

Common trends in northeast Atlantic squid time series

A.F. Zuur^{a,*}, G.J. Pierce^b

^aStatistics Group, FRS Marine Laboratory, PO Box 101, Victoria Road, Aberdeen, AB11 9DB, UK

^bDepartment of Zoology, University of Aberdeen, Tillydrone Avenue, Aberdeen AB24 2TZ, UK

Received 10 December 2002; accepted 14 August 2003

Abstract

In this paper, dynamic factor analysis is used to estimate common trends in time series of squid catch per unit effort in Scottish (UK) waters. Results indicated that time series of most months were related to sea surface temperature measured at Millport (UK) and a few series were related to the NAO index. The DFA methodology identified three common trends in the squid time series not revealed by traditional approaches, which suggest a possible shift in relative abundance of summer- and winter-spawning populations.

Crown Copyright © 2004 Published by Elsevier B.V. All rights reserved.

Keywords: Squid time series; Common trends; Dynamic factor analysis; Sea surface temperature; NAO index

1. Introduction

The genus *Loligo* supports numerous fisheries around the world. *Loligo forbesi* is the most important cephalopod fishery resource in Scottish (UK) waters and has been the focus of many studies of life history, distribution and abundance (Boyle and Pierce, 1994; Collins et al., 1997, 1999; Pierce et al., 1998, 2001; Waluda and Pierce, 1998; Bellido et al., 2001). Two recent publications have addressed questions about temporal trends in abundance of *L. forbesi*: Sims et al. (2001) showed that the timing of migration into the English Channel, as revealed by survey catches, could be related to sea surface temperature (SST), with higher temperatures being associated with earlier migration. Pierce and Boyle (2003) analysed temporal

patterns in Scottish fishery data, demonstrating that SST influenced winter abundance in the North Sea. Nevertheless, while relationships between distribution patterns and environmental factors in squid are well-known and traditional auto-regressive integrated moving average (ARIMA) models have been fitted to abundance time series, reliable fishery forecasts remain out of reach.

In a more general context, there have been many instances in which ARIMA models (Ljung, 1987) have been applied to fishery time series with varying degrees of success (e.g. Fogarty, 1989; Stergiou, 1989, 1991; Pajuelo and Lorenzo, 1995; Stergiou and Christou, 1996; Stergiou et al., 1997; Park, 1998; Pierce and Boyle, 2003; Georgakarakos et al., 2002). However, ARIMA models require long stationary time series but most fisheries time series are short and non-stationary. To overcome the last problem, time series can be de-trended or integrated series can be analysed. However, in most fisheries studies

* Corresponding author.

E-mail address: a.zuur@marlab.ac.uk (A.F. Zuur).

the main question is related to the pattern of trends over time in abundance. ARIMA models are not designed for this purpose.

Zuur et al. (2003a) introduced a technique called dynamic factor analysis (DFA) and used it to estimate common trends and effects of explanatory variables in a zoobenthic time-series data set. Zuur et al. (2003b) gave a non-technical introduction and applied the method on landings per unit effort *Nephrops* series from Europe. A slightly modified version of DFA was used in Mendelsohn and Schwing (1997, 2002) to estimate common trends in sea surface temperature and wind stress time series. Here, DFA is used to estimate common trends of catch per unit effort for squid (*Loligo forbesi*) in ICES fishery subdivisions IVa (northern North Sea) and VIa (west coast of Scotland), using data from Pierce and Boyle (2003).

