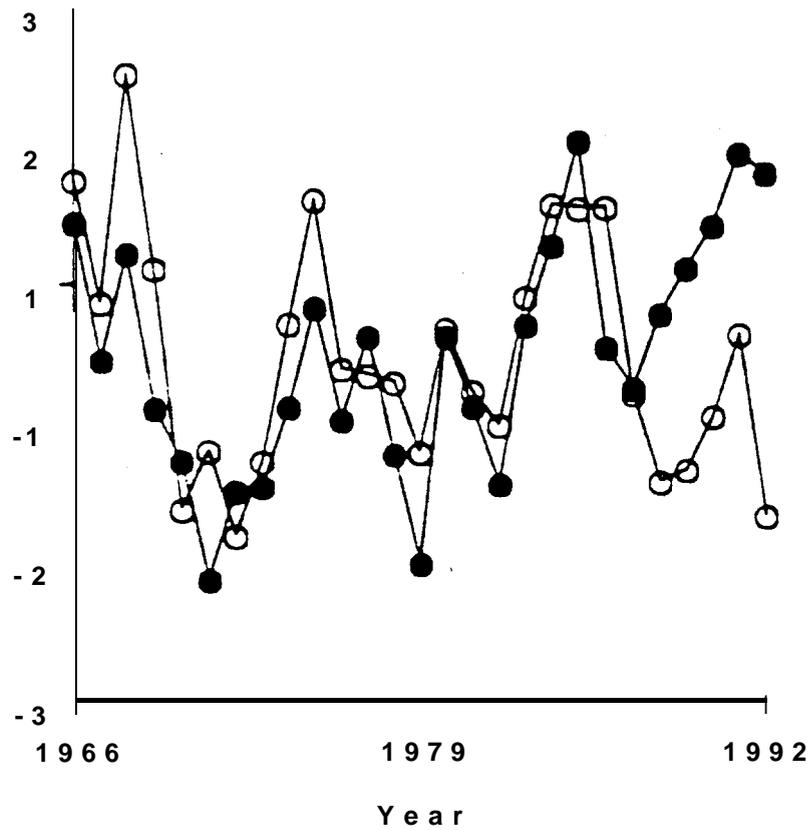
University College of North Wales, U.K.

The sampling constraints imposed by the CPR survey being operated from “ships of opportunity” include:

1. Changes in tow routes.
2. Changes in sampling intensity.
3. Changes in tow speed.

Changing tow routes may introduce spurious time-series signals. A technique for avoiding or at least reducing this problem is to define areas for analysis on the basis of spatial homogeneity of plankton abundance. When such an area was defined to the NW of the UK a significant relationship was found between mean annual plankton abundance and the mean annual Gulf Stream position (1966- 1991). The plankton abundances for the last 4 years of the series (1988- 1991) were however significantly less than that predicted on the basis of the Gulf Stream position (Figure 1). Changes in sampling intensity have been profound within the series. In general there has been a marked reduction in sampling intensity in the last 30 years. Since plankton abundance values are not normally distributed, but instead are highly skewed to the left, there will be a greater probability of underestimating plankton abundance when fewer samples are taken. A “bootstrapping” method was used to show that the anomalous annual abundance values were not the consequence of poor sampling. Finally the mean annual tow speed was shown to have remained constant in this area throughout the series and so was not the consequence of the 4 anomalous points.

Figure 1. The annual abundance of copepods (O---O) and the mean annual Gulf Stream position (.-.), both normalized to zero mean and unit variance. For the entire 27 years of the time series, these two parameters were significantly positively correlated.



Treatment of Missing Data in CPR Data

L. Howell and M. Nicholson

MAFF Fisheries Laboratory, Lowestoft, U. K.

Summary

This paper looks at methods for estimating annual means from an incomplete month-by-year matrix of plankton data. Four different methods are presented, and the different assumptions underlying them are discussed. The methods are demonstrated using month-by-year mean \log_{10} (abundances) of *Calanus*, stages V-VI, from area B 1 in the North Sea.

Introduction

Knowing the annual abundance of plankton is important for a number of environmental applications. For example, plankton abundance is thought to be a good indicator of climate change, (Dickson et al., 1992). Also Cushing (1984) noticed a correlation between cod recruitment and the abundance of *Calanus*.

For the European shelf, the Continuous Plankton Recorder (CPR) survey provides a long time series on plankton abundance and species composition (The CPR Survey Team, 1992). One of the objectives of the survey is to provide descriptions of year to year changes in plankton abundance (Colebrook, 1975). The data within a month are aggregated (Colebrook, 1975), and the yearly signal constructed from the sum of the monthly abundance indices. Hence the annual abundance of plankton in year y , P_y , is

$$P_y = \sum_{m=1}^{12} \mu_{ym}$$

estimated by

$$\hat{P}_y = \sum_{m=1}^{12} p_{ym}$$

where μ_{ym} is the true, and p_{ym} is the observed plankton abundance in the m 'th month and y 'th year.

This procedure will be satisfactory if there are data in every month. In the historical record of data collected by the CPR survey this requirement is not always satisfied. Here, we examine and compare various methods for estimating P_y from incomplete data by inferring the abundances in the unobserved months from the observed months.

The methods will be demonstrated using month-by-year mean \log_{10} (abundances) of the zooplankton *Calanus*, stages V-VI, from area B1 in the North Sea (Figure 1) for the period 1949 to 1991. The data are tabulated in Table 1.

Methods of Analysis

We will consider four methods. The first has been used in previous analyses of these data, and is described in Colebrook (1975). The second and third assume an underlying fictional form to describe m_{ym} . The fourth method uses a method described by Mandel (1993) which assumes a linear model to estimate the missing cells. We now describe each method in detail.

Method 1: Ratio Estimator.

This method is described by Colebrook (1975). The technique consists of multiplying the long-term monthly mean corresponding to a missing month in a given year by the ratio of the sum of the non-missing values in that year to the

sum of the corresponding long-term monthly means.

Writing

$$d_{ym} = 0 \text{ if } p_{ym} \text{ is missing} \\ = 1 \text{ else,}$$

then

where

$$P_{(\cdot)m} = \frac{\sum_{y=1}^Y P_{ym} \delta_{ym}}{\sum_{y=1}^Y \delta_{ym}}$$

$$P_{y(\cdot)} = \sum_{m=1}^{12} P_{ym} \delta_{ym}$$

and

$$P_{(\cdot)} = \sum_{m=1}^{12} P_{(\cdot)m} \delta_{ym} \cdot$$

The brackets around the subscripts signify that summation is over non-missing data. Note that $P_{(\cdot)}$ will be different for every m and y . Table 2 gives the estimates obtained for 1958 and 1991, each having one missing month.

Method 2: Fitting an underlying Normal Distribution function.

If the shape of the production curve within a year follows a Normal distribution function centred on t with duration determined by the scale parameters, we have

then we can write

$$\log[\mu_{ym}] = \beta_0 + \beta_1 m + \beta_2 m^2$$

where

$$\beta_0 = \log\left(\frac{P_y}{\sqrt{2\pi\sigma^2}}\right) - \frac{\tau^2}{2\sigma^2}$$

$$\beta_1 = \frac{\tau}{\sigma^2}$$

and

$$\beta_2 = -\frac{1}{2\sigma^2}.$$

where $b_2 \leq 0$ to satisfy $s^2 \geq 0$

Therefore, assuming

$$\log(p_{ym}) = \log[\mu_{ym}] + \varepsilon_{ym}$$

with

$$\varepsilon_{ym} \approx N(0, \sigma_e^2)$$

then a quadratic regression of $\log(p_{ym})$ on m will yield estimates of t , s^2 and P_y and hence m_{ym} from the identities

$$\tau = -\frac{\beta_1}{2\beta_2}$$

$$\sigma^2 = -\frac{1}{2\beta_2}$$

and

$$P_y = \sqrt{-\frac{\pi}{\beta_2}} e^{\beta_0 - \frac{\beta_1^2}{4\beta_2}}$$

This model is fitted separately to each year, provided there are data for at least four months.

Figures 1a and 1b show this model fitted to the data from 1958 and 1991 respectively. The estimated values are given in Table 2.

Method 3: Normal Distribution Function With Constant Variance.

Method 2 can be modified to make use of any information common to all years. For example, it may be possible to assume that the duration of plankton production, measured by s , is constant. Then, writing

$$\mu_{ym} = P_y \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(m-\tau_y)^2}{\sigma^2}}$$

where P_y and τ_y are different in each year and s^2 is now common to all years, we have

$$\log[\mu_{ym}] = \beta_{0y} + \beta_{1y}m + \beta_2 m^2$$

where

$$\beta_{0y} = \log\left(\frac{P_y}{\sqrt{2\pi\sigma^2}}\right) - \frac{\tau_y^2}{2\sigma^2}$$

$$\beta_{1y} = \frac{\tau_y}{\sigma^2}$$

and

$$\beta_2 = -\frac{1}{2\sigma^2}$$

where $b_2 \leq 0$.

Then, as before, a quadratic regression of $\log(p_{ym})$ on m , pooled over all years, will yield estimates of t_y , s^2 and P_y from the identities:

$$\tau_y = -\frac{\beta_{1y}}{2\beta_2}$$

$$\sigma^2 = -\frac{1}{2\beta_2}$$

and

$$P_y = \sqrt{-\frac{\pi}{\beta_2}} e^{\beta_{0y} - \frac{\beta_{1y}^2}{4\beta_2}}.$$

Figures 2a,b show the result of fitting this model to the data for 1958 and 1991. The fit for 1991 looks more sensible than the fit obtained by method 2. The estimated values are given in Table 2.

Method 4: Row Linear Model.

In an analysis of variance context, we can write

$$\mu_{ym} = \mu + Year_y + Month_m + (YearMonth)_{ym}.$$

Clearly this model cannot be fitted without replication within each month and year. However, Mandel (1961) suggested a model with limited interaction, of the form

$$\mu_{ym} = \mu + Year_y + Month_m + \lambda\alpha_y Month_m$$

which can be fitted without replication. This can be written as

$$\mu_{ym} = \mu + Year_y + \beta_y Month_m.$$

Mandel extended this model to deal with missing cells (Mandel, 1993) by exploiting the regression of P_{ym} on $Month_m$ for the non missing cells in year y , and $Month_m$ is estimated from those years with no missing data.

Figures 3a,b show the regression data for 1958 and 1991. The estimated values are given in Table 2

Comparison of Methods Using Cross Validation.

We compare the methods on the basis of their mean squared prediction error (MSPE), which we estimate using cross validation. In cross validation, each observation in a data set is omitted in turn. The model is fitted to the remaining observations, providing an estimate of the omitted value. The estimated MSPE is calculated from

$$MSPE = \frac{1}{MY} \sum_{y=1}^Y \sum_{m=1}^M (p_{ym} - p_{(ym)})^2$$

where $p_{(ym)}$ is the predicted value of p_{ym} when p_{ym} is omitted.

The comparisons were made using the data from the sixteen years which were complete. The MSPE was estimated under four scenarios of missing data:

1. One month missing.
 2. The first three months of each year simultaneously missing
 3. The fifth, sixth and seventh months of each year simultaneously missing
- and
4. Six months missing at random. This was simulated one hundred times for each year. The MSPE for all the simulations was then calculated for each method,

Scenario 3 where data are missing at the peak of plankton production might be expected to be more damaging (have larger MSPE) than Scenario 2.

Table 3 gives the estimated MSPEs for each Scenario for each method.

Discussion

On the basis of the MSPE, methods 1 and 3 were the weakest models. Method 3 did particularly badly on scenario 2 where the first three months were missing.

The performance of the other three methods was similar. Method 2 was slightly superior to the others. On the scale of \log_{10} numbers of *Calanus*, the MSPE for methods 2,4 and 5 is about 20%. Each method has its own set of assumptions, advantages and disadvantages. For example, Method 2 only exploits data from one year, both an advantage and disadvantage. Method 4 cannot be used unless there is a subset of complete years.

References

Colebrook, J.M. (1975) The continuous plankton recorder survey: automatic data processing methods. *Bull. mar. Ecol.* 8, 123-142.

Cushing, D.H. 1984. The gadoid outburst in the North Sea. *J. Cons. int. Explor. Mer.* 39:159-166.

Dickson, R. R., Colebrook, J.M. and Svendsen E. 1992. Recent changes in the summer plankton of the North Sea. *ICES mar. Sci. Symp.*, 195:232-242.

Mandel, J. (1961) Non-additivity in two-way analysis of variance. *J. Am. Statist. Ass.*, 56:878-888.

Mandel, J. (1993) The analysis of two-way tables with missing values. *Appl. Statist.* 42:85-93

The CPR Survey Team (1992) Continuous plankton records: the North Sea in the 1980s. *ICES mar. Sci. Symp.* 195 243-248.

Table 2 Predicted values of \log_{10} Calanus.

	Method 1	Method 2	Method 3	Method 4
1958 (October)	1.37	0.93	1.08	1.38
1991 (March)	0.94	0.70	1.09	0.92

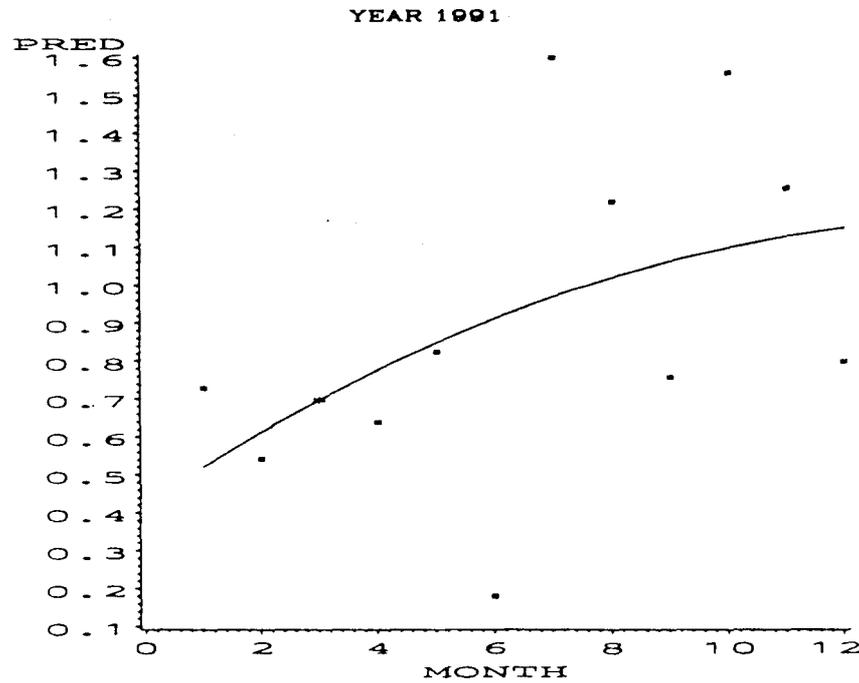
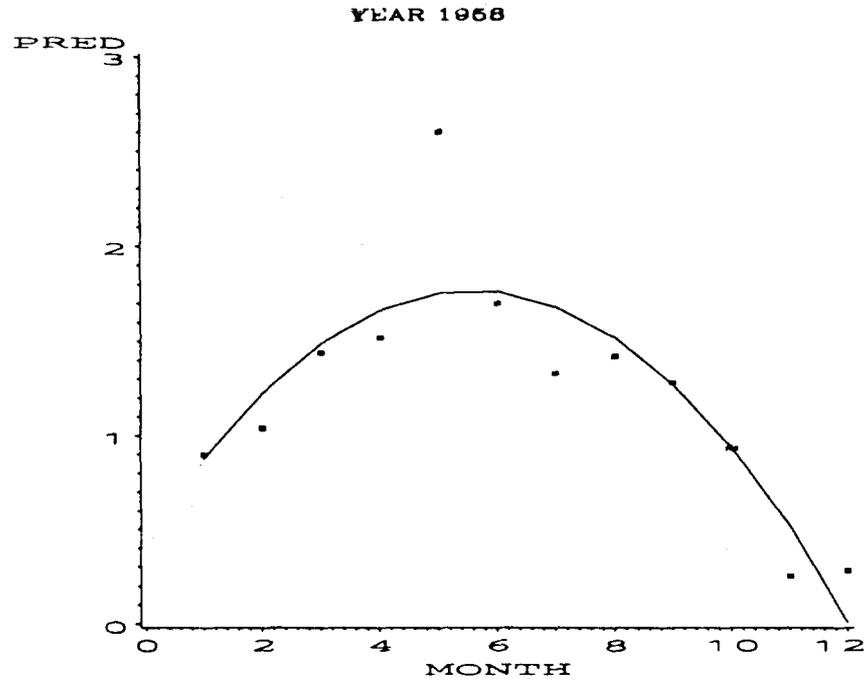
Table 3 Estimated MSPES of \log_{10} Calanus.

	Method 1	Method 2	Method 3	Method 4
Scenario 1	0.198	0.278	0.215	0.238
Scenario 2	0.208	0.657	0.204	0.246
Scenario 3	0.260	0.325	0.280	0.333
Scenario 4	0.210	0.471	0.246	0.274

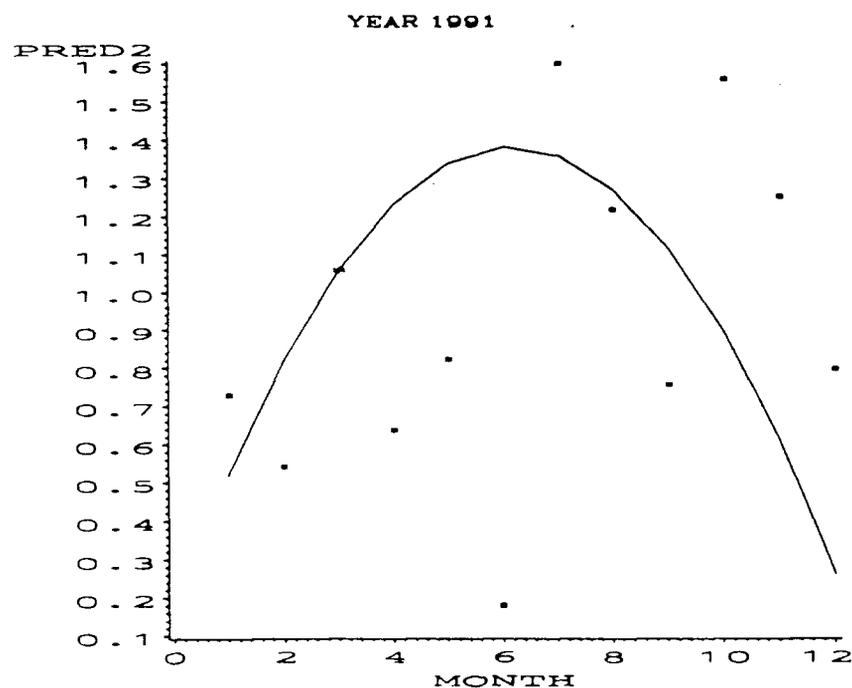
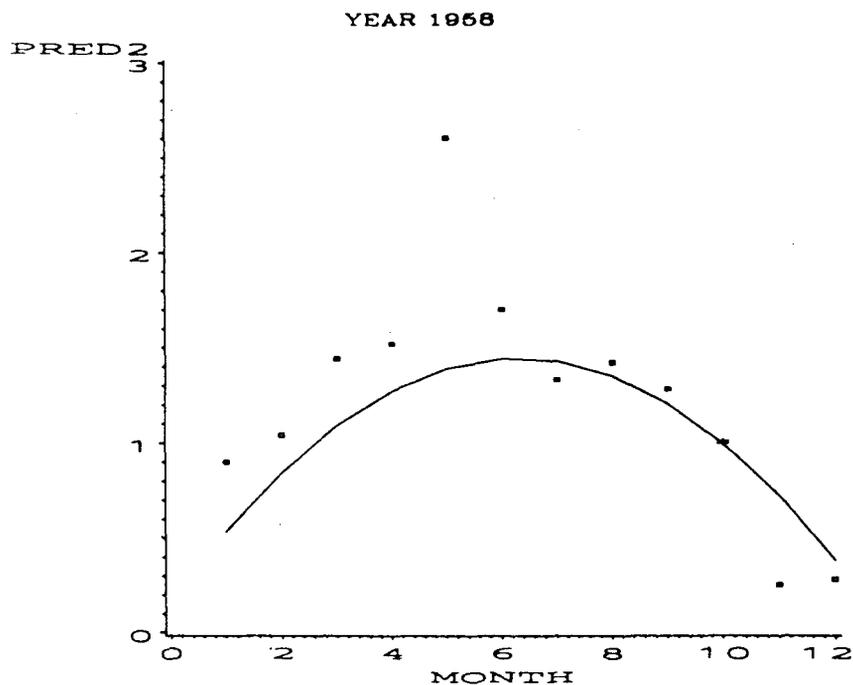
Table 1: Month-by-year mean densities of *Calanus*, stages V-VI, from the North Sea for the period 1949 to 1991.