2. Material and methods

2.1. Data used

The data consist of monthly average trawler landings per unit effort (LPUE) of Squid (*Loligo forbesi*) for the two ICES fisheries subdivisions IVa (northern North Sea) and VIa (West Coast of Scotland; see Fig. 1) from 1970 to 1999. Because squid is basically a by-catch of the whitefish fishery, fishing effort is assumed to be independent of squid abundance. Furthermore, the relatively high value of squid means that catches are rarely discarded. Hence we assume that LPUE is equivalent to CPUE and will be a reasonable index of abundance (Pierce and Boyle, 2003). As to possible explanatory variables, various sea surface temperature (SST) series from around the UK are available, namely Millport, Peterhead, Fair

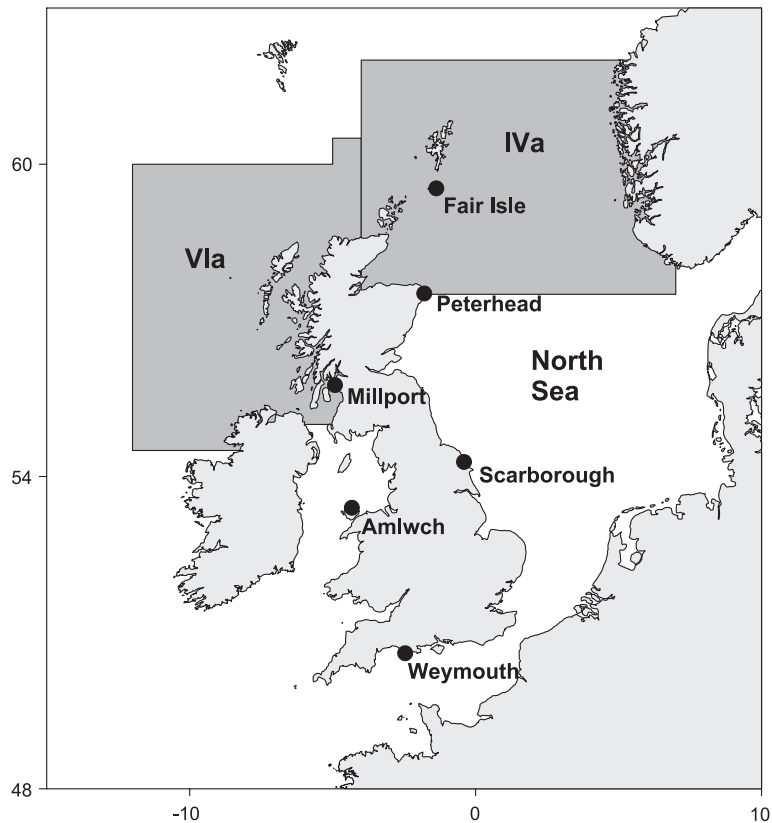


Fig. 1. Geographical location of the ICES blocks IVa and VIa.

Isle, Scarborough, Amwlch and Weymouth (Fig. 1). A few of these SST series were available from 1970 onwards but unfortunately most of them were recorded later (1978 onwards). In a preliminary analysis, dynamic factor analysis was applied to these six winter sea surface time series. Results indicated that all these time series can be described by one common trend. Therefore, the SST measured at Millport, which is the most complete and longest time series, is used here. Another explanatory variable used in the analysis is the North Atlantic Oscillation (NAO) index (Hurrell, 1995). This index is defined as the difference between sea level atmospheric pressure at the Azores and Iceland. A relatively high index indicates strong westerly winds; a low index indicates weak westerly winds.

2.2. Seasonal and trend analysis

To analyse seasonal data, one of the usual time-series models is of the form:

$$Y(t) = T(t) + S(t) + I(t) + e(t)$$

Where $Y(t)$ is a univariate time series, $T(t)$ is the trend, $S(t)$ the time-dependent seasonal component, $I(t)$ can contain cycles, explanatory variables or autoregressive terms and $e(t)$ is the error. Technical details on this model, also called the STL procedure) can be found in [Hastie and Tibshirani \(1990\)](#) and [Cleveland et al. \(1990\)](#). If $I(t)$ contains explanatory variables, STL (Seasonal and Trend decomposition using Loess) is basically a composition of two smoothing functions, $T(t)$ and $S(t)$, a parametric component $I(t)$ and an unexplained term called noise. If $I(t)$ is a cyclic component, STL is a combination of three smoothing functions and a noise term. Technically, these smoothing functions are estimated by the so-called backfitting algorithm ([Hastie and Tibshirani, 1990](#)). To separate the different time effects, each smoothing function has a certain amount of smoothing. Although ad hoc procedures are in place, it is possible that for example the trend and seasonal components are confounded. Dynamic factor analysis can be seen as a multivariate extension of the STL model $Y(t) = T(t) + I(t) + e(t)$. In principle, the dynamic factor model can also be used to estimate common

trends, common seasonal patterns and common cycles simultaneously ([Harvey, 1989](#)), but it is likely that such a model results in computational problems ([Zuur et al., 2003a](#)).

2.3. Seasonality

An alternative data analysis approach is to remove the seasonal component of the data prior to the analysis. Although de-seasonalising can be done in various ways ([Brockwell and Davis, 1996](#)), none of the approaches are without problems. For example, for monthly data one option is to calculate monthly averages using data from all years. The time series are then corrected by simply subtracting the appropriate monthly average. However, this has the disadvantage that it does not allow for changes in the monthly patterns. The same problem occurs if the monthly data are modelled as a parametric cosine function or as dummy (nominal) variables ([Harvey, 1989](#)). Yet another option is used in [Shumway \(2000\)](#), who presented an example of a 43-year river flow time series. The aim of that example was to derive monthly mean parameters under serial correlation, and common monthly profiles were estimated.

The two squid CPUE series and the SST and NAO series were available on a monthly basis. Cross-correlations between all series were significantly different from zero at the 5% level. However, this is only because all series show a strong seasonal pattern. To deal with the seasonal aspects of the data, a similar approach as in [Shumway \(2000\)](#) was adopted. CPUE squid time series for each month were created, resulting in 12 time series for each of the two areas. The first time series contains all CPUE values measured in January in area IVa, the second time series all CPUE series measured in February, etc. Hence, we have 24 time series. It is then interesting to analyse which of the monthly time series follow similar patterns over time, and which not. Explanatory variables like SST and the NAO can be incorporated into the model and show the effects in certain months. Obviously, this data analysis approach is not without problems either; the common trends obtained by DFA applied to the 24 time series are now common patterns in monthly time series (or in a subset) and

therefore confounding between long-term trends in the original two time series and seasonal changes is still possible.

2.4. Normality

Although DFA takes account of serial correlation, it is based on normality of the data. Zuur et al. (2003b) stated that normality of the data is beneficial in DFA, but it is not of fundamental importance. The same holds for linear regression models (Quinn and Keough, 2002). To obtain data that are approximately normally distributed, various data transformations were applied. QQ-plots and histograms were made for the original 24 time series, and for square root, cubic root and \log_{10} transformed data. All plots indicated that a transformation results in ‘more normally distributed’ data, the square root being favourable in most cases. The additional benefit of a transformation is that it reduces the influence of large observations.

Zuur et al. (2003a) showed that the factor loadings obtained by applying DFA on standardised time series and on non-standardised time series are identical up to a matrix multiplication. This matrix is diagonal and contains the standard deviations of the original time series. For the ‘what is going on’ question it is advisable to standardise the time series because interpretation of the factor loadings (and therefore the common trends) will be easier if the time series are standardised, especially if the original time series show different variation. If the aim of the analysis is prediction on the original scale of the data, then either a post-matrix multiplication is required if the series were standardised, or the original data can be analysed by DFA.

2.5. Dynamic factor analysis

A detailed statistical description of dynamic factor analysis is given in Zuur et al. (2003a). Because of the technical nature of that paper, an introduction on an intermediate statistical level was given in Zuur et al. (2003b). Here we give an explanation in layman’s terms. To explain the underlying principle of dynamic factor analysis, we first need to describe a related technique called structural time-series models (Harvey, 1989). For short time series measured on an

annual basis the structural time-series model, in words, is given by:

$$\begin{aligned} 1 \text{ time series} &= \text{constant} + \text{trend} \\ &+ \text{explanatory variables} + \text{noise} \end{aligned} \quad (1)$$

The trend is modelled as a smoothing curve and is therefore not restricted to being a straight line. Technically, it is modelled as a random walk (Chatfield, 1989; Harvey, 1989) and estimation is carried out with the Kalman filter/smoothing algorithm. The ‘explanatory variables’ component is modelled as in linear regression, namely via $\mathbf{b}\mathbf{x}_t$, where \mathbf{x}_t contains the value of the k explanatory variables at time t and \mathbf{b} contains the k unknown regression coefficients (see Appendix for further technical details). The constant is the intercept term in a regression model.