	month											
	1	2	3	4	5	6	7	8	9	10	11	12
1949	0.3		1.0	1.9	1.3				1.6			0.1
1950	0.5	1.1	0.7	2.3	2.2	1.5	1.5	0.6		1.1	0.3	
1951		0.7		1.8	2.1	1.7	1.2	1.1	1.7	0.9	0.5	0.4
1952	0.9	0.4	1.3	2.6	1.5	2.0	0.8	0.6	1.2	0.9	1.1	
1953	1.1	0.7	1.2	1.4	1.6	1.1	1.5	1.0	1.0	1.1	0.8	0.3
1954		0.3	0.8		0.0		1.3	0.6		0.6	0.4	0.4
1955	0.4	0.8	1.0	1.5	2.2	1.4		0.8		1.0	0.5	0.2
1956	0.5	0.5		2.2	0.0	1.3	1.9	1.0		0.8	0.3	
1957	0.5	1.6	1.2	1.7	1.5	2.4	0.1	0.5	0.8	1.1	0.9	1.3
1958	0.9	1.0	1.4	1.5	2.6	1.7	1.3	1.4	1.3		0.3	0.3
1959	1.0		1.5	1.6	2.1	1.2	1.3	1.0	0.4	0.6	0.4	0.2
1960	0.3		1.7	2.6	0.5		1.9			1.7	0.9	
1961	0.8	1.4		1.5	2.7	1.0		0.5	0.9	1.2	0.8	
1962	0.6	0.3		1.4	2.1	1.4	1.3	0.5	0.7	0.9	1.0	0.2
1963	0.5	0.0	0.3	1.5	2.3	1.1	1.6	1.7	1.2	1.3	1.2	0.4
1964	0.9	1.6	0.0	2.1	1.1	2.2	1.9	0.4	1.4	1.0	1.3	0.3
1965	1.1	1.1	1.0	1.2	2.0	2.3	1.0	1.8	1.5	1.1	0.9	0.4
1966	0.6	1.4	1.5	1.3	2.2	1.5	1.3	1.4		1.6	1.1	
1967	0.8	0.0	0.9	1.1	2.1	0.9	1.0	1.1	1.1	1.2	0.2	0.0
1968	0.7	1.2	0.9	0.9	1.6	1.8	1.0	0.5		1.9	0.5	0.8
1969	0.8	0.5	0.8	1.0	2.2	2.3	1.5	1.1	1.8	0.5	0.5	0.0
1970	0.3	0.5	1.7	1.9	1.8	2.4	1.6	1.1	1.2	1.1	0.8	0.1
1971	0.7	0.9	1.0	1.1	2.1	1.2	1.5	0.8	0.5	0.9	0.2	0.2
1972	0.0	0.9	0.7	1.6		1.6	1.7	1.0	0.8	1.0	0.5	0.1
1973		0.2	1.0	1.1	1.9	2.2	1.9	0.6	1.2	1.7	1.0	0.0
1974	0.1	0.2	0.7	0.5	1.5			1.0	1.5	0.4	1.1	0.4
1975	0.1	1.1	1.0	1.8		2.3		2.7	0.7	0.9	0.4	0.0
1976	0.5	1.0	1.6	0.7	2.2	2.1	2.3	1.1	1.5	1.2	0.0	1.1
1977	0.1		1.0	1.3	1.2	2.1		1.0	1.2	1.7	1.0	0.1
1978	0.1	0.5	1.6	1.0	2.2	1.3	1.7	0.9	1.0	1.9	0.7	0.3
1979	0.2	1.8	1.6	0.6	2.0	1.2	1.4	1.6	1.3	1.4	0.5	0.2
1980	0.3	0.7	1.2	0.7	1.3	2.0	1.3	0.7	1.0	0.7	0.3	0.3
1981	0.2	0.6	1.1	0.9	1.1	0.6	0.7	0.7	1.5	1.1	0.4	0.1
1982	0.1	0.8	0.3	1.1	0.7	1.2	1.3	0.8	1.5	1.6	0.4	0.2
1983	0.2	0.3	0.7	1.4	1.5	1.6	1.4	1.1	1.4	1.4	0.8	0.5
1984	0.2	0.2	1.1	0.8	2.3	1.9	1.0		1.5	1.0	0.7	0.8
1985	0.3	0.8	1.2	2.1	1.7	0.9	0.7	0.1	1.5	1.3	0.7	0.7
1986	0.0		1.1	0.3	2.0	0.6	1.2	0.7	0.0	1.1	0.7	0.8
1987	0.3	0.9	1.0	0.3	2.6	0.8	0.1	1.3	1.0	0.8		
1988	0.4		0.9		1.6		1.4	1.0	1.3	0.8	0.6	0.5
1989	0.4	1.1				0.4	0.2		0.6	0.6	0.9	1.2
1990	0.7	0.9			1.1	0.8		0.9	0.9	0.9	0.4	
1991	0.7	0.5		0.6	0.8	0.2	1.6	1.2	0.8	1.6	1.3	0.8

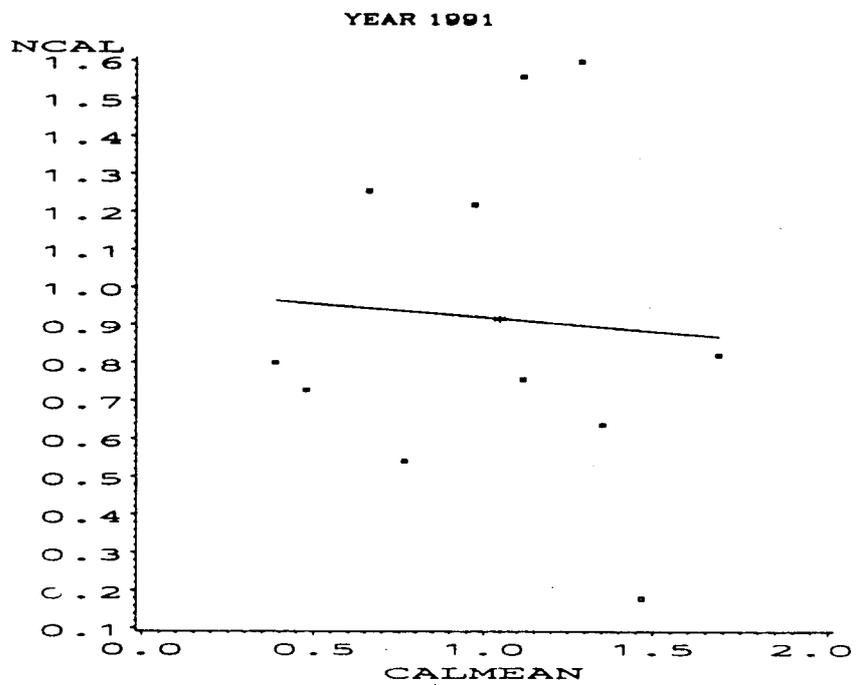
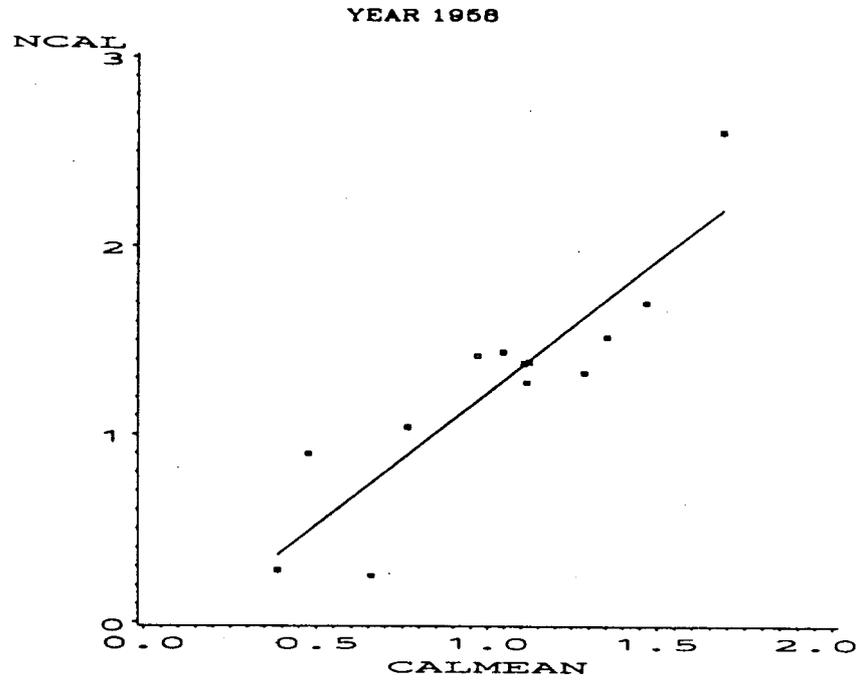
Figures 1a and 1b: Normal distribution function (method 2) fitted to data from 1958 and 1991 respectively.



Figures 2a and 2b: Normal distribution function with constant variance (method 3) fitted to data from 1958 and 1991 respectively.



Figures 3a and 3b: Regression data for 1958 and 1991



IOC Workshop Report No. 124
Annex II - page 30

CPR Survey: plankton analysis, data archiving and retrieval

H.G. Hunt

Sir Alister Hardy Foundation for Ocean Science, Plymouth, U.K.

When the Continuous Plankton Recorder is towed at a depth of 10m, water enters through a 1.27cm square aperture in the nosecone and is slowed to about a thirtieth of its original speed by the enlargement of the tunnel to a size of 5.08cm by 10.16cm at the point of filtration. It is then filtered through a slowly moving band of bolting silk, of 23.6meshes to the cm, before being covered by a second band to form a "sandwich" which is wound onto a spool in a tank into which a dilute solution of formalin diffuses to fix the plankton. The silk transport is driven by a propellor at the rear of the machine as it is being towed through the water. For the CPR Survey it has been arranged such that 10.16cm of silk passes the filtering point, and 3 cubic metres of water is filtered, for every 10 miles of tow. The machine is loaded with silk marked and numbered at 5.08cm intervals so that when it is returned after towing, the amount of silk used can be calculated.

The crew of the towing ship returns a completed log form from which a clean set of navigational data is entered to the computer and run through a tow processing program. This program produces three pages of hardcopy output and a file of data.

Page one contains a printout of the input data together with the mileage towed, the mileage sampled, the miles per division of marked silk (M.P.D) and the number of divisions per 10-mile sample (D. P. B).

The D.P.B. is used to produce the second page of output which contains the set of cutting points for the tow. This is given to the person who is to cut up the silk into the 10-mile samples for distribution to the plankton analysts.

Page three of the output contains the set of sampling information for the samples to be analysed. The information consists of the tow identification data and, for each sample, the sample id., the latitude and longitude of the centre point of the 10-mile sample, the coordinates of the sample from a set of 10 latitude x 2° latitude "standard rectangles, an indication as to whether the centre-point of the sample was sampled during the day or night and the local time the sample was taken.

The file output from the program contains a full set of sampling information for the tow ready for input to the sample archive. This consists of, for each sample, the allocated month and year (used for certain internal CPR Survey processing functions), the tow id., the validity of the tow, the sample id., the latitude in decimal minutes (north negative), the longitude in decimal minutes (west negative), the local time of sampling and the day, month and year of the sample.

Note that because a continuous sampling method is used by the CPR, of the plankton carried on each 10-mile sample silk, 75% comes from the 10-mile section of tow to which the sample is assigned with 12.5% coming from each of the preceding and succeeding five miles of tow.

	Phytoplankton	Zooplankton
Species	119	118
Genus	42	63
Family	1	6
Sub-order	1	3
Order	-	16
Class	2	14
Phylum	-	6
Total	165	226

Table 1. Summary of taxonomic entities identified in the CPR Survey

Of these some 28 phytoplankton and 35 zooplankton taxa are either widely or locally abundant. Of the rest, some are of low abundance but present all the time, whereas others are but rarely caught. Some are only analysed as present or absent, e.g. *Phaeocystis* spp., a nuisance phytoplankton which disintegrates on hitting the silk and covers it with a slimy coating of mucilage.

The analysis of the plankton is carried out in 4 stages:

1. Colour analysis before the silk is cut into samples. This consists of matching the colour on the silk with prepared colour cards reflecting a dilution scale of acetone extractions of the green chlorophylls.
2. Phytoplankton analysis -20 x 295µm diameter fields along the two diagonals of the filtering silk (0.0001 of the sample).
3. Zooplankton analysis - a zig-zag track, 2.06mm wide, across both the sampling and covering silk (0.2 of the sample).
4. Zooplankton whole sample analysis for the larger organisms

The data is entered to analysis sheets and when the tow is complete a check on the analysis is carried out by the most experienced analyst in the team as to whether any re-analysis has to be carried out.

At the completion of a month's analysis, the sampling and plankton data is run with a suite of programs to give sampling tables, tables of occurrence of rare taxa, a tile of log standard area data for the 68 most abundant taxa and a file of plankton data for input to the data archive.

The traditional main data archive is file-based with four files per year:

1. A file containing the navigational data and the length of silk filter mesh used for each tow. This is the information from which the sampling information is obtained.
2. A file containing the sampling information for each tow detailing position (latitude, longitude and allocated "standard rectangle"), local time, day or night, date and analyst id. of whoever analysed the sample.
3. A file containing the plankton information for the tows analysed in January to June.
4. A tile containing the plankton information for the tows analysed in July to December.

Each tow within the plankton files has a limited set of the sampling attributes; the rectangle coordinates and night/day allocation. The plankton information is held as coded counts

Data for individual or grouped taxonomic entities can be retrieved in one of three ways:

IOC Workshop Report No. 124
Annex II - page 32

1. Based on 1° latitude by 2° longitude "standard rectangles". The main use of the data retrieved in this way is for producing distribution charts for biogeographical studies.
2. Based on defined groups of the standard rectangles into "standard areas" and used for the analysis of large-scale variation in time and space. Such data has been used recently in the interpretation of possible effects of climatic/hydrographic interactions on plankton populations.
3. Based on any defined set of convex polygons (in practice, usually rectangles) which are most suitable for fine resolution analysis in restricted areas or along transects.

The data archive has now been input to an Oracle Relational Data Base System, held on a Sun workstation and accessible from PCs on the Plymouth Marine Laboratory Local Area Network.

The database, CPRSBASE, consists of a set of relations, or tables, maintained by the Data manager of SAHFOS.

Tables 2 to 8 give a description of the tables which currently comprise CPRSBASE.

ANALYST	
START_YEAR	The year that the analyst first started analysing samples.
INTER_STOP_YEAR1	The year that the analyst stopped analysing samples for the first time.
INTER_START_YEAR2	The year that the analyst started analysing samples for the second time.
INTER_STOP_YEAR2	The year that the analyst stopped analysing samples for the second time.
INTER_START_YEAR3	The year that the analyst started analysing samples for the third time.
STOP_YEAR	The year that the analyst finally stopped analysing samples.
ANALYST_ID	Unique analyst number
ANALYST_NAME	Analyst's full name

Table 2

TREATMENT	
TREATMENT_ID	Unique number identifying data processing treatment
PLANKTON_GROUP	(C)olour index, (P)hytoplankton or (Z)ooplankton
PROPORTION OF SAMPLE	Proportion of sample analysed as a subsample
TREATMENT_SAMPLES	Day, Night or All samples to be used in the data processing treatment
TREATMENT_GROUP	Name of the group identified by the treatment id.

Table 3.

PLANKTON_CATEGORY	
PLANKTON_GROUP	(C)olour index, (P)hytoplankton or (Z)ooplankton
TAXON_CATEGORY	Code for the counting category
CATEGORY_MINIMUM	Lower bound for the counting category
CATEGORY_MAXIMUM	Upper bound for the counting category
CATEGORY_MEAN	Accepted mean value for the category

Table 4.

PLANKTON_INFO	
TOW_NAME	Alphanumeric tow identifier
SAMPLE_ID	Sample number along the tow
TAXON_ID	Taxon identifier
TAXON_CATEGORY	Code for the counting category

SAMPLE INFO	
TOW_NAME	Alphanumeric tow identifier
SAMPLE_ID	Sample number along the tow
SAMPLE_LATITUDE	Latitude in minutes of the sample (North is negative)
SAMPLE_LONGITUDE	Longitude in minutes of the sample (West is negative)
SAMPLE_DATE	Local time and date of the sample
ANALYST_ID	Numeric identifier of the analyst who analysed the sample

Table 5.

TOW EVENTS	
TOW_NAME	Unique alphanumeric identifier of tow
EVENT_CODE	Code for the events shooting and hauling the CPR and altered courses along the tow
EVENT_LATITUDE	Latitude in minutes for each event (North is negative)
EVENT_LONGITUDE	Longitude in minutes for each event (West is negative)
EVENT_DATE	Time (GMT) and date for each event
SILK_READING	Silk readings for the events hauling and shooting the CPR

Table 6.

TOW_INFO	
TOW_MONTH_SERIAL_NUMBER	Number of the tow within a month
CPR_MONTH	Month allocated to the tow
CPR_YEAR	Year allocated to the tow
TOW_NAME	Unique alphanumeric tow identifier
SHIP_NAME	Name of ship making the tow
RECORDER_ID	Number of the CPR used for the tow
INSIDE_MECHANISM	Number of the inside mechanism used in the CPR used for the tow
PROP_SETTING	Propellor setting of the CPR used for the tow
TOW_VALIDITY	Validity of the tow (e.g. OK)
SHIP_SPEED	Average speed of the ship over the tow
MILES_PER_DIVISION	Ratio of miles per 2in section of silk sampled over the tow
TOW MILEAGE	Miles towed over the tow
SAMPLED MILEAGE	Miles sampled over the tow

Table 7.

TOW_MESSAGES	
TOW_NAME	Alphanumeric tow identifier
COMMENT_1	7-character comment from the record information file
COMMENT_2	12-character comment from the record information file
MESSAGE	80-character message from the record information file

TAXON_CATALOGUE	
TAXON_ID	Unique taxon numeric identifier
PRINT_ORDER	Print order of taxon within the CPR data processing system
TREATMENT_ID	Number identifying the data processing treatment for the taxon
TAXON_LABEL	Unique 8-character label for the taxon
CHECKNAME	Unique 15-character shortened name of the taxon
TAXON_NAME	Unique name of the taxon

Table 8.

Identification of interannual variability in multidimensional time series

A. Jassby

Division of Environmental Studies, University of California, U.S.A.

Chlorophyll *a* distributions in the San Francisco Bay-Delta Estuary were used to illustrate the investigation of mechanisms underlying year-to-year variability. Over the past two decades, fish and invertebrate populations at several trophic levels have undergone large year-to-year changes, including severe declines in some cases. Hydrodynamic variability is thought to lie at the heart of these changes, particularly variability in the net flow of water from the landward to the seaward portion (*Q*_{out}) and the export of water from the landward portion to state and federal water projects (*Q*_{exp}). The actual mechanisms are diverse, but a link through the food supply to the base of the food web is thought to play a role. The mechanisms underlying chlorophyll variability are therefore of interest.

Chlorophyll has been measured in the landward portion (Suisun Bay plus the Delta) since 1968 at approximately 30 stations on a monthly basis. A first step in the analysis is to aggregate the data into space-time bins for each year in some consistent fashion. The exact choice of bins will reflect the questions being addressed, the length of the observations, and various constraints imposed by the nature of the ecosystem. In this case, data was binned on the basis of season (winter, spring, summer, autumn) and specific conductivity (<0.25, 0.25-2, 2-10, >10 mS/cm). The estuarine turbidity maximum (ETM) is typically found from 2-10 mS/cm (1-6 ppt salinity).

Are there any underlying patterns or modes of variability that together constitute the interannual variability represented by the bins? The 16 bins can be thought of as a multivariate time series, in which case PCA provides a natural method for uncovering underlying modes. First, the eigenvalues of the covariance matrix were tested for significance with a Monte Carlo test (Preisendorfer's Rule N). On this basis, the first two principal components (PCs), accounting for 75% of the overall variability, were retained for rotation. The PROMAX rotation (*k*=2) generally performs best in uncovering the modes for geophysical fields.

The first mode, itself accounting for 50% of the total, was due primarily to variability in the summer ETM (Fig. 1). The amplitude time series (scores), i.e., the intensity of the mode from year to year, suggested a long-term decline, but the temporal behaviour could not be substantiated because of the small number of observations. Data were not available for all bins during the more extreme years; during the 1982-83 ENSO, for example, the landward estuary was largely fresh water for a long period. As a result, the PCA was carried out with only 12 observations. In order to investigate the causal mechanisms behind this mode, more observations are needed. Because the summer ETM does not contribute to variability in the second mode, the summer ETM series itself was used as a surrogate for the first PC.

The relation between summer ETM as the response, and both *Q*_{out} and *Q*_{exp} as predictors, was investigated with the Additivity and Variance Stabilization (AVAS) algorithm. AVAS determines smooth transformations for the response and predictors such that the relation is additive and the variance of the transformed response is independent of response size. In our example, a logarithmic-like transform for the response, a negative linear effect for *Q*_{exp}, and a unimodal response for *Q*_{out} with a peak at 500 m³/s were determined (Fig. 2). The results make extremely good physical sense. When *Q*_{out}=500, the ETM is positioned in Suisun Bay, a subembayment with broad shoals and therefore a favorable light regime for phytoplankton. Upstream and downstream of Suisun Bay, phytoplankton in the ETM are confined to narrow, deep turbid channels where primary productivity is often negative. The negative effect of *Q*_{exp} represents the diversion of phytoplankton out of the estuary upstream of the ETM, suggesting an important contribution of riverine phytoplankton to the ETM. This has since been confirmed using material balances. The second PC, which is not discussed here, represents spring blooms in the upper estuary and can be tied to hydraulic residence time.

To summarize, the data are first binned in a way that reflects the question, the amount of data, and constraints imposed by the ecosystem. PCA is then applied to the bins as variables, the significant PCs are identified through some means, and a PROMAX (*k*=2) rotation is applied. The response of the PC amplitudes to forcing variables of interest are then explored through the use of nonparametric additive models, such as AVAS or generalised additive models. These models can then be parametrized in the form of classical linear or generalised linear models for inference purposes. The program of analysis outlined here, when used in an exploratory sense, 'has proved to be versatile and well-adapted to short time series.

IOC Workshop Report No. 124
Annex II - page 38

Bibliography

Hastie, T., and R. Tibshirani. 1991. Generalized additive models. Chapman and Hall.

Jassby, A. D., C.R. Goldman, and T.M. Powell. 1990. Interannual fluctuations in primary production: direct physical effects and the trophic cascade at Castle Lake, California. *Limnol. Oceanogr.* 35:1021-1038.

Jassby, A.D., C.R. Goldman, and T.M. Powell. 1992. Trend, seasonality, cycle, and irregular fluctuations in primary productivity at Lake Tahoe, California, Nevada, USA. *Hydrobiologia* 246:195-203.

Figure 1. Variability in the summer ETM

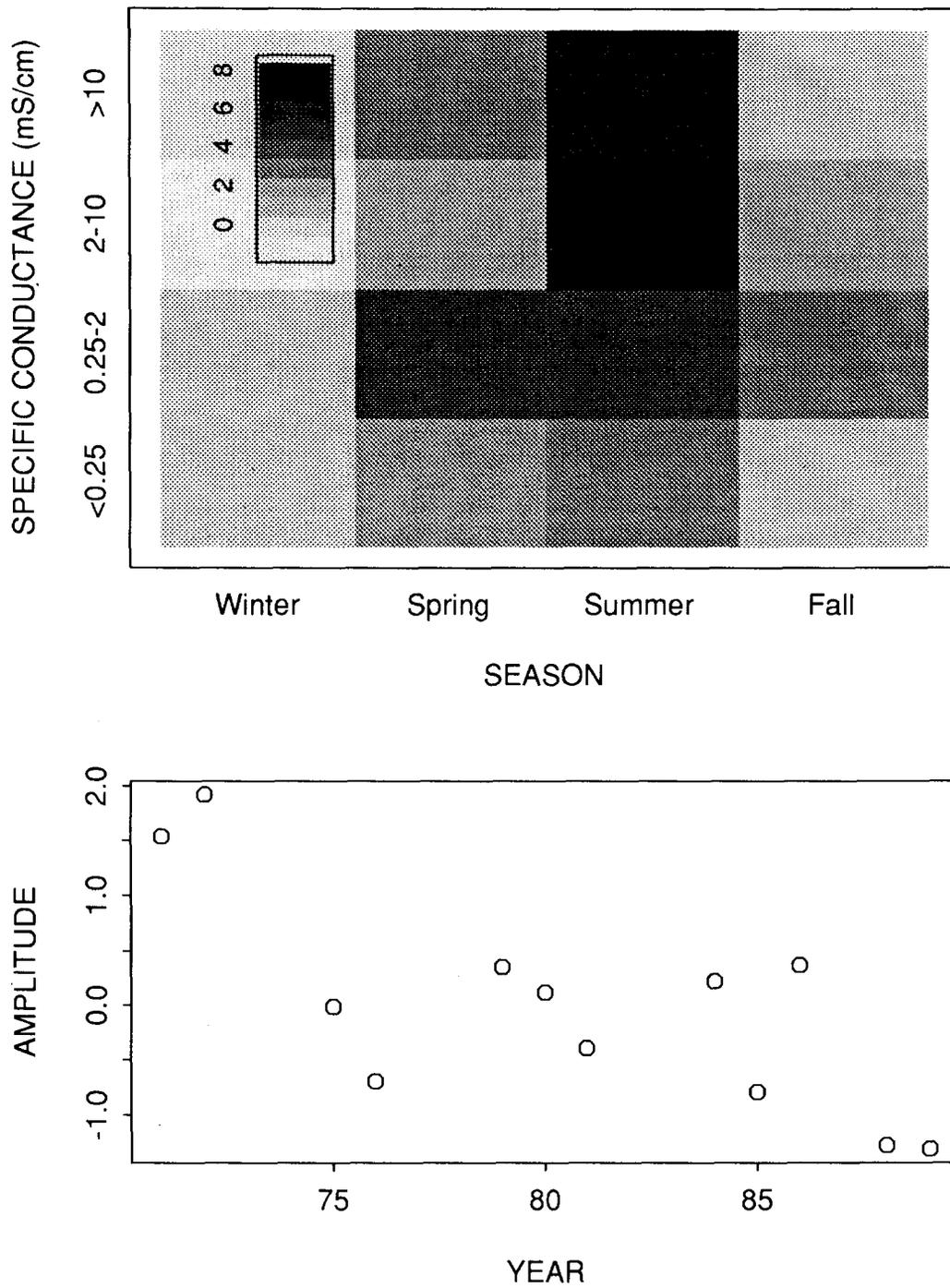
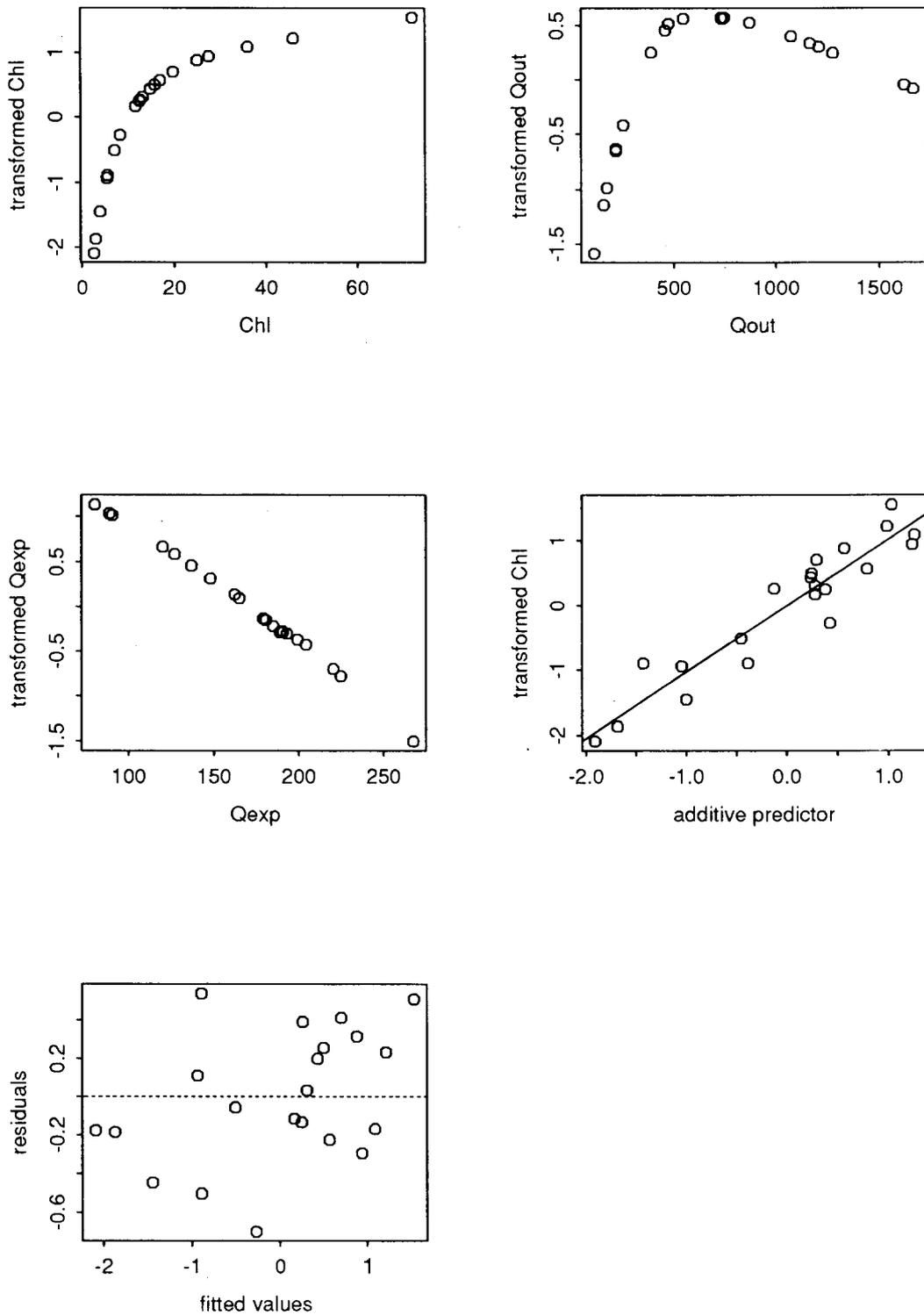


Figure 2. Additivity and Variance Stabilization (AVAS) algorithm: ChI vs. Qout, Qexp.



Non-Gaussian time series modelling

E. McKenzie

University of Strathclyde, U. K

My work in non-gaussian time series in the last 10 years has been in three main areas:-

- 1). development of new models with specific marginal distributions and correlation structures for use in simulation;
- 2) devising procedures for assessing the independence of residuals in fitted models, a necessary requirement in simulation. and
- 3) modelling actual non-gaussian time series,

An example of (1) is a simple model for a process which exhibits some of the characteristics of rainfall time series, viz sequences of dependent positive observations interrupted by sequences of zeros. In the basic model the positive series was AR(1) in form and the run lengths of the strings of zeros were geometric. However, this can be generalized in many ways and some were indicated.

As an example of (3) the modelling of a binary time series which was one of a large number of daily menstrual diaries in a large WHO data set from an intimation study on contraceptives. The probability of changing from one state to the other on any day was modelled explicitly as a function of current state, how long the woman had been in that state and the product of these. The modelling of binary sequences is relevant in the CPR series.

Two series from the CPR survey were modelled;: 34 years (1958-91) of monthly data on *Ceratum fusus* in two areas (B 1 and B2) of the North Sea. The data contain just less than 50% zeros and B 1 has 10% missing values, B2 somewhat less. proposed model has the form: $Y_t = B_t X_t$ where $\{B_t\}$ is a binary process with $P(B_t=1) = P_t = 1 - P(B_t=0)$, and $\{X_t\}$ is a process of positive values. In essence, $\{X_t\}$ is a measure of the abundance of the organisation in that area, but, for various reasons (including pancity) may not be recorded in the sampling that month (i.e. $B_t=0$). It is assumed that $\{B_t\}$ and $\{X_t\}$ can be modelled separately at least initially. Logistic regression was used to model $\{B_t\}$ and in both areas a strong seasonality and dependence on the previous observation were modelled. The final model in B 1 was:

$$\ln\left(\frac{P_t}{1 - P_t}\right) = -0.19 - 1.43 \cos \omega t - 1.90 \sin \omega t - 0.19 \cos 4\omega t + 0.51 \sin 4\omega t + 0.56 Y_{t-1}$$

$$\text{where } \omega = \frac{2\pi}{12}$$

Note that P_t increases (decreases) with Y_{t-1}

IOC Workshop Report No. 124
Annex II - page 42

Modelling the positive series $\{X_t\}$ was tried in two ways. First, with X_t having a Gamma marginal distribution whose mean μ_t satisfied $\ln(\mu_t) = \beta^T Z_t$, where Z_t are known regressors reflecting seasonality, etc. Fitted models were found to be poor representatives of some of the long term trend movements, especially in the late 80's. Including a factor for each year resolved this difficulty but left a model which might have good explanatory properties but did have poor predictive ones. Also, the fit to the Gamma was not very convincing.

The second approach is via transformations of X_t . Since variation clearly increases with level an obvious and easy starting point is logarithmic transformation: $W_t = \ln(X_t)$. It is clear from the plots that the logged series exhibits trending and changes of level and a very strong seasonal pattern. A model which has proven useful in such circumstances is the ARIMA (0, 1, 1) (0, 1, 1)₁₂ which has the form:

$$\nabla \nabla_{12} W_t = (1 - \theta_1 B)(1 - \theta_2 B^{12}) a_t$$

where $\{a_t\}$ is white noise, $\Delta W_t = W_t - W_{t-1}$ and $\nabla_{12} W_t = W_t - W_{t-12}$. The estimation routine did not converge satisfactorily and yielded the estimates

$$\hat{\theta}_1 = 0.86 \text{ and } \hat{\theta}_2 = 0.992$$

It is the proximity of $\hat{\theta}_2$ to its upper boundary (of 1) which causes problems. It strongly suggests the seasonal differencing is 'too much' in that it has removed a deterministic trend. The residuals suggest the model is close to correct: final model for B was:

$$W_t = -0.21 - 0.51 \cos \omega t - 0.66 \sin \omega t - 0.34 \cos 2\omega t + 0.22 \sin 2\omega t + Z_t$$

$$\text{where } \nabla Z_t = (1 - 0.87 B) a_t$$

$$\text{and } \hat{\sigma}_a^2 = 1.003$$

Note model form: $W_t = \ln X_t = f(t) + Z_t$

$f(t)$: provides seasonal pattern within years; need to fit it as deterministic form suggests it is a very stable one.

Z_t : non-stationary process, contributes level-changing effect.

Also effects are multiplicative: $X_t = e^{W_t} = e^{f(t)} \cdot e^{Z_t}$; so if Z_t causes a rise (say) in level, amplitudes of the sinusoid change accordingly.

The residuals suggest an adequate fit by usual tests i.e. correlation, magnitude and pattern. Also, quartile plot strongly supports normality of the residuals.

Area B² has similar analysis and alignments but now ARIMA (0, 1,1) (0, 1,1) ₁₂ seems to fit well -we get;

$$\nabla \nabla_{12} W_t = (1 - 0.74B)(1 - 0.91 B^{12}) a_t$$

and

$\hat{\sigma}_a^2 = 1.029$, and again residuals suggest this is a good model.

The cross correlation function of the two sets of residuals (i.e. from B1 and B2) was estimated. It reveals significant values at lags 0 and 1, i.e. measures in B 1 and B2 in same month are correlated and measures in B 1 in current month are correlated with those which occurred in B2 six months earlier. It could be used to start building a bivariate model for this species in these two areas.