If there are N time series, the model in (1) can be applied to each time series. However, this means that N model validations have to be carried out, each containing residual plots, model fits, numerical output, etc. Furthermore, interactions between the N series are ignored. Dynamic factor analysis overcomes these two problems. Instead of N time series, a linear combination of M common trends is estimated, where M is (ideally) much smaller than N . The underlying model, in words, is given by:

$$\begin{aligned} N \text{ time series} &= \text{constant} \\ &+ \text{linear combination of } M \text{ common trends} \\ &+ \text{explanatory variables} + \text{noise} \end{aligned}$$

To illustrate the principle, the formulation for a model containing two common trends is presented. Suppose the data set contains 5 time series ($N=5$). The dynamic factor model with a constant, two common trends ($M=2$) and one explanatory variable is given by:

$$y_{1t} = c_1 + a_{11} z_{1t} + a_{12} z_{2t} + b_1 x_t + \text{noise}_{1t}$$

$$y_{2t} = c_2 + a_{21} z_{1t} + a_{22} z_{2t} + b_2 x_t + \text{noise}_{2t}$$

$$y_{3t} = c_3 + a_{31} z_{1t} + a_{32} z_{2t} + b_3 x_t + \text{noise}_{3t}$$

$$y_{4t} = c_4 + a_{41} z_{1t} + a_{42} z_{2t} + b_4 x_t + \text{noise}_{4t}$$

$$y_{5t} = c_5 + a_{51} z_{1t} + a_{52} z_{2t} + b_5 x_t + \text{noise}_{5t}$$

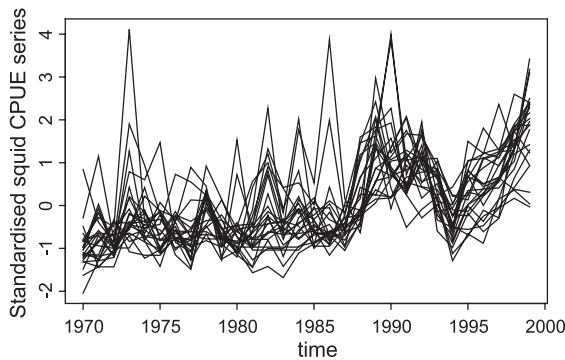


Fig. 2. Square-root transformed, standardised squid CPUE time series for months 1–12 of ICES blocks IVa and VIa. Because the time series are standardised, the y-axis is unitless and the range of the values along the y-axis is approximately between -3 and 3 .

where: y_{1t} is the value of the first time series at time t , z_{1t} and z_{2t} represent the values of the two common trends at time t , x_t is the value of the explanatory variable, a_{i1} and a_{i2} are factor loadings and show which of the time series are related to a particular common trend, b_i and its standard error indicate the influence of the explanatory variable on the i th series and c_i is a constant parameter as in a linear regression model. If the time series are standardised, the constant parameters are 0.

2.6. Matrix diagonality

The dynamic factor model can be written in matrix notation as:

$$y_t = \mathbf{A} z_t + \mathbf{B} x_t + e_t \quad (2)$$

where the $N \times M$ matrix \mathbf{A} contains the factor loadings, \mathbf{B} (of dimension $N \times L$) the regression parameters and e_t the noise components. Zuur et al. (2003a) included the vector of constant parameters into the regression component. It is assumed that $e_t \sim N(\mathbf{0}, \mathbf{R})$. The common trends in DFA are restricted to being smooth functions over time via the random walk formulation (see Appendix). Note that the M common trends represent ‘trends’ that are common to each series. The magnitude and sign of the factor loadings determine how these trends are related to the original time series.