Figure 1. Area B1 : Logged Series and Smoothed Trend

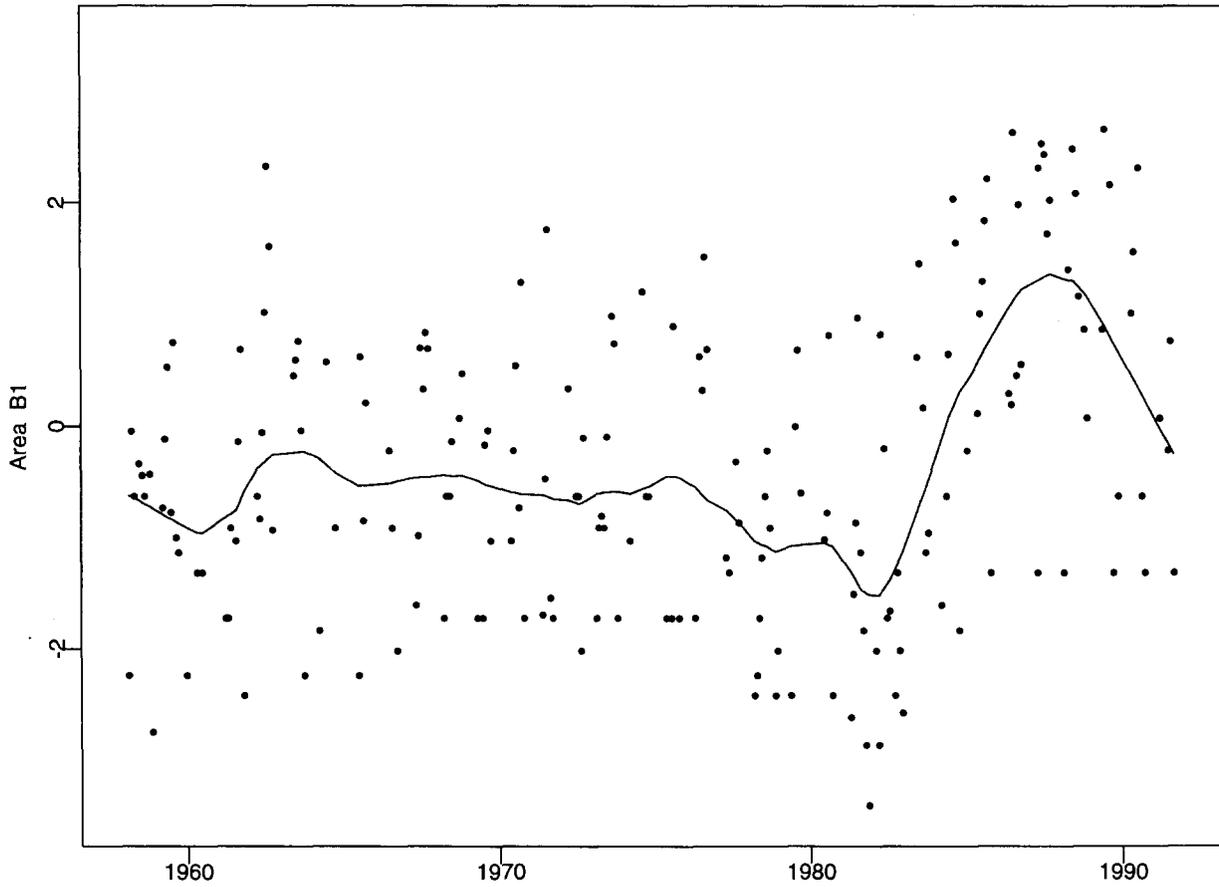


Figure 2. Area B2 : Logged Series and Smoothed Trend

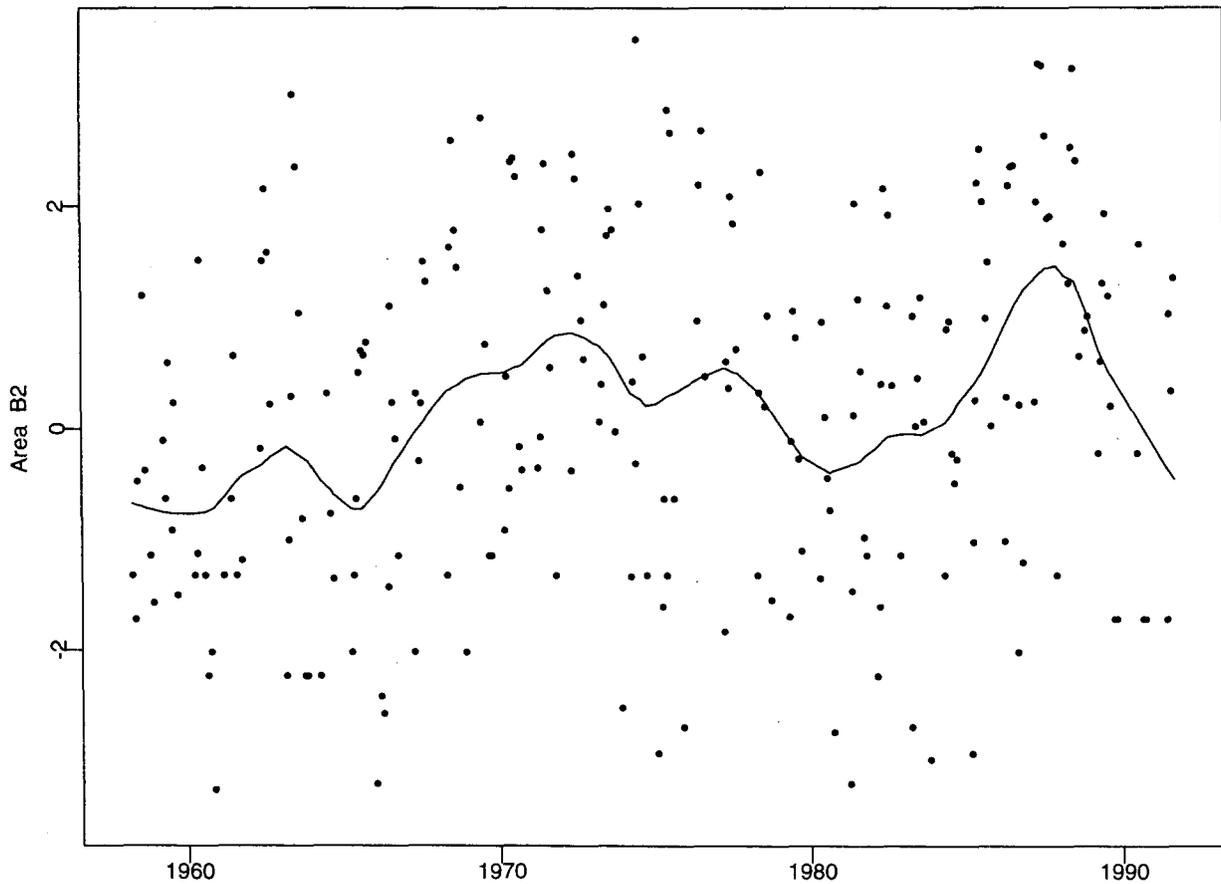


Figure 3. Area B1 : Standardized Residuals from Model

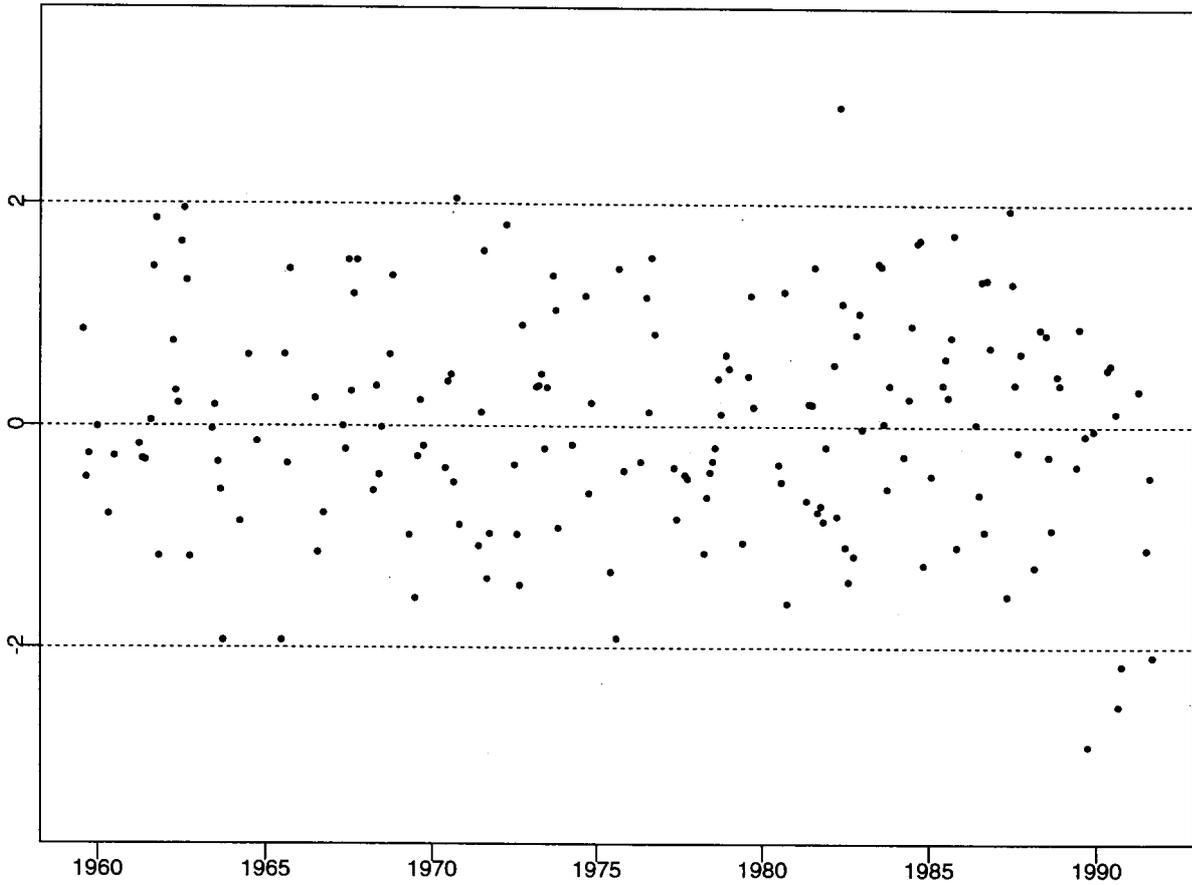


Figure 4. Area B2 : Standardized Residuals from Model

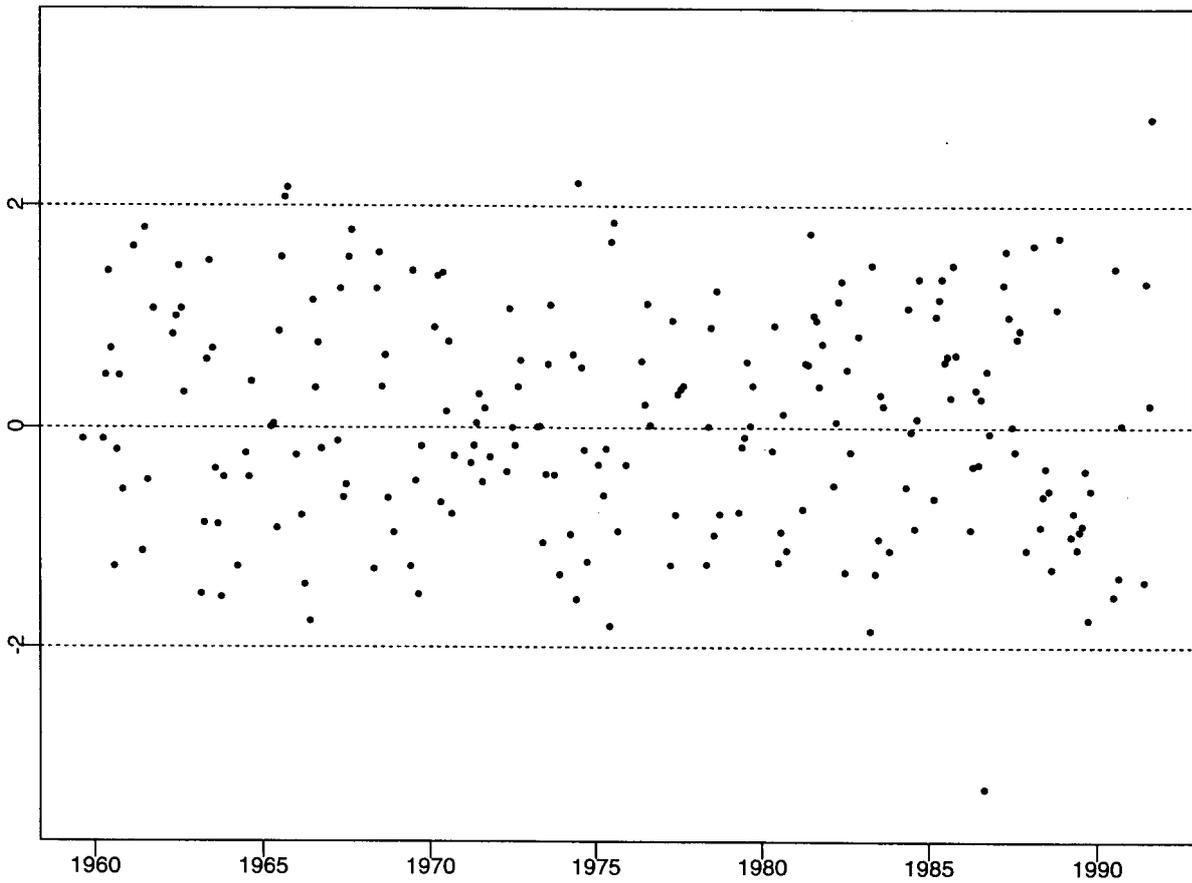


Figure 5. Area B1 : ACRF of Standardized Residuals from Model

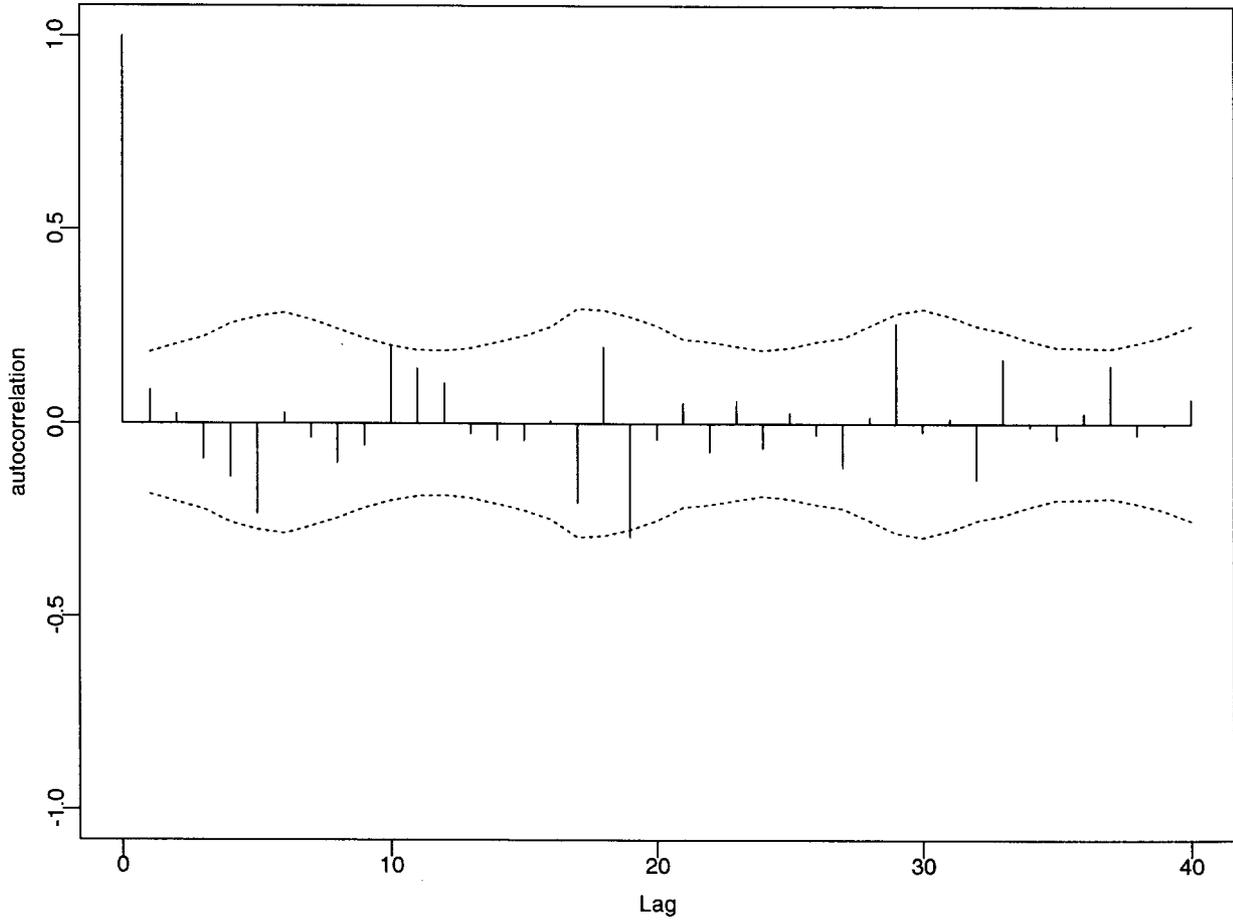
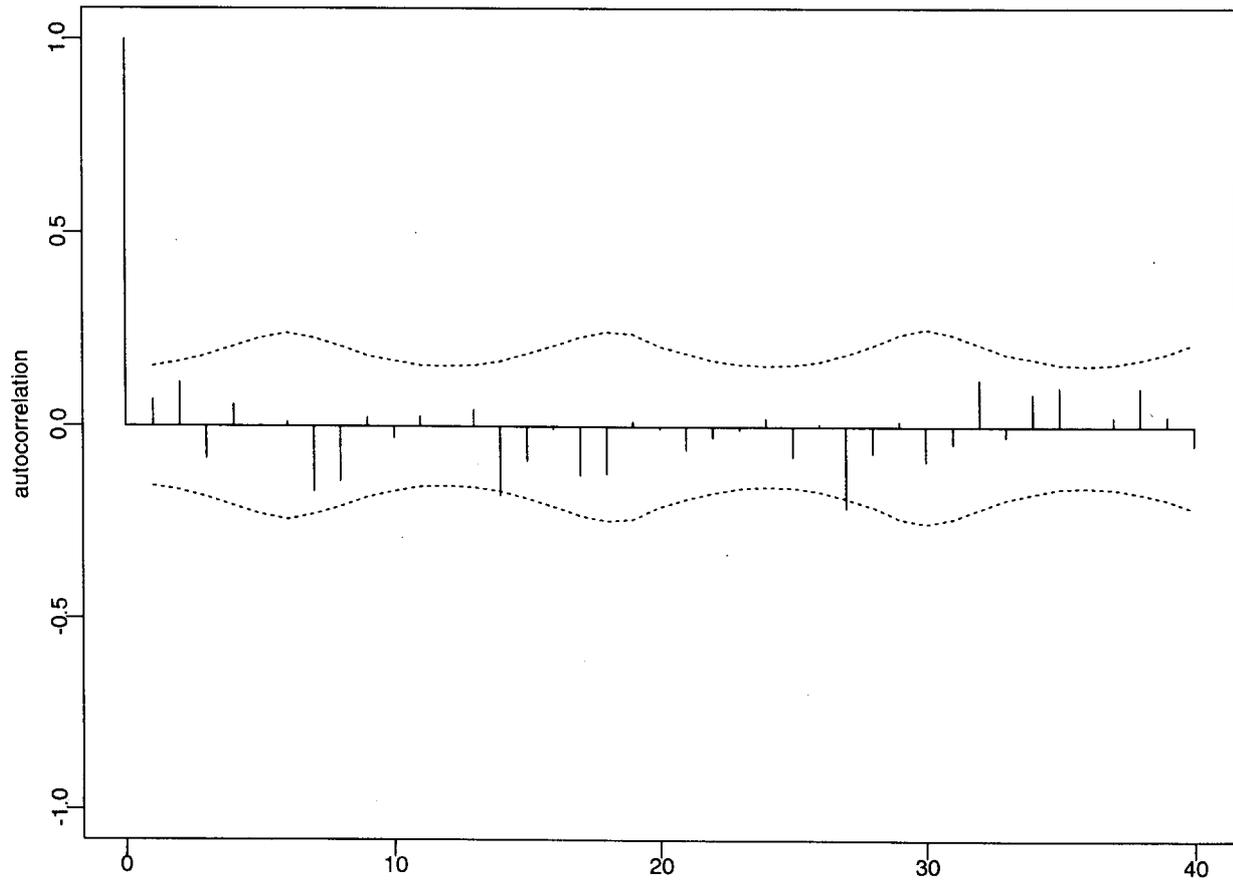


Figure 6. Area B2 : ACRF of Standardized Residuals from Model



Comments on mapping techniques

B. Planque and F. Ibanez

Laboratoire d'Écologie du Plancton Marin, Villefranche-sur-Mer, France

1. CPR data and mapping

CPR data are derived from sequential transects in a range of directions on a large area, up to Oceanic-basin scale. Delays in the sampling between the different locations - within and between transects - is a significant factor which affects CPR results at daily, monthly, and long-time scales.

Subdividing the sampled area into smaller regions (for instance the 10 latitude and 2° longitude CPR boxes), and averaging the data in each of these, has provided ecologically meaningful results in the past. This method of adjacent squares enabled the visualization of the main discontinuities in the species distribution. It has also permitted to analyse and compare very long time series. However, this procedure cannot be interpreted as a mapping technique.

2. The spatio-temporal scales

The search for an optimal representation of the distribution of plankton organisms depends on defining an interval of time in which only the spatial variability - as opposed to temporal variability - can be considered important; i.e. a period during which, the true spatial distribution of the species is not radically modified by time-dependant processes.

This raises the question:

How is the variability of CPR data distributed in time and space?

The variogram function

From samples collected during the year 1991, we have examined the spatial variability of the abundance of *Calanus finmarchicus* stage CV and adults in the North Atlantic. This was done by using the 'structure function' called semi-variogram that describes the level of variability at different scales. The experimental semi-variogram function (also called variogram) is given by:

$$g(h) = \frac{1}{2N(h)} \sum_i (x_i - x_{i+h})^2$$

where $N(h)$ is the number of sampling locations separated by a distance h , x_i and x_{i+h} are the abundance values at locations separated by the distance h .

The experimental variogram provides a measure of the spatial dependency (or spatial autocorrelation) between sampling locations. It is thus a powerful tool to describe the spatial and temporal variability. A theoretical model - fitted to the experimental variograms - can be used in the kriging, a method that provides some of the most accurate estimations in case of spatially autocorrelated data (Matheron, 1974).

Results

Our first purpose is to determine if the spatial variability changes with seasons?

Fig. 1a shows spatial variograms calculated for each season. The Winter variogram is very flat, indicating a poor structured spatial distribution. The same is true in autumn (high values of the variogram for long-distances are meaningless given the poor number of degrees of freedom). In contrast, during spring and summer, the variograms are much more interesting: the spatial variability is pronounced and high values occur at almost 200 miles for spring. In both spring and summer, there is a discontinuity at almost 450 miles, which is the maximum length of each tow. Most of the spatial pattern is visible in this scale. At a higher spatial scale, the regular increase of the variogram function during spring and summer is probably the result of large scale spatial trends. Consequently to these results it seems obvious that mapping may differ for each season. However, given the number of transects, the sampling interval unit could be reduced to each month.

Fig. 1 b shows the monthly variogram functions. The January variogram is cyclic with low values at about every 150

miles probably corresponding to patchiness scale. The February variogram does not display any spatial structure. The March variogram is characterised by a maximum value at a scale of 150-200 miles, indicating that spatial structures exist below this scale. April, May, June and July variograms are very similar, with a pronounced high variability at the scale of 200 miles. August, September and October variograms are irregular, and November displays a high variability at a distance of 90-100 miles. These results indicate that specific months should be selected and gathered to draw valid charts with homogeneous spatial structures.

The second purpose is to determine if the spatial variability changes between day and night and if the day/night effects are constant all through the year.

Fig. 2a shows the variograms calculated during winter 1991 for day samples only or night samples only. There is a strong difference (almost twice the variance) in variability between day and night samples separately with the variance being higher during the night. This result suggests charts for day and night could be constructed separately.

In Fig. 2b and 2c. we constructed variograms by comparing samples taken by day to samples taken during the night (day/night) or by comparing each day samples to every other day samples and each night samples to every other night sample (day/day and nigh/night). This figures shows that in both periods (April-June and July-September) when day and night samples are gathered, there is an increase in the spatial variance at short scales (<200 miles). This suggests that the estimation of the spatial variance is biased by local daily changes in abundance probably induced by *Calanus finmarchicus* diel vertical migrations.

3. Future mapping strategy

After this examination of the spatio-temporal variability, a mapping strategy is proposed: the geographical representation must include several transects. The maximum number of transect that can be gathered depends on the assumption that the temporal variability remains negligible when compared to the spatial variability.

A stepwise procedure has to be followed:

1. A spatial variogram is calculated as soon as a new transect is added to the preceding one.
2. It is tested if the new variogram differs significantly from the previous one.
 - if not, data of the new transect are incorporated into the group of samples used for mapping and the procedure starts again with a new transect.
 - if yes, data of the new transect are rejected and the map can be constructed from data provided by previous transects. The new transect will constitute the basis of another map.

Testing whether variograms differ significantly is a difficult statistical problem, because all the values of the variables are highly correlated. A jack-knife procedure is proposed:

Given N transects, calculate N variograms on N- 1 transects:

- 1) 1,2,3, ...N-1
- 2) 1,2,3, ...N-2,N
- 3) 1,2,3, ...N-3,N-1,N

N) 2,3,4,N

For each distance class h, there are N values for the experimental variogram. The distribution of these values can be included into statistical test and differences between variograms can be statistically tested. These statistical test can be adapted to test the eventual change of variability introduced by adding new transects.

4. The anisotropy problem

Having recognised a temporally homogeneous period, the problem of anisotropy in the field of sampling has to be considered. It is obvious that variability in space may differ considerably from one orientation to another (Ibanez and Boucher, 1987): the latitudinal deviation may for instance be much more important than the longitudinal one.

Variograms for two orthogonal dimensions should be calculated. The idea is to recognise the scale at which the spatial influence of the station becomes negligible. The two distances serve to define the true size of the meshes of the grid necessary for kriging.

5. Conclusion

Interactions between three dimensions (two in space and one in time) characterise the CPR data, and trying to create charts independent of time variation is a true challenge. This note provides a suggested plan for future analysis. The sensitivity of the test allowing the acceptance of new data has to be carefully considered.

The next step of the analysis will be the elaboration of a program such that, given any date (for the day or for the night), a user could obtain the map of any variable.

A comparison of these maps through time (three dimensional analysis) will constitute the aim of further research.

References

- Ibanez F. and Boucher J. 1987. Anisotropie des populations zooplanctoniques clans la zone frontale de Mer Ligure. *Oceanol. Acts*, 10,208-216.
- Matheron G. 1974. Application de la théorie des variables régionalisées. *Cab. C.M.M. Fontainebleau*, 5, 212pp.

Figure 1. (a) Spatial variograms of *Calanus finmarchicus* abundance during the day for Spring, Summer, Autumn and Winter 1991. Small scales variability is maximum in Spring. (b) Monthly spatial variograms (day samples only). Spatial variability is maximum during the period April-July with spatial structures of about 150-200 miles size.

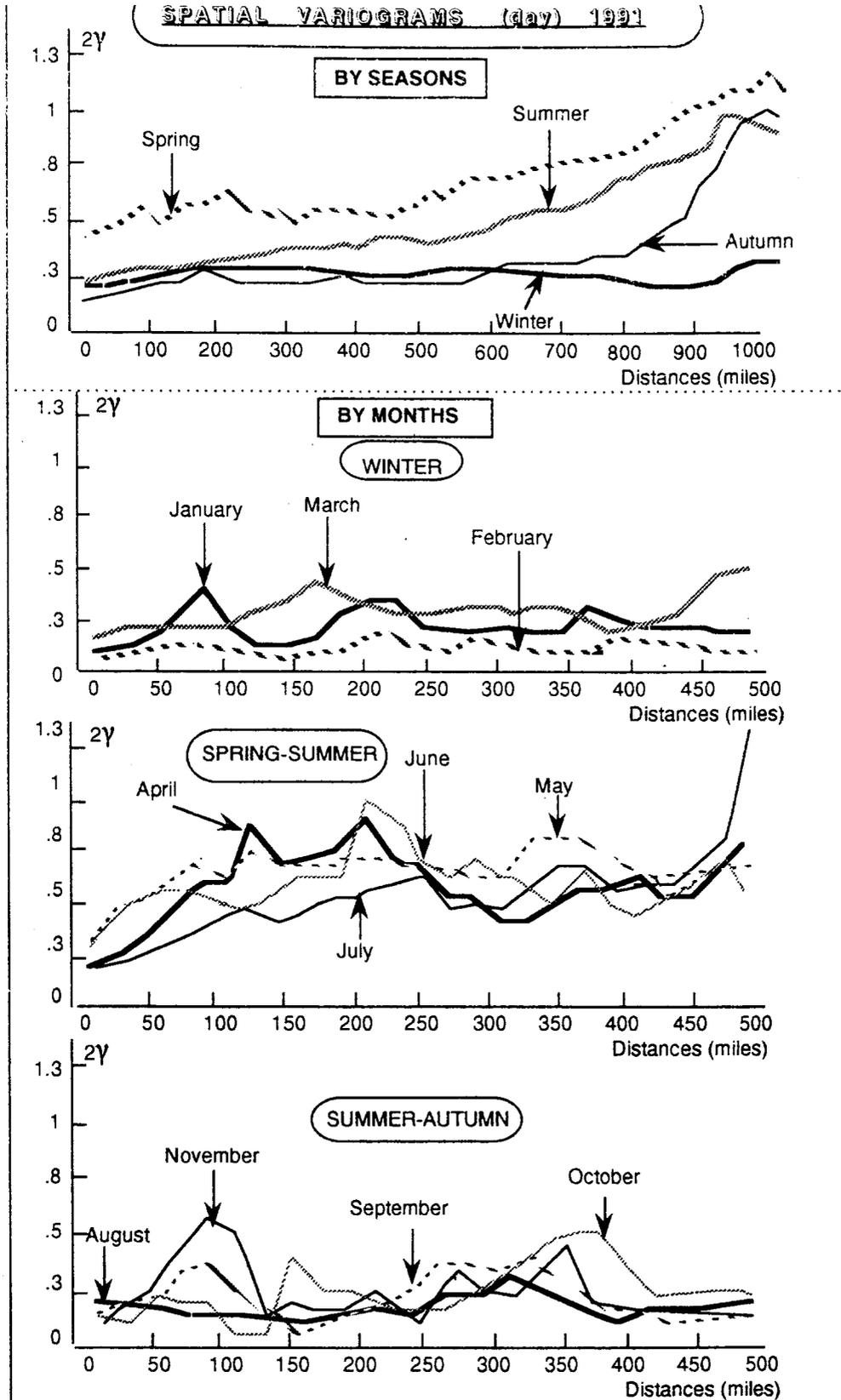
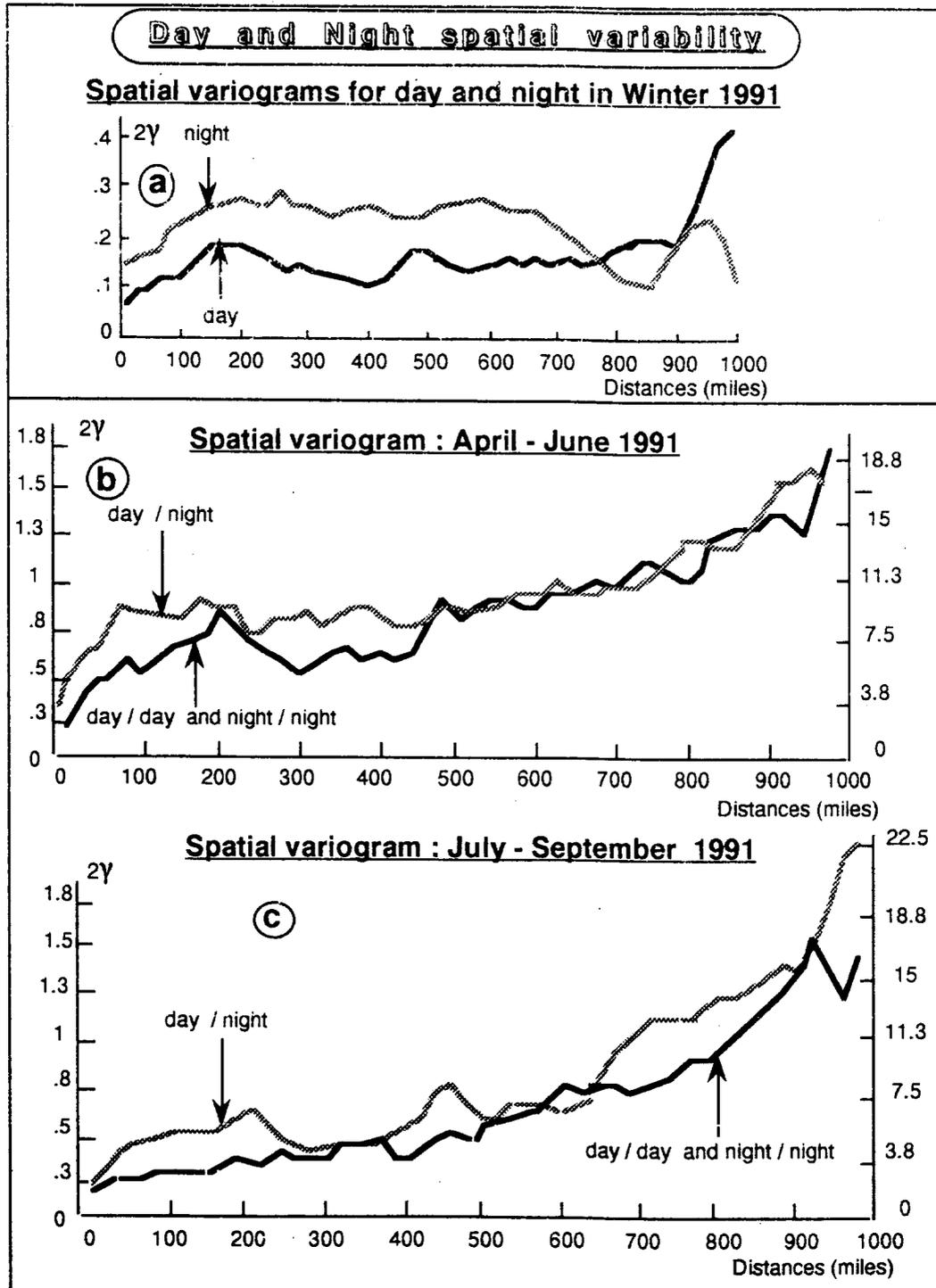


Figure 2. Comparison of spatial variability during day and night. (a) Variograms calculated from day samples only (gray) and night samples only (black) for winter 1991. (b) Variograms calculated by comparing day samples to night samples (gray) or by comparing each day sample to every other day samples and each night samples to every other night samples (black) for the period April-June. (c) same as b but for the period July-September.



Detecting changes in the composition of a multi-species time series.

A. R. Solow

Marine Policy Centre, Woods Hole Oceanographic Institution, U.S.A

Summary

This paper is concerned with extracting a trend from a time series of the composition (i.e. relative abundance) of a multi-species community. The approach is based on modifying a nonparametric method for extracting a trend from ordinary (i.e.; non-compositional) multiple time series for use with compositions

A common approach to extracting a trend from multiple time series is by principal component analysis (PCA). The goal of PCA is to produce orthogonal linear combinations of the original time series with variances decreasing from the maximum possible. In many cases, the first few such components account for a substantial fraction of the total variance and are typically attributed to trend. This attribution is based on an implicit association of high variance with trend. While this association is reasonable in some cases, in others it is not.

An alternative to PCA is minimum/maximum autocorrelation factor analysis (MAFA). The goal of MAFA is to produce orthogonal linear combinations of the original time series with smoothness, as measured by lag-one autocorrelation, decreasing from the maximum possible. Thus, MAFA is based on an implicit association of smoothness with trend.

As with other multivariate methods, MAFA cannot be applied to compositional time series without modification. The modification proposed in this work is to convert the original compositional time series to log contrasts, to apply MAFA to the transformed series, and to transform back to the original scale by a logistic transformation

The problem of determining the number of factors to retain in the analysis is addressed via a sequential randomization procedure.

Two examples of the application of this approach are described: one concerning changes in the composition of Georges Bank fish stocks and the other concerning changes in the Gulf of Maine zooplankton community.

1. Introduction

In ecology, interest commonly centers on detecting a change in the composition of a multi-species community. This paper outlines a non-parametric approach to extracting a common trend from a time series of compositional data and describes two applications of this approach. The approach is based on a method of trend extraction for ordinary multiple time series proposed by Shapiro and Switzer (1989), modified for application to compositional data along the lines suggested by Aitchison (1983).

2. Extracting a trend from a time series of compositions

Let $X(t) = (X_i(t), i = 1, \dots, p)'$, $t = 1, \dots, n$, be a multiple time series. To begin with, suppose that these are ordinary (i.e. not compositional) time series that have been centered by subtracting their respective means. In some situations, interest centers on identifying a common signal or trend in these data. One approach is to use principal component analysis (e.g. Jolliffe, 1986). The goal of principal component analysis (PCA) is to construct a set of orthogonal linear combinations of the original time series (called principal components or PCs) with variance decreasing from the maximum possible. In many applications, the first few PCs account for a large proportion of the total variance in the data and are taken to represent trend. An objection to this use of PCA is that it does not exploit the order of the data, in the sense that the weights used to form the PCs are invariant to permutation of order.

A number of alternative approaches have been developed (e.g. Box and Tiao, 1977; Pena and Box, 1987; Shapiro and Switzer, 1989). Instead of the variance criterion used in PCA, these focus on some measure of predictability or smoothness. The least formal of these methods is minimum/maximum autocorrelation factor analysis (MAFA) proposed by Shapiro and Switzer (1989). The goal of MAFA is to construct a set of orthogonal linear combinations of the original time series (MAFs) with smoothness, as measured by lag-one autocorrelation, decreasing from the maximum possible.

Let X be the n -by- p matrix whose columns are the original time series and let D be the $(n-1)$ -by- p matrix whose columns are the first differences of the original time series. The covariance matrices of the original time series and their first differences are $C=(X'X)/n$ and $V=(D'D)/(n-1)$, respectively. The first MAF has the form:

$$Y_1(t) = \mathbf{a}' \mathbf{X}(t) \quad (1)$$

Since:

$$(\mathbf{a}' \mathbf{V} \mathbf{a})/(\mathbf{a}' \mathbf{C} \mathbf{a}) = 2(1 - r_1) \quad (2)$$

where r_1 is the lag-one autocorrelation of $Y_1(t)$, it follows that, to maximise lag-one autocorrelation, the weight vector \mathbf{a} should be proportional to the eigenvector of the matrix $\mathbf{c}^{-1} \mathbf{v}$ corresponding to the smallest eigenvalue of $\mathbf{c}^{-1} \mathbf{v}$ (which is equal to $2(1 - r_1)$). In the applications described in the next section, the constant of proportionality was chosen to ensure that $Y_1(t)$ has unit variance and that the weight on $X_1(t)$ was positive. In general, the n -by- p matrix whose columns are the MAFs is:

$$\mathbf{Y} = \mathbf{X} \mathbf{A} \quad (3)$$

where the columns of \mathbf{A} are proportional to the eigenvectors of $\mathbf{c}^{-1} \mathbf{v}$.

Because the matrix of $\mathbf{c}^{-1} \mathbf{v}$ is not symmetric, Shapiro and Switzer (1989) proposed the following indirect method for extracting its eigenvectors. A key feature of MAFA is that its results are invariant to affine transformations of X . It follows that the columns of the weight matrix \mathbf{A} are proportional to the eigenvectors of the covariance matrix of the first differences of the scaled (i.e. by dividing by the square root of the corresponding eigenvalue) PCs of C . In other words, the matrix \mathbf{A} can be found by applying PCA to the first differences of the scaled PCs of C .

As with PCA, MAFA has two rather different uses: as an index definition procedure and as a smoothing algorithm. In many cases, only the first few MAFs exhibit significant autocorrelation. These constructed variables can be treated as index series, representing the common behavior.

The time ordering of the residuals is randomly permuted and the permuted residuals are added to the projections onto the first MAF. The second MAF is extracted from the constructed data and its lag-one autocorrelation is found. The process is repeated a large number of times and the significance of the second MAF is assessed from the position of its lag-one autocorrelation among the ordered lag-one autocorrelations found by randomisation. As before, the assessment of significance should be one-sided. The procedure terminates when the first non-significant MAF is found.

When the multiple time series are compositions (i.e. non-negative real numbers with unit sum), it is necessary to modify the procedure outlined above. Aitchison (1983) proposed a modification of PCA for compositional data and this approach will form the basis of modifying MAFA.

The modification proposed by Aitchison (1983) is to replace the data matrix X by a new matrix Z using the transformation:

$$Z_i(t) = \log (X_i(t) / (X_1(t) \dots X_{p+1}(t))^{1/p}) \quad (5)$$

Because the rows of Z have zero sum, the smallest eigenvalue of the corresponding covariance matrix will be zero. The corresponding eigenvector is discarded and the analysis proceeds as outlined above using the remaining $p-1$ eigenvectors. The inverse transformation:

$$x_i(t) = \exp(Z_i(t)) / (\exp(Z_1(t)) + \dots + \exp(Z_{p-1}(t))) \quad (6)$$

can be used to return to the compositional scale.

3. Applications

The first application described in this section concerns the composition of the Georges Bank fish stock. Over the past thirty years, the fish population on Georges Bank has been subjected to major perturbations - most notably, the arrival of foreign fishing trawlers in 1972 and their subsequent exclusion in 1975- and there is great interest in the behavior of the population over this period. The National Marine Fisheries Service (NMFS) has undertaken autumn surveys of

Georges Bank since 1963. Figure 1 shows the proportions of four categories of fish - groundfish and flounders, other fish, principal pelagics, and dogfish and skates - for the period 1963-1988.

A standard ecological approach to monitoring the composition of a multi-species community is through a diversity index (e.g. Magurran, 1988). For example, the Shannon index of diversity in period t is:

$$H(t) = - \sum_{i=1}^p X_i(t) \log X_i(t)$$

In Figure 2, $H(t)$ is plotted against t for the data shown in Figure 1. In qualitative terms, diversity appears to increase from the beginning of the observation period until around 1970, after which it declines slowly before rebounding in 1988.

To better understand the changes in stock composition, the method outlined in the previous section was also applied to these data. The weight matrix is:

$$A = \begin{matrix} & \begin{matrix} 0.56 & -1.87 & -0.58 \\ 0.44 & 1.93 & 1.22 \end{matrix} \\ \begin{matrix} 0.03 & 0.58 & -1.07 \\ -1.03 & -0.63 & 0.42 \end{matrix} & \end{matrix}$$

The columns of A have a zero sum. This follows from their orthogonality to the discarded eigenvector corresponding to the zero eigenvalue, which is proportional to an n -vector of 1's. The lag-one autocorrelations of the three MAF's are 0.84, 0.19 and -0.03. The randomization procedure described above was used to assess the significance of these MAF's. Of 1000 random permutations of the original time series, none had a first MAF with lag-one autocorrelation larger than 0.84, so the first MAF was retained. Of 1000 constructed time series, 217 had a second MAF lag-one autocorrelation larger than 0.19, so the second and third MAF's were discarded.

The first MAF was smoothed using a simple 5-point moving average. The results are shown in Figure 3. The MAF, which represents the difference between a weighted sum of categories 1 and 2 and category 4, declines steadily over this period. The series of transformed proportions were projected onto the smoothed first MAF and retransformed according to (6). The results, which are shown in Figure 1, indicate a switch from flounders and groundfish and other fish to dogfish and skates, with the proportion of principle pelagic species remaining approximately constant.

The second application described in this section concerns the composition of the zooplankton community in the Gulf of Maine. The time series of the relative abundances of five zooplankton species *Calanus*, *Pseudocalanus*, *Centropages typicus*, *Oithona*, and *Metridia lucens*, for the period 1962-1989 are shown in Figure 4. These data were also collected by the NMFS. The diversity index $H(t)$ is plotted against t for these data in Figure 2. In this case, the weight matrix is:

$$A = \begin{matrix} & \begin{matrix} 1.25 & 0.63 & 0.04 & -2.39 \\ 0.65 & 0.76 & -0.96 & 2.60 \\ -0.18 & -2.27 & 1.93 & 1.90 \\ -1.49 & -0.15 & -1.13 & -2.29 \\ -0.23 & 1.06 & 0.13 & 0.18 \end{matrix} \end{matrix}$$

The lag-one autocorrelations of the four MAF's are 0.59, 0.30, 0.18, and -0.52. The significance level of the first MAF estimated from 1000 random perturbations of the original time series was 0.011, so the first MAF was retained. The significance level of the second MAF estimated from 1000 constructed time series was 0.201, so the remaining MAF's were discarded.

Figure 5 shows the first MAF, smoothed using a simple 5-point moving average. This MAF which represents the difference between a weighted sum of categories 1 and 2 and category 4, exhibits quasi-periodic behaviour. There is also evidence of this behaviour in $H(t)$, although the two series are out of phase. The time series of the transformed proportions were projected onto the smoothed first MAF and re-transformed. The results which are shown in Figure 4, indicate quasi-periodic shifts between *Calanus* and the *Oithona*, *M. lucens* complex. Although the weight on *M. lucens* in the first MAF is small, its strong correlation with *Oithona* causes its projection to be strongly correlated with that of

Oithona.