As to the modelling of the covariance matrix \mathbf{R} , the easiest approach is to use a diagonal matrix. An

alternative option is to use a positive-definite, symmetric, non-diagonal matrix for \mathbf{R} . Off-diagonal elements of \mathbf{R} represent the joint information in two response variables, which cannot be explained with the other terms. Models with similar error structures exist in linear regression, namely generalised least squares models (Johnston and Dinardo, 1997).

2.7. Model fits

Various models can be applied: models containing one, two, three or more common trends; models with and without explanatory variables and constant parameters; models with a diagonal or non-diagonal error covariance matrix \mathbf{R} . To decide which model fits the data ‘best’, Zuur et al. (2003a, b) used Akaike’s information criterion (AIC). The AIC is defined as the difference between a measure of fit (maximum likelihood) and the number of parameters (number of trends, explanatory variables and structure of \mathbf{R}). If more parameters are used, the model fit is better, but the penalty for the extra parameters is higher as well. The AIC tries to find a balance between these two components. Further details on the AIC and alternative model selection methods were discussed in both papers. In this paper, the AIC is used as well and the dynamic factor analysis model with the smallest AIC values is taken to be the ‘best’ candidate model.

Results presented in this paper were obtained with the software package Brodgar (<http://www.brodgar.com>).

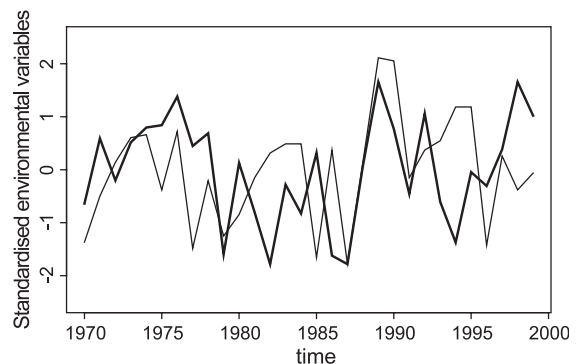


Fig. 3. Sea surface temperature at Millport (thick line) and the NAO index. Both time series were standardised and therefore the y-axis is unitless and the range of the values along the y-axis is approximately between -3 and 3 .

Table 1
Four dynamic factor models used in the analysis

No	Model	R
1	24 CPUE series = constant + M common trends + noise	diagonal
2	24 CPUE series = constant + M common trends + explanatory variables + noise	diagonal
3	24 CPUE series = constant + M common trends + noise	non-diagonal
4	24 CPUE series = constant + M common trends + explanatory variables + noise	non-diagonal

R is the error covariance matrix.

2.8. Results

The transformed and standardised squid CPUE series are presented in Fig. 2. Note that from 1987, all CPUE series follow approximately similar patterns. Furthermore, there are a few CPUE time series having large values in 1973, 1986 and 1991. These values do not seem to be outliers because more than

one time series had relatively high values in these years. As to explanatory variables, we took averages of January, February and March for the SST series and NAO index and these are presented in Fig. 3. We compared these series with averages of other months but the patterns were similar.

To estimate the underlying common trends, four sets of dynamic factor models were used (Table 1). In the first set, the 24 squid CPUE series are modelled as a constant plus a linear combination of M common trends and a noise term, which is assumed to be normally distributed, with expectation $\mathbf{0}$ and a diagonal covariance matrix R . Different values of M were used (labelled as model 1a, 1b, etc.). We then added explanatory variables, and each combination of explanatory variables and common trends was tried (models 2a–2s). In model sets 3 and 4 the same approach was followed as in 1 and 2, except that the covariance matrix was modelled as a non-diagonal, symmetric positive definite matrix. The AIC value of each model is given in Table 2. The AIC values of all models containing a non-diagonal matrix

Table 2
AIC values obtained by applying dynamic factor models 1 to 4 on the CPUE Squid time series