4. Discussion

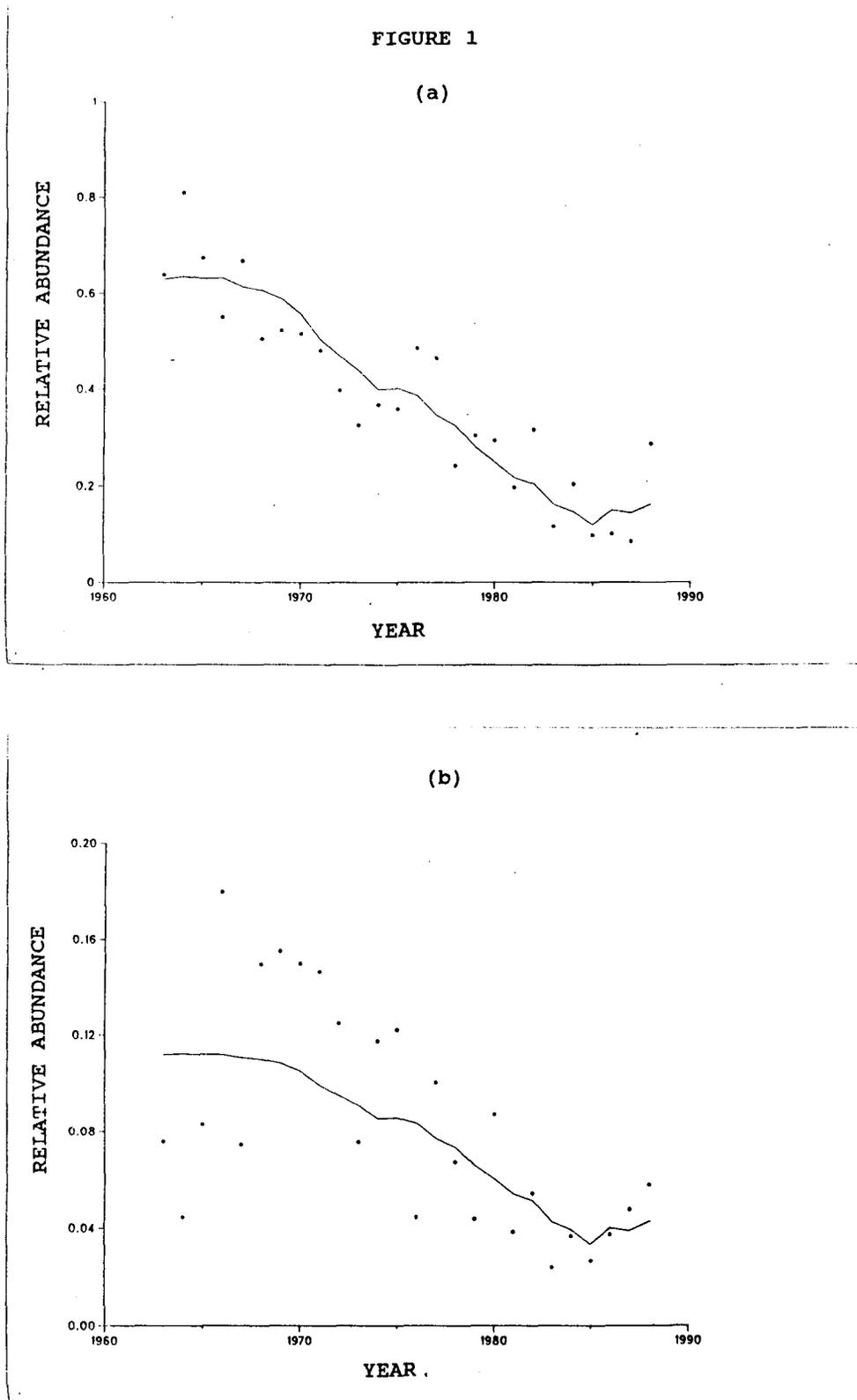
While the method outlined in this paper has considerable intuitive appeal, it is also justified in part optimality results for MAFA described by Shapiro & Switzer (1989). Specifically, they showed that the first MAF maximises the signal-to-noise ratio in the case where lagged covariances are a fixed fraction of unlagged covariances for both the multivariate signal and the multivariate noise. Even without this kind of formal justification, the use of lag-one autocorrelation as a smoothness criterion is clearly preferable to the use of variance.

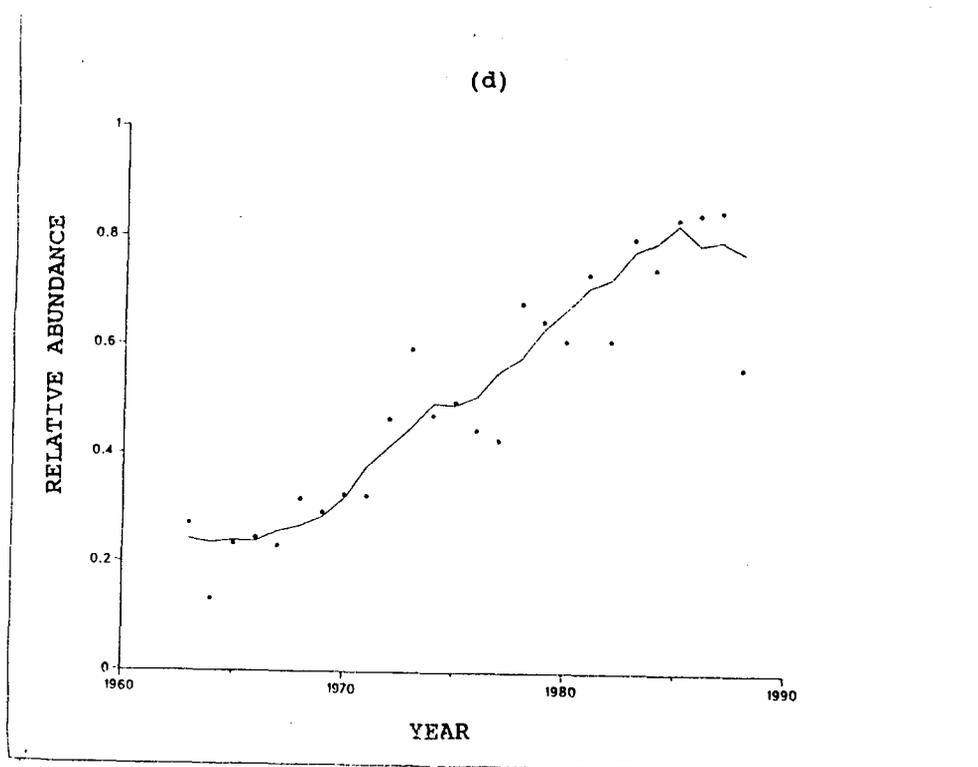
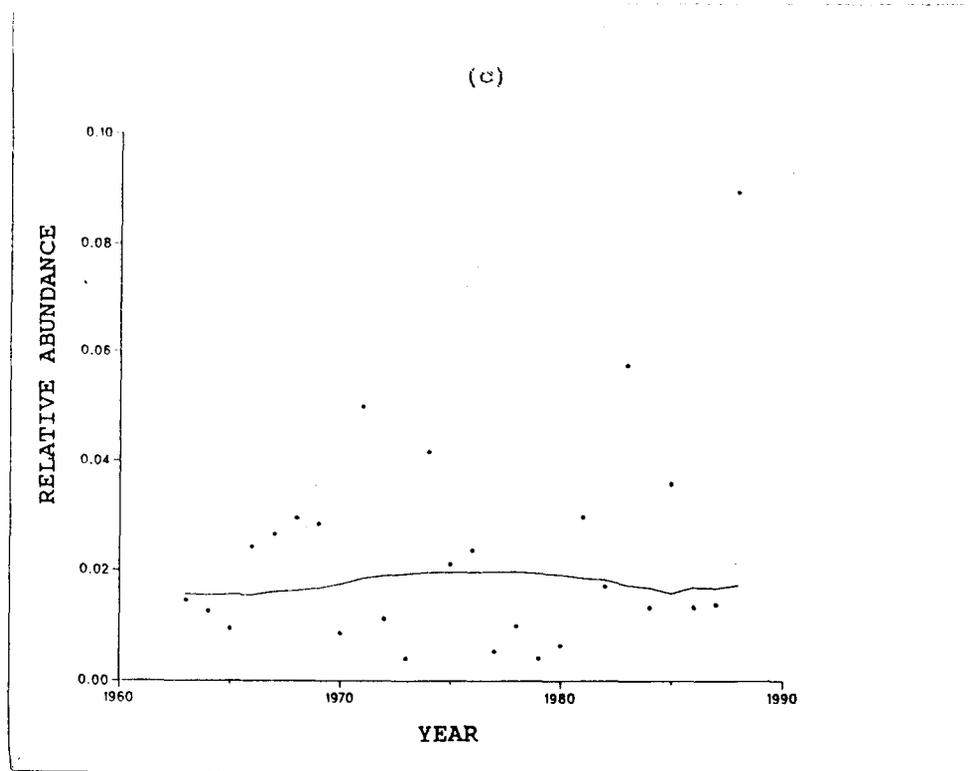
The method worked well in both applications described in the previous section, capturing the main features of change in composition. In contrast, the approach based on Shannon's diversity index provided almost no information about the nature of this change. An important advantage of the method described in this paper is that, by focusing on a small number of index series, it facilitates an understanding of relationships between the composition of a community and other factors. This understanding may be reached by purely empirical methods (e.g. by correlating the retained MAF's to time series of environmental variables). For example, the behaviour of the first MAF extracted from the zooplankton data appears to be related to variability in sea surface temperature, although more careful analysis is needed to establish this relationship. It may also be possible to make use of the interpredability of the retained MAF's. For example, the nature of the turnover of the fish stock on Georges Bank (from commercial species to non-commercial species) is consistent with the effects of over-exploitation, while a turnover in the opposite direction would not be.

References

- Aitchinson, J. 1983. Principal component analysis of compositional data. *Biometrika*, 70:57-65.
- Box, G. E. P., & Tiao, G.C. 1977. A cononical analysis of multiple time series. *Biometrika*, 64:355-365.
- Jolliffe, I.T. 1986. Principle Component Analysis. Springer-Verlag, New York.
- Magurran, A.E. 1988. Ecological Diversity and its Measurement. Princeton University Press, Princeton.
- Pena, D., & Box, G.E.P. 1987. Identifying a simplifying structure in time series. *Journal of the American Statistical Association*, 82:836-843.
- Shapiro, D. E., & Switzer, P. 1989. Extracting time trends from multiple monitoring sites. Technical Report 132, Department of Statistics, Stanford University.

Figure 1. Annual time series of Georges Bank fish stock composition and estimated trends 1963-1988. (a) flounders and groundfish, (b) other fish, (c) principle pelagics, (d) dogfish and skates.





IOC Workshop Report No. 124
Annex II - page 58

Figure 2. Shannon's diversity index for Georges Bank fish stock (solid) and Gulf of Maine zooplankton (dashed).

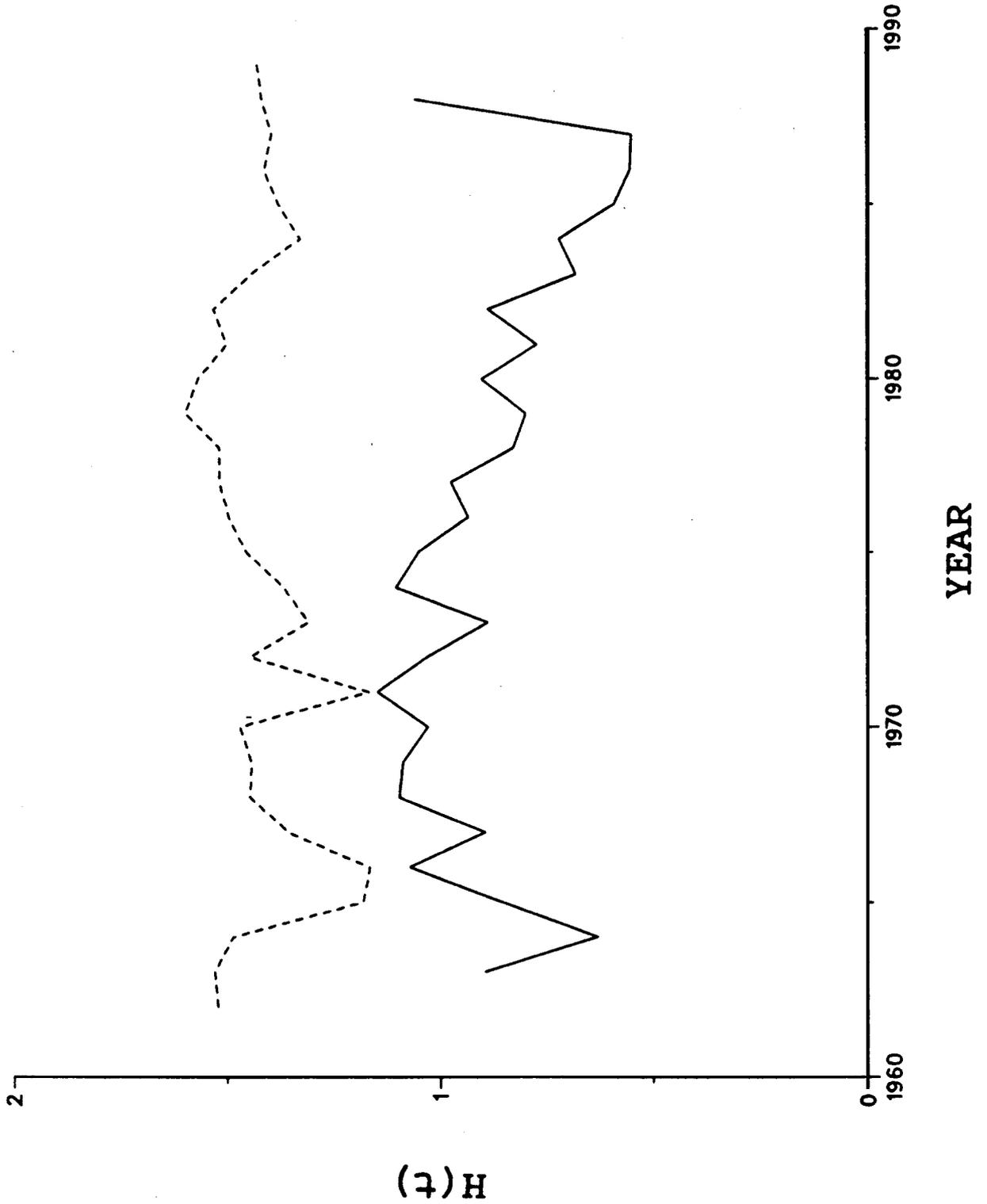


Figure 3. First MAF for Georges Bank fish stock smoothed by 5-point moving average,

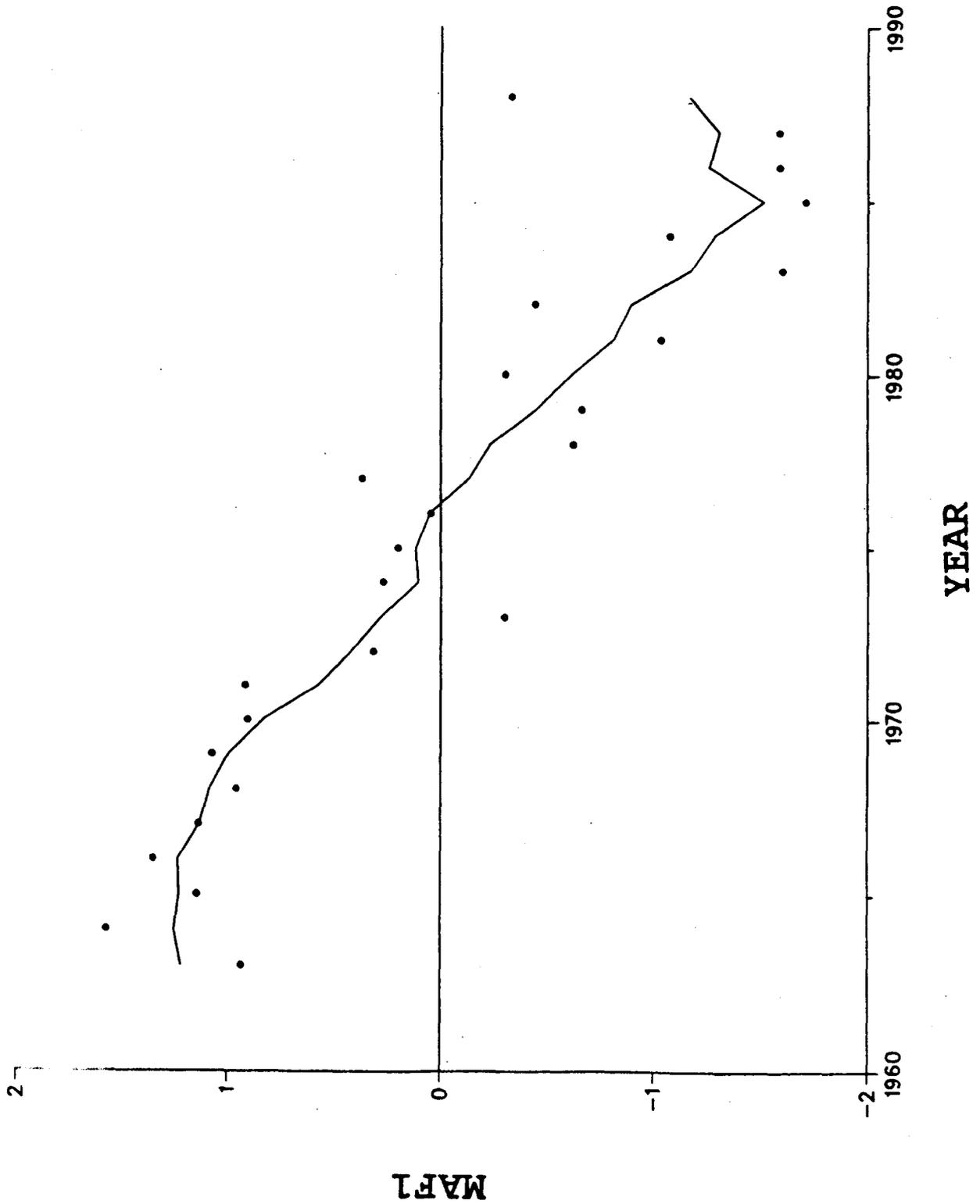
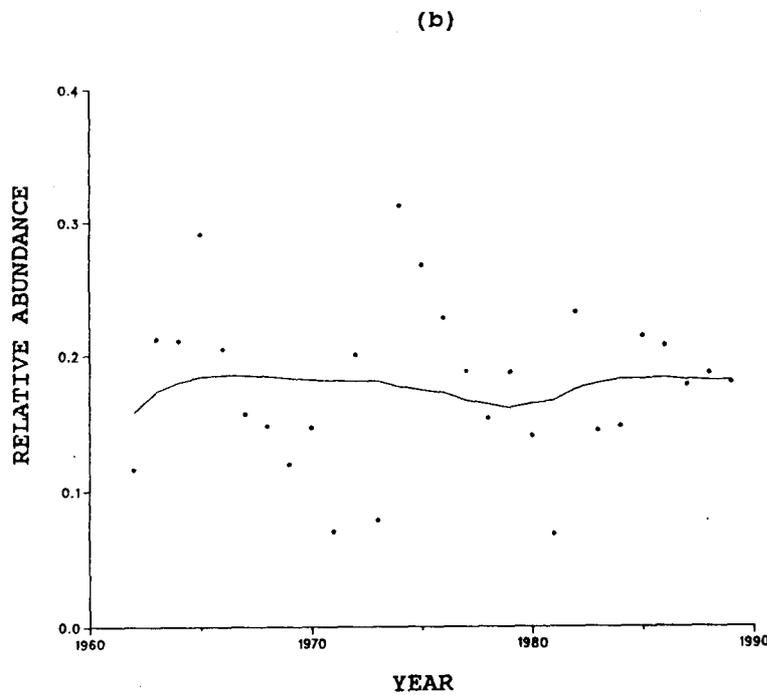
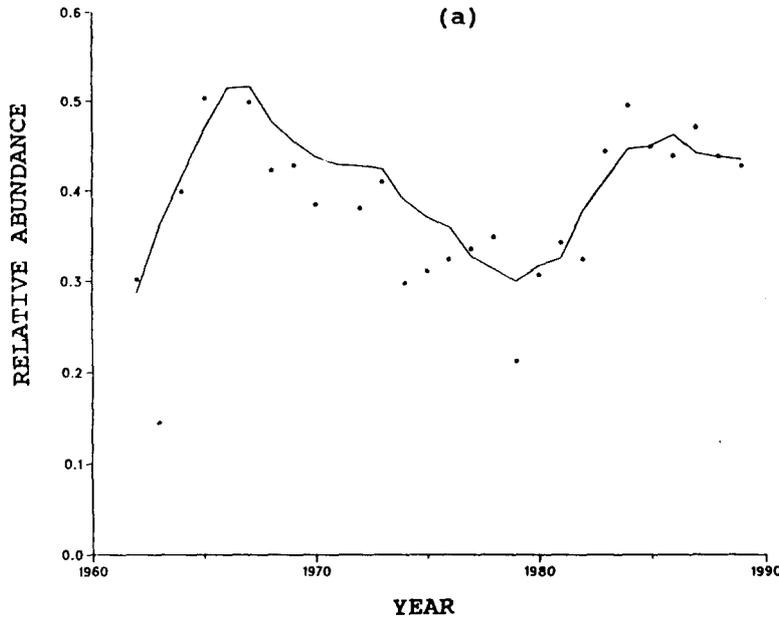
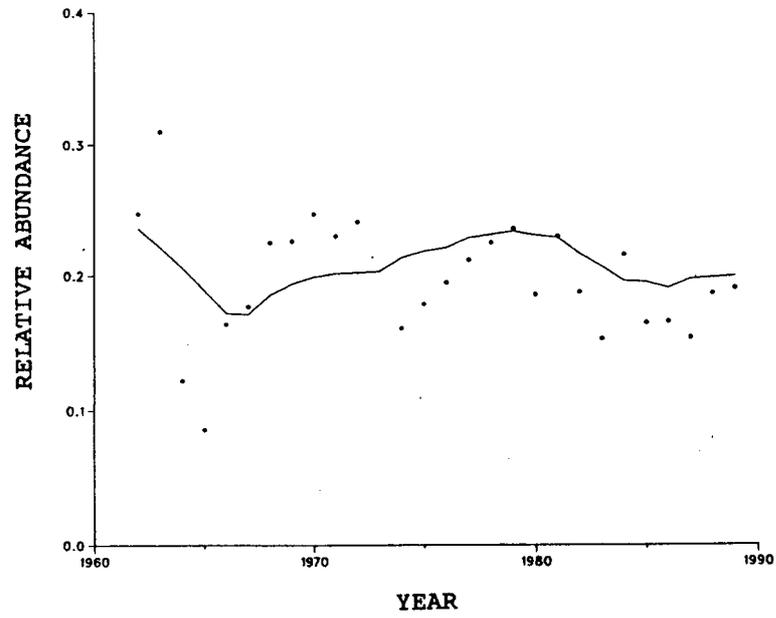


Figure 4. Annual time series of Gulf of Maine zooplankton composition and estimated trends 1962-1989. (a) *Calanus* (b) *Pseudocalanus* (c) *Centropages typicus* (d) *Oithona* (e) *Metridia lucens*.

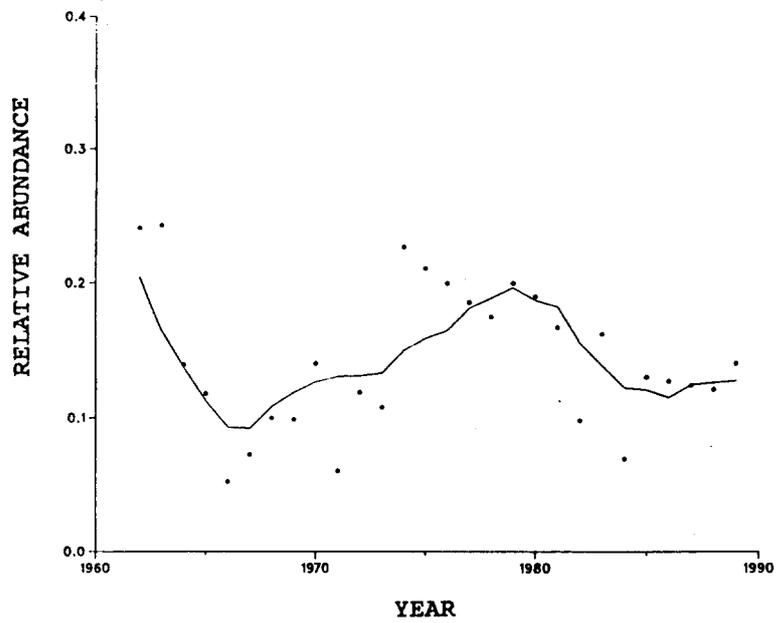
FIGURE 4



(c)



(d)



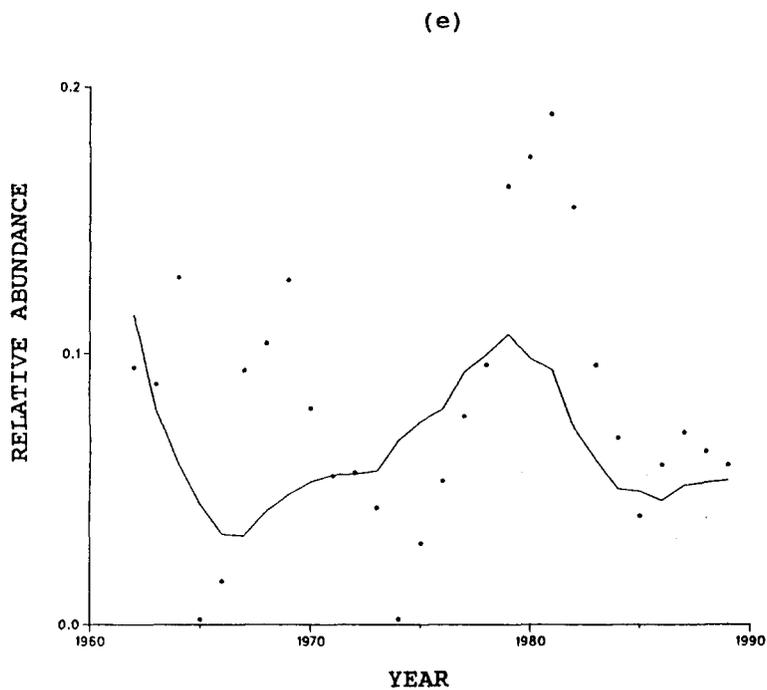
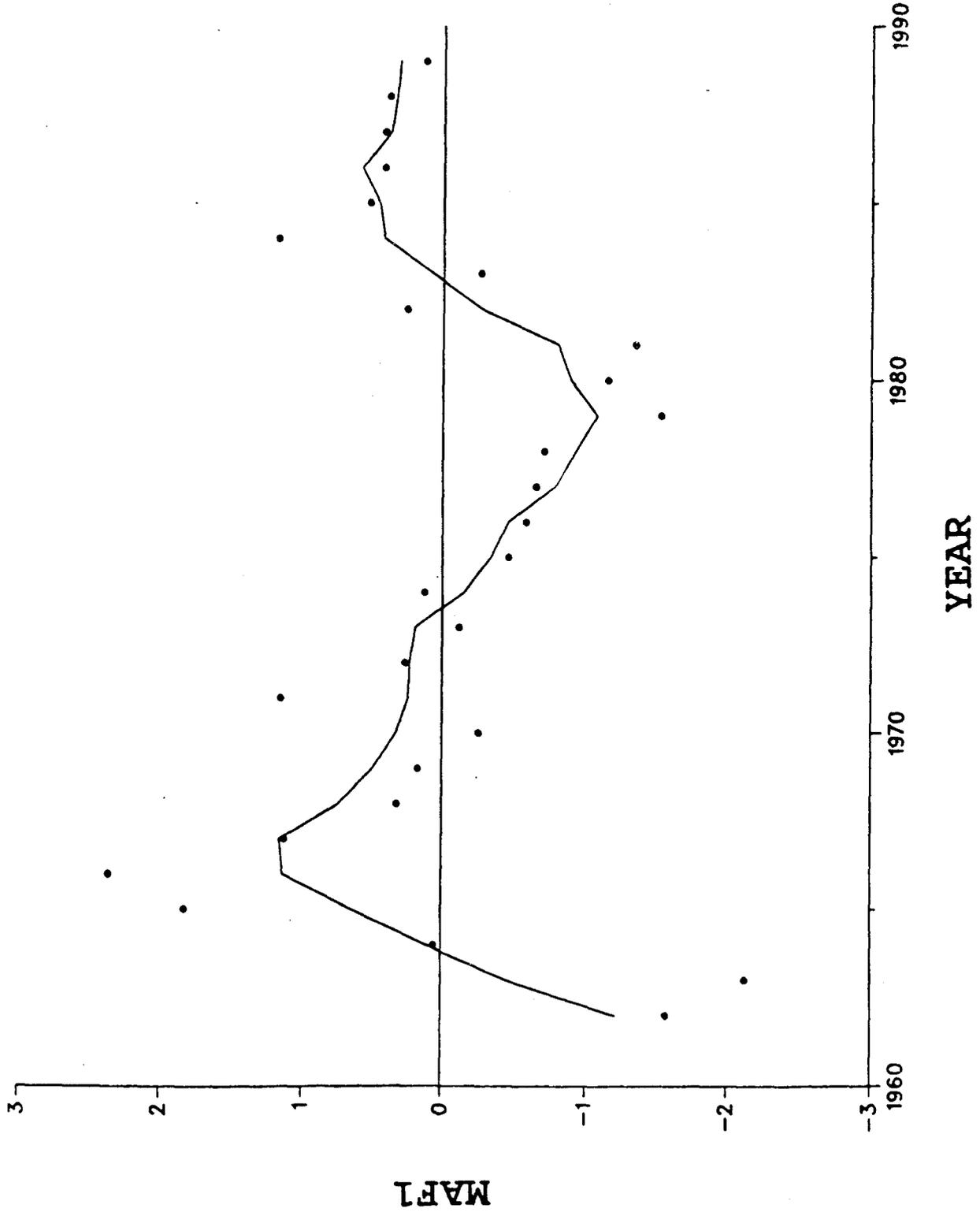


Figure 5. First MAF for Gulf of Maine zooplankton smoothed by 5-point moving average.



Gulf Stream position and the plankton of the European Shelf

A. H. Taylor

Plymouth Marine Laboratory, U.K.

The year-to-year variations seen in data from the Continuous Plankton Recorder are part of a pattern of climatic changes stretching across the North Atlantic. This has been demonstrated by comparing the plankton trends with a time-series of the latitude of the north wall of the Gulf Stream (Fig. 1, Taylor and Stephens, 1980; Taylor *et al.*, 1992). Charts showing the position of the north wall have been published for every month since 1966 in the periodicals *Gulf Stream Monthly Summary*, *Gulf Stream* and *oceanographic Monthly Summary*. The latitude of the wall was read from each chart at the longitudes 79°W to 65°W and an index of its monthly position was calculated by means of principal components analysis. The annual values of this index are the broken lines on Fig. 1.

The similarities to the fluctuations in the abundance of plankton shown in Fig. 1 implies that the estimated shifts of the north wall are real, even though these represent less than a quarter of the variance of the positions read from the charts. The relationship indicates a climatic connection spanning the complete width of the North Atlantic and, as such, supports the view that the planktonic trends are driven externally rather than being a consequence of the ecosystem's nonlinear dynamics.

The relationship appears in a wide range of oceanographic regimes: the central North Sea, in the Baltic outflow over the Norwegian trench and in the NE Atlantic. Only in those areas of the Shelf where there is little seasonal stratification does it seem to be completely absent. This implies that the connection is purely atmospheric, operating through the annual cycle of stratification, a suggestion that is supported by observations from Lake Windermere. Summer zooplankton abundance in Windermere (George and Harris, 1985) also shows a clear relationship to the annual time-series of the position of the north wall, and the weekly temperature profiles from the lake show this to be associated with the onset of stratification in the spring. Figure 2 shows the depth of maximum temperature gradient during the first week in June compared with the Gulf Stream index series. When the north wall is displaced northwards weather conditions in the spring and autumn tend to be more anticyclonic with a reduced number of cyclones, and spring mixed layer depths are shallower. In Lake Windermere this leads to a reduced abundance of zooplankton while in the seas surrounding the U.K. it leads to increased abundance of zooplankton.

References

- George, D.G. and Harris, G.P. 1985 The effect of climate on long-term changes in the crustacean zooplankton biomass of Lake Windermere, U.K. *Nature*, 316,536-539.
- Taylor, A.H. and Stephens, J.A. 1980 Latitudinal displacements of the Gulf Stream (1966 to 1977) and their relation to changes in temperature and zooplankton abundance in the NE Atlantic. *Oceanol. Acts*, 3, 148-149.
- Taylor, A. H., Colebrook, J. M., Stephens, J.A. and Baker, N.G. 1992 Latitudinal displacements of the Gulf Stream and the abundance of plankton in the NE Atlantic. *J. Mar. Biol. Ass.*, 72,919-921.

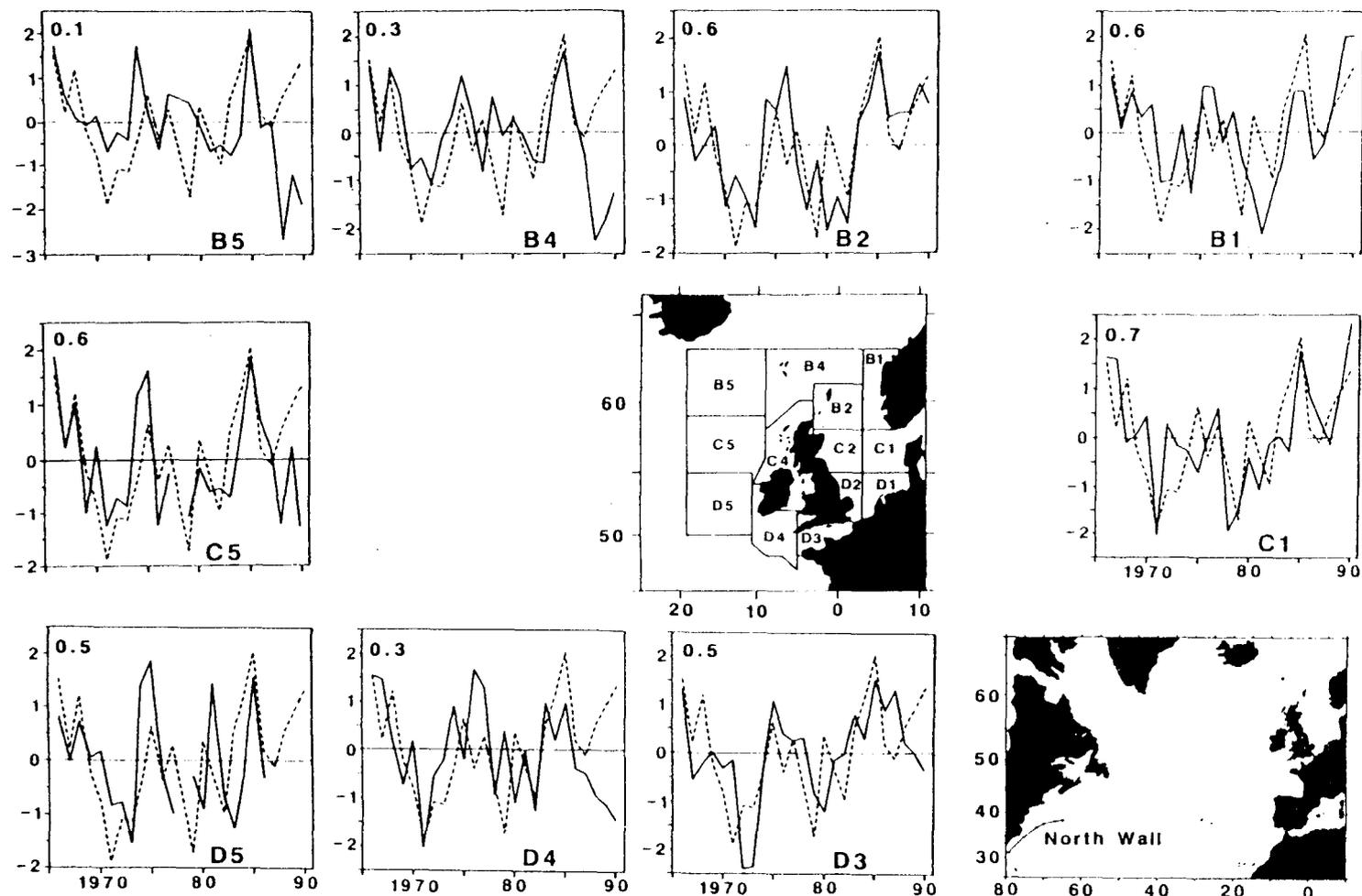
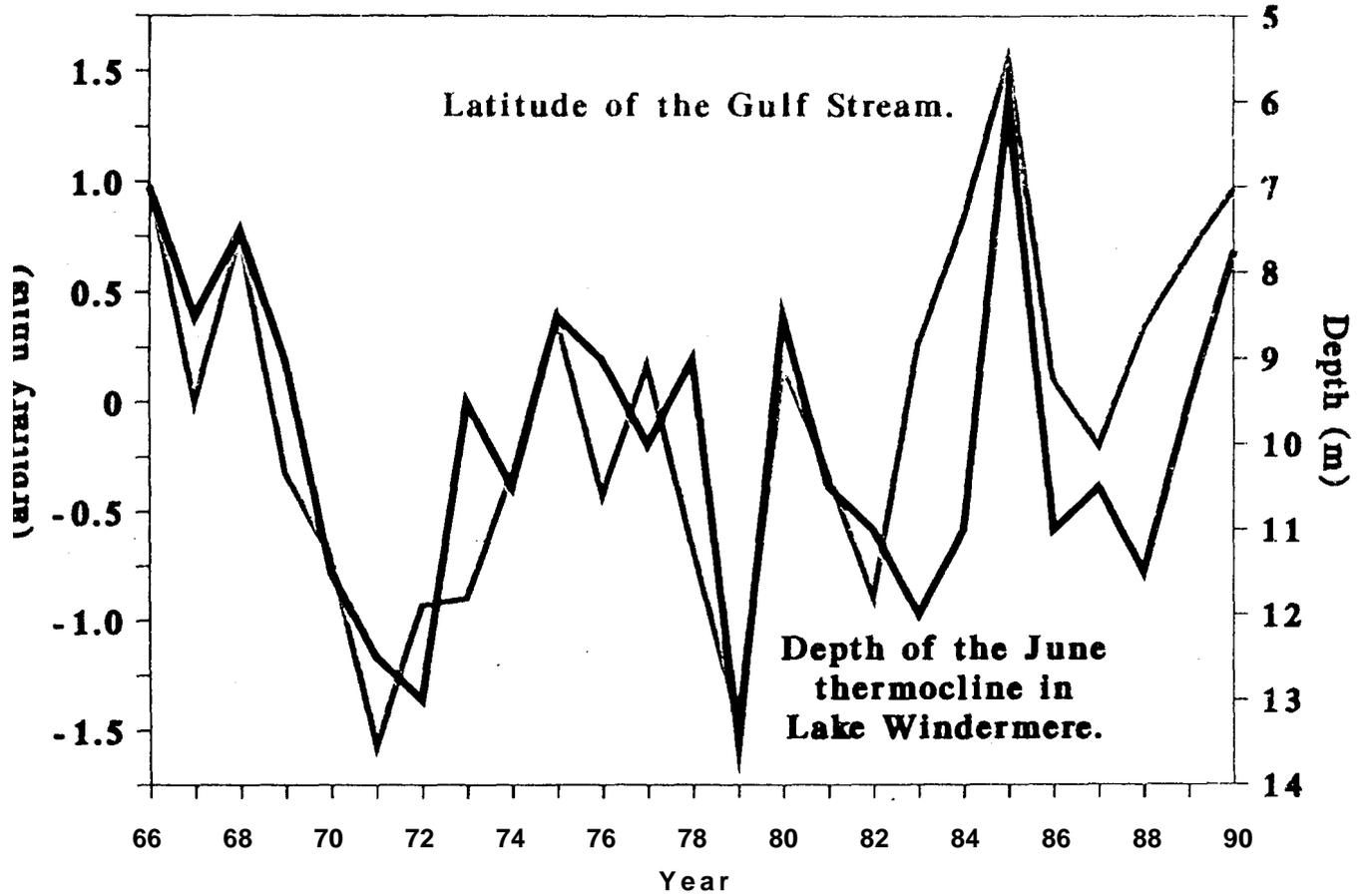


Figure 1. Annual means of the latitude of the north wall of the Gulf Stream (1966 to 1990, broken line) compared with logarithm of the number of copepods in areas of the Continuous Plankton Recorder Survey (solid line). Each graph has been standardized to have zero mean and unit variance and is accompanied by the correlation coefficient between the two time-series (the 5% and 1% significance levels are 0.4 and 0.5, respectively). Correlation coefficients for the areas C2, C4, D1 and D2 which are not shown are 0.2, 0.3, 0.4 and -0.1, respectively. Charts showing the positions of the areas and of the north wall of the Gulf Stream are included.

Figure 2 Summer zooplankton biomass in Lake Windermere 1966-1991 (George & Harris 1985) compared with north-south displacements of the Gulf Stream.



IOC Workshop Reports

The Scientific Workshops of the Intergovernmental Oceanographic Commission are sometimes jointly sponsored with other intergovernmental or non-governmental bodies. In most cases, IOC assumes responsibility for printing, and copies may be requested from:

Intergovernmental Oceanographic Commission - UNESCO
1, rue Miollis, 75732 Paris Cedex 15, France

No.	Title	Languages	No.	Title	Languages	No.	Title	Languages
1	CCOP-IOC, 1974, Metallogenesis, Hydrocarbons and Tectonic Patterns in Eastern Asia (Report of the IDOE Workshop on); Bangkok, Thailand, 24-29 September 1973 UNDP (CCOP), 138 pp.	E (out of stock)	18	IOC/UNESCO Workshop on Syllabus for Training Marine Technicians: Miami, 22-26 May 1978 (UNESCO reports in marine sciences, No. 4 published by the Division of Marine Sciences, UNESCO).	E (out of stock), F, S (out of stock), R	36	IOC/FAO Workshop on the Improved Uses of Research Vessels; Lisbon, 28 May-2 June 1984.	E
2	CICAR Ichthyoplankton Workshop, Mexico City, 16-27 July 1974 (UNESCO Technical Paper in Marine Sciences, No. 20).	E (out of stock) S (out of stock)	19	IOC Workshop on Marine Science Syllabus for Secondary Schools; Llantwit Major, Wales, U.K., 5-9 June 1978 (UNESCO reports in marine sciences, No. 5, published by the Division of Marine Sciences, UNESCO).	E (out of stock), E, S, R, Ar	36	Papers submitted to the IOC/FAO Suppl. Workshop on the Improved Uses of Research Vessels; Lisbon, 28 May-2 June 1984.	E
3	Report of the IOC/GFCM/ICSEM International Workshop on Marine Pollution in the Mediterranean; Monte Carlo, 9-14 September 1974.	E, F E (out of stock)	20	Second CCOP-IOC Workshop on IDOE Studies of East Asia Tectonics and Resources; Bandung, Indonesia, 17-21 October 1978.	E	37	IOC/UNESCO Workshop on Regional Co-operation in Marine Science in the Central Indian Ocean and Adjacent Seas and Gulfs; Colombo, 8-13 July 1985.	E
4	Report of the Workshop on the Phenomenon known as 'El Niño'; Guayaquil, Ecuador, 4-12 December 1974.	E (out of stock) S (out of stock)	21	Second IDOE Symposium on Turbulence in the Ocean; Liège, Belgium, 7-18 May 1979.	E, F, S, R	38	IOC/ROPME/UNEP Symposium on Fate and Fluxes of Oil Pollutants in the Kuwait Action Plan Region; Basrah, Iraq, 8-12 January 1984.	E
5	IDOE International Workshop on Marine Geology and Geophysics of the Caribbean Region and its Resources; Kingston, Jamaica, 17-22 February 1975.	E (out of stock) S	22	Third IOC/WMO Workshop on Marine Pollution Monitoring; New Delhi, 11-15 February 1980.	E, F, S, H	39	CCOP (SOPAC)-IOC-IFREMER-ORSTOM Workshop on the Uses of Submersibles and Remotely Operated Vehicles in the South Pacific; Suva, Fiji, 24-29 September 1985.	E
6	Report of the CCOP/SOPAC-IOC IDOE International Workshop on Geology, Mineral Resources and Geophysics of the South Pacific; Suva, Fiji, 1-6 September 1975.	E	23	WESTPAC Workshop on the Marine Geology and Geophysics of the North-West Pacific; Tokyo, 27-31 March 1980.	E, R	40	IOC Workshop on the Technical Aspects of Tsunami Analysis, Prediction and Communications; Sidney, B.C., Canada, 29-31 July 1985.	E
7	Report of the Scientific Workshop to Initiate Planning for a Co-operative Investigation in the North and Central Western Indian Ocean, organized within the IDOE under the sponsorship of IOC/FAO (IOFC)/UNESCO/EAC; Nairobi, Kenya, 25 March-2 April 1976.	E, F, S, R	24	WESTPAC Workshop on Coastal Transport of Pollutants; Tokyo, 27-31 March 1980.	E (out of stock)	40	First International Tsunami Suppl. Workshop on Tsunami Analysis, Prediction and Communications, Submitted Papers; Sidney, B.C., Canada, 29 July - 1 August 1985.	E
8	Joint IOC/FAO (IPFC)/UNEP International Workshop on Marine Pollution in East Asian Waters; Penang, 7-13 April 1976.	E (out of stock)	25	Workshop on the Inter-calibration of Sampling Procedures of the IOC/WMO UNEP Pilot Project on Monitoring Background Levels of Selected Pollutants in Open-Ocean Waters; Bermuda, 11-26 January 1980.	E (superseded by IOC Technical Series No. 22)	41	First Workshop of Participants in the Joint FAO/IOC/WHO/AEA/UNEP Project on Monitoring of Pollution in the Marine Environment of the West and Central African Region (WACAF/2); Dakar, Senegal, 28 October-1 November 1985.	E
9	IOC/CMG/SCOR Second International Workshop on Marine Geoscience; Mauritius, 9-13 August 1976.	E, F, S, R	26	IOC Workshop on Coastal Area Management in the Caribbean Region; Mexico City, 24 September-5 October 1979.	E, S	43	IOC Workshop on the Results of MEDALPEX and Future Oceanographic Programmes in the Western Mediterranean; Venice, Italy, 23-25 October 1985.	E
10	IOC/WMO Second Workshop on Marine Pollution (Petroleum) Monitoring; Monaco, 14-18 June 1976.	E, F E (out of stock) R	27	CCOP/SOPAC-IOC Second International Workshop on Geology, Mineral Resources and Geophysics of the South Pacific; Noumea, New Caledonia, 9-15 October 1980.	E	44	IOC-FAO Workshop on Recruitment in Tropical Coastal Demersal Communities; Ciudad del Carmen, Campeche, Mexico, 21-25 April 1986.	E (out of stock) S
11	Report of the IOC/FAO/UNEP International Workshop on Marine Pollution in the Caribbean and Adjacent Regions; Port of Spain, Trinidad, 13-17 December 1976.	E, S (out of stock)	28	FAO/IOC Workshop on the effects of environmental variation on the survival of larval pelagic fishes. Lima, 20 April-5 May 1980.	E	44	IOC-FAO Workshop on Suppl. Recruitment in Tropical Coastal Demersal Communities, Submitted Papers; Ciudad del Carmen, Campeche, Mexico, 21-25 April 1986.	E
11	Collected contributions of invited lecturers and authors to the IOC/FAO/UNEP International Workshop on Marine Pollution in the Caribbean and Adjacent Regions; Port of Spain, Trinidad, 13-17 December 1976.	E (out of stock), S	29	WESTPAC Workshop on Marine Biological Methodology; Tokyo, 9-14 February 1981.	E	45	IOCARIBE Workshop on Physical Oceanography and Climate; Cartagena, Colombia, 19-22 August 1986.	E
12	Report of the IOCARIBE Interdisciplinary Workshop on Scientific Programmes in Support of Fisheries Projects; Fort-de-France, Martinique, 28 November-2 December 1977.	E, F, S	30	International Workshop on Marine Pollution in the South-West Atlantic; Montevideo, 10-14 November 1980.	E (out of stock) S	46	Reunión de Trabajo para Desarrollo del Programa 'Ciencia Oceánica en Relación a los Recursos No Vivos en la Región del Atlántico Sud-occidental'; Porto Alegre, Brazil, 7-11 de abril de 1986.	S
13	Report of the IOCARIBE Workshop on Environmental Geology of the Caribbean Coastal Area; Port of Spain, Trinidad, 16-18 January 1978.	E, S	31	Third International Workshop on Marine Geoscience; Heidelberg, 19-24 July 1982.	E, F, S	47	IOC Symposium on Marine Science in the Western Pacific: The Indo-Pacific Convergence; Townsville, 1-6 December 1966.	E
14	IOC/FAO/WHO/UNEP International Workshop on Marine Pollution in the Gulf of Guinea and Adjacent Areas; Abidjan, Côte d'Ivoire, 2-9 May 1978.	E, F	32	UNU/IOC/UNESCO Workshop on International Co-operation in the Development of Marine Science and the Transfer of Technology in the context of the New Ocean Regime; Paris, 27 September-1 October 1982.	E, F, S	48	IOCARIBE Mini-Symposium for the Regional Development of the IOC-UN (OETB) Programme on 'Ocean Science in Relation to Non-Living Resources (OSNLR)'; Havana, Cuba, 4-7 December 1986.	E, S
15	CCPS/FAO/IOC/UNEP International Workshop on Marine Pollution in the South-East Pacific; Santiago de Chile, 6-10 November 1978.	E (out of stock)	32	Papers submitted to the Suppl. UNU/IOC/UNESCO Workshop on International Co-operation in the Development of Marine Science and the Transfer of Technology in the Context of the New Ocean Regime; Paris, 27 September-1 October 1982.	E	49	AGU-IOC-WMO-CCPS Chapman Conference: An International Symposium on 'El Niño'; Guayaquil, Ecuador, 27-31 October 1986.	E
16	Workshop on the Western Pacific, Tokyo, 19-20 February 1979.	E, F, R	33	Workshop on the IREP Component of the IOC Programme on Ocean Science in Relation to Living Resources (OSLR); Halifax, 26-30 September 1963.	E	50	CCALR-IOC Scientific Seminar on Antarctic Ocean Variability and its Influence on Marine Living Resources, particularly Krill (organized in collaboration with SCAR and SCOR); Paris, France, 2-6 June 1987.	E
17	Joint IOC/WMO Workshop on Oceanographic Products and the IGOS Data Processing and Services System (IDPSS); Moscow, 9-11 April 1979.	E	34	IOC Workshop on Regional Co-operation in Marine Science in the Central Eastern Atlantic (Western Africa); Tenerife, 12-17 December 1963.	E, F, S	51	CCOP/SOPAC-IOC Workshop on Coastal Processes in the South Pacific Island Nations; Lae, Papua-New Guinea, 1-8 October 1987.	E
17	Papers submitted to the Joint Suppl. IOC/WMO Seminar on Oceanographic Products and the IGOS Data Processing and Services System; Moscow, 2-6 April 1979.	E	35	CCOP/SOPAC-IOC-UNU Workshop on Basic Geo-scientific Marine Research Required for Assessment of Minerals and Hydrocarbons in the South Pacific; Suva, Fiji, 3-7 October 1983.	E			

No.	Title	Languages	No.	Title	Languages	No.	Title	Languages
52	SCOR-IOC-UNESCO Symposium on Vertical Motion in the Equatorial Upper Ocean and its Effects upon Living Resources and the Atmosphere; Paris, 6-10 May 1985.	E	74	IOC-UNEP Review Meeting on Oceanographic Processes of Transport and Distribution of Pollutants in the Sea; Zagreb, Yugoslavia, 15-18 May 1989.	E	96	IOC-UNEP-WMO-SAREC Planning Workshop on an Integrated Approach to Coastal Erosion, Sea Level Changes and their Impacts; Zanzibar, United Republic of Tanzania, 17-21 January 1994.	E
53	IOC Workshop on the Biological Effects of Pollutants; Oslo, 11-29 August 1986.	E	75	IOC-SCOR Workshop on Global Ocean Ecosystem Dynamics; Solomons, Maryland, USA, 29 April-2 May 1991.	E	96	IOC-UNEP-WMO-SAREC	E
54	Workshop on Sea-Level Measurements in Hostile Conditions; Bidston, UK, 28-31 March 1988	E	76	IOC/WESTPAC Scientific Symposium on Marine Science and Management of Marine Areas of the Western Pacific; Penang, Malaysia, 2-6 December 1991.	E	Suppl. 1	Planning Workshop on an Integrated Approach to Coastal Erosion, Sea Level Changes and their Impacts; Submitted Papers	
55	IBCCA Workshop on Data Sources and Compilation, Boulder, Colorado, 18-19 July 1988.	E	77	IOC-SAREC-KMFRI Regional Workshop on Causes and Consequences of Sea-Level Changes on the Western Indian Ocean Coasts and Islands; Mombasa, Kenya, 24-28 June 1991.	E	96	IOC-UNEP-WMO-SAREC	E
56	IOC-FAO Workshop on Recruitment of Penaeid Prawns in the Indo-West Pacific Region (PREP); Cleveland, Australia, 24-30 July 1988.	E	78	IOC-CEC-ICES-WMO-ICSU Ocean Climate Data Workshop Goddard Space Flight Center, Greenbelt, Maryland, USA, 18-21 February 1992.	E	Suppl. 2	Planning Workshop on an Integrated Approach to Coastal Erosion, Sea Level Changes and their Impacts; Submitted Papers	
57	IOC Workshop on International Co-operation in the Study of Red Tides and Ocean Blooms; Takamatsu, Japan, 16-17 November 1987.	E	79	IOC/WESTPAC Workshop on River Inputs of Nutrients to the Marine Environment in the WESTPAC Region; Penang, Malaysia, 26-29 November 1991.	E	97	IOC Workshop on Small Island Oceanography in Relation to Sustainable Economic Development and Coastal Area Management of Small Island Development States; Fort-de-France, Martinique, 8-10 November 1993.	E
58	International Workshop on the Technical Aspects of the Tsunami Warning System; Novosibirsk, USSR, 4-5 August 1989.	E	80	IOC-SCOR Workshop on Programme Development for Harmful Algae Blooms; Newport, USA, 2-3 November 1991.	E	98	CoMSBlack '92A Physical and Chemical Intercalibration Workshop; Erdemli, Turkey, 15-29 January 1993.	E
58	Second International Workshop on the Technical Aspects of Tsunami Warning Systems, Tsunami Analysis, Preparedness, Observation and Instrumentation. Submitted Papers; Novosibirsk, USSR, 4-5 August 1989.	E	81	Joint IAPSO-IOC Workshop on Sea Level Measurements and Quality Control; Paris, 12-13 October 1992.	E	99	IOC-SAREC Field Study Exercise on Nutrients in Tropical Marine Waters; Mombasa, Kenya, 5-15 April 1994.	E
59	IOC-UNEP Regional Workshop to Review Priorities for Marine Pollution Monitoring Research, Control and Abatement in the Wider Caribbean; San José, Costa Rica, 24-30 August 1989.	E, F, S	82	BORDOMER 92: International Convention on Rational Use of Coastal Zones. A Preparatory Meeting for the Organization of an International Conference on Coastal Change; Bordeaux, France, 30 September-2 October 1992.	E	100	IOC-SOA-NOAA Regional Workshop for Member States of the Western Pacific - GODAR-II (Global Oceanographic Data Archeology and Rescue Project); Tianjin, China, 8-11 March 1994.	E
60	IOC Workshop to Define IOCARIBE-TRODERP proposals; Caracas, Venezuela, 12-16 September 1989.	E	83	IOC Workshop on Donor Collaboration in the Development of Marine Scientific Research Capabilities in the Western Indian Ocean Region; Brussels, Belgium, 12-13 October 1992.	E	101	IOC Regional Science Planning Workshop on Harmful Algal Blooms; Montevideo, Uruguay, 15-17 June 1994.	E
61	Second IOC Workshop on the Biological Effects of Pollutants; Bermuda, 10 September-2 October 1988.	E	84	Workshop on Atlantic Ocean Climate Variability; Moscow, Russian Federation, 13-17 July 1992.	E	102	First IOC Workshop on Coastal Ocean Advanced Science and Technology Study (COASTS); Liège, Belgium, 5-9 May 1994.	E
62	Second Workshop of Participants in the Joint FAO-IOC-WHO-IAEA-UNEP Project on Monitoring of Pollution in the Marine Environment of the West and Central African Region; Accra, Ghana, 13-17 June 1988.	E	85	IOC Workshop on Coastal Oceanography in Relation to Integrated Coastal Zone Management; Kona, Hawaii, 1-5 June 1992.	E	103	IOC Workshop on GIS Applications in the Coastal Zone Management of Small Island Developing States; Barbados, 20-22 April 1994.	E
63	IOC/WESTPAC Workshop on Co-operative Study of the Continental Shelf Circulation in the Western Pacific; Bangkok, Thailand, 31 October-3 November 1989.	E	86	International Workshop on the Black Sea; Varna, Bulgaria, 30 September - 4 October 1991.	E	104	Workshop on Integrated Coastal Management; Dartmouth, Canada, 19-20 September 1994.	E
64	Second IOC-FAO Workshop on Recruitment of Penaeid Prawns in the Indo-West Pacific Region (PREP); Phuket, Thailand, 25-31 September 1989.	E	87	Taller de trabajo sobre efectos biológicos del fenómeno «El Niño» en ecosistemas costeros del Pacífico Sudeste; Santa Cruz, Galápagos, Ecuador, 5-14 de octubre de 1989.	S only (Summary in E, F, S)	105	BORDOMER 95: Conference on Coastal Change; Bordeaux, France, 6-10 February 1995.	E
65	Second IOC Workshop on Sardine/Anchovy Recruitment Project (SARP) in the Southwest Atlantic; Montevideo, Uruguay, 21-23 August 1989.	E	88	IOC-CEC-ICSU-ICES Regional Workshop for Member States of Eastern and Northern Europe (GODAR Project); Obninsk, Russia, 17-20 May 1993.	E	105	Conference on Coastal Change: Proceedings; Bordeaux, France, 6-10 February 1995	E
66	IOC ad hoc Expert Consultation on Sardine/Anchovy Recruitment Programme; La Jolla, California, USA, 1989.	E	89	IOC-ICESEM Workshop on Ocean Sciences in Non-Living Resources; Perpignan, France, 15-20 October 1990.	E	Suppl.		
67	Interdisciplinary Seminar on Research Problems in the IOCARIBE Region; Caracas, Venezuela, 28 November-1 December 1989.	E (out of stock)	90	IOC Seminar on Integrated Coastal Management; New Orleans, USA, 17-18 July 1993.	E	106	IOC/WESTPAC Workshop on the Paleographic Map; Bali, Indonesia, 20-21 October 1994.	E
68	International Workshop on Marine Acoustics; Beijing, China, 26-30 March 1990.	E	91	Hydroblack'91 CTD Intercalibration Workshop; Woods Hole, USA, 1-10 December 1991.	E	107	IOC-ICSU-NIO-NOAA Regional Workshop for Member States of the Indian Ocean - GODAR-III; Dona Paula, Goa, India, 6-9 December 1994.	E
69	IOC-SCAR Workshop on Sea-Level Measurements in the Antarctica; Leningrad, USSR, 28-31 May 1990.	E	92	Réunion de travail IOCEA-OSNLR sur le Projet « Budgets sédimentaires le long de la côte occidentale d'Afrique » Abidjan, Côte d'Ivoire, 26-28 juin 1991.	F	108	UNESCO-IHP-IOC-IAEA Workshop on Sea-Level Rise and the Multidisciplinary Studies of Environmental Processes in the Caspian Sea Region; Paris, 9-12 May 1995.	E
69	IOC-SCAR Workshop on Sea-Level Measurements in the Antarctica; Submitted Papers; Leningrad, USSR, 28-31 May 1990.	E	93	IOC-UNEP Workshop on Impacts of Sea-Level Rise due to Global Warming; Dhaka, Bangladesh, 16-19 November 1992.	E	Suppl.	Workshop on Sea-Level Rise and the Multidisciplinary Studies of Environmental Processes in the Caspian Sea Region; Submitted Papers; Paris, 9-12 May 1995.	
70	IOC-SAREC-UNEP-FAO-IAEA-WHO Workshop on Regional Aspects of Marine Pollution; Mauritius, 29 October - 9 November 1990.	E	94	BMTC-IOC-POLARIMAR International Workshop on Training Requirements in the Field of Eutrophication in Semi-Enclosed Seas and Harmful Algal Blooms; Bremerhaven, Germany, 29 September - 3 October 1992.	E	109	First IOC-UNEP CEPOL Symposium; San José, Costa Rica, 14-15 April 1993.	E
71	IOC-FAO Workshop on the Identification of Penaeid Prawn Larvae and Postlarvae; Cleveland, Australia, 23-28 September 1990.	E	95	SAREC-IOC Workshop on Donor Collaboration in the Development of Marine Scientific Research Capabilities in the Western Indian Ocean Region; Brussels, Belgium, 23-25 November 1993.	E	110	IOC-ICSU-CEC Regional Workshop for Member States of the Mediterranean - GODAR-IV (Global Oceanographic Data Archeology and Rescue Project) Foundation for International Studies, University of Malta, Valletta, Malta, 25-28 April 1995.	E
72	IOC/WESTPAC Scientific Steering Group Meeting on Co-Operative Study of the Continental Shelf Circulation in the Western Pacific; Kuala Lumpur, Malaysia, 9-11 October 1990.	E						
73	Expert Consultation for the IOC Programme on Coastal Ocean Advanced Science and Technology Study; Liège, Belgium, 11-13 May 1991.	E						

No.	Title	Languages	No.	Title	Languages	No.	Title	Languages
111	Chapman Conference on the Circulation of the Intra-Americas Sea; La Parguera, Puerto Rico, 22-26 January 1995.	E	117	Joint IOC-CIDA-Sida (SAREC) Workshop on the Benefits of Improved Relationships between International Development Agencies, the IOC and other Multilateral Intergovernmental Organizations in the Delivery of Ocean, Marine Affairs and Fisheries Programmes; Sidney B.C., Canada, 26-28 September 1995.	E	122	IOC-UNEP-PERSGA-ACOPS-IUCN Workshop on Oceanographic Input to Integrated Coastal Zone Management in the Red Sea and Gulf of Aden Jeddah, Saudi Arabia, 8 October 1995.	E
112	IOC-IAEA-UNEP Group of Experts on Standards and Reference Materials (GESREM) Workshop; Miami, USA, 7-8 December 1993.	E	118	IOC-UNEP-NOAA-Sea Grant Fourth Caribbean Marine Debris Workshop; La Romana, Santo Domingo, 21-24 August 1995.	E	123	Second IOC Regional Science Planning Workshop on Harmful Algal Blooms in South America; Mar del Plata, Argentina, 30 October - 1 November 1995.	E/S
113	IOC Regional Workshop on Marine Debris and Waste Management in the Gulf of Guinea; Lagos, Nigeria, 14-16 December 1994.	E	119	IOC Workshop on Ocean Colour Data Requirements and Utilization; Sydney B.C., Canada, 21-22 September 1995.	E	124	GLOBEC-IOC-SAHFOS-MBA Workshop on the Analysis of Time Series with Particular Reference to the Continuous Plankton Recorder Survey; Plymouth, U.K., 4-7 May 1993.	E
114	International Workshop on Integrated Coastal Zone Management (ICZM) Karachi, Pakistan; 10-14 October 1994.	E	120	International Training Workshop on Integrated Coastal Management; Tampa, Florida, U.S.A., 15-17 July 1995.	E			
115	IOC/GLOSS-IAPSO Workshop on Sea Level Variability and Southern Ocean Dynamics; Bordeaux, France, 31 January 1995.	E	121	Atelier régional sur la gestion intégrée des zones littorales (ICAM); Conakry, Guinée, 12-22 décembre 1995.	F			
116	IOC/WESTPAC International Scientific Symposium on Sustainability of Marine Environment: Review of the WESTPAC Programme, with Particular Reference to ICAM Bali, Indonesia, 22-26 November 1996	E						