Model	M	AIC	Explanatory variable	Model	M	AIC	Explanatory variable
1a	1	1503.33		2o	2	1363.32	NAO & SST
1b	2	1356.53		2p	3	1358.22	NAO & SST
1c	3	1306.94		2q	4	1318.83	NAO & SST
1d	4	1292.54		2r	5	1300.83	NAO & SST
1e	5	1284.34		2s	6	1305.72	NAO & SST
1f	6	1281.05		3a	1	1231.58	
1g	7	1289.30		3b	2	1210.05	
2a	1	1487.59	NAO	3c	3	1205.67	
2b	2	1357.69	NAO	3d	4	1215.39	
2c	3	1329.04	NAO	4a	1	1210.22	NAO
2d	4	1313.13	NAO	4b	2	1190.80	NAO
2e	5	1300.53	NAO	4c	3	1192.54	NAO
2f	6	1295.75	NAO	4d	1	1204.95	SST
2g	7	1297.16	NAO	4e	2	1185.62	SST
2h	1	1520.64	SST	4f	3	1171.46	SST
2i	2	1360.38	SST	4g	4	1180.05	SST
2j	3	1315.28	SST	4h	1	1188.00	NAO & SST
2k	4	1298.30	SST	4i	2	1164.88	NAO & SST
2l	5	1287.12	SST	4j	3	1154.19	NAO & SST
2m	6	1293.33	SST	4k	4	1163.63	NAO & SST
2n	1	1506.58	NAO & SST				

M is the number of common trends. The maximum number of common trends varies between different types of models because an increasing AIC indicates a sub-optimal model (applying models with larger values of M once the AIC is increasing will lead to sub-optimal models as judged by the AIC).

R were much smaller, which indicates that there is a certain amount of information in the CPUE series that cannot be explained with common trends and explanatory variables. The AIC values indicate that model 4j, containing three common trends, the NAO and SST as explanatory variable and a non-diagonal matrix R , was the optimal model. Results obtained by this model are discussed next.

The estimated regression parameters for the SST and NAO series are given in Tables 3 and 4. Estimated parameters with relatively large t -values (in absolute sense) indicate strong relationships. For area IVa, the t -values indicated that SST is significantly related (at the 5% level) to eight CPUE series. The larger t -values were obtained for the spring months. For area VIa, SST was also significantly related (at the 5% level) to eight CPUE series. The NAO index was only significantly related (at the 5% level) to five of the 24 time series, and mainly to the IVa series.

The estimated common trends are presented in Fig. 4. The first common trend is characterised by an increase from 1984 onwards. The second trend shows a steady increase between 1970–1985 and a sharp decrease until 1994, followed by an increase. The third common trend increases until 1990 and sharply decreases thereafter. Factor loadings can be used to see which common trends are related to the original time series (Fig. 4, right-hand figures). The factor loadings can either be presented in a scatter plot or as vertical lines along each axis. We followed

Table 3
Estimated regression parameters and t -values for the SST series

IVa series	estimated value	t -value	VIa series	Estimated value	t -value
1	0.22	2.19	1	0.10	1.18
2	0.21	1.28	2	0.27	2.36
3	0.33	3.26	3	0.42	3.31
4	0.37	4.41	4	0.28	2.12
5	0.30	3.58	5	0.23	1.96
6	0.33	3.27	6	0.28	2.06
7	0.42	3.81	7	-0.04	-0.25
8	0.25	2.22	8	0.17	1.31
9	0.31	2.43	9	0.55	4.89
10	0.22	1.71	10	0.36	2.46
11	0.22	1.85	11	0.45	4.64
12	0.18	1.66	12	0.32	3.75

The cross-correlation between SST and the NAO is 0.26.

Table 4
Estimated regression parameters and t -values for the NAO series

IVa series	estimated value	t -value	VIa series	Estimated value	t -value
1	-0.02	-0.19	1	-0.02	-0.20
2	0.09	0.52	2	-0.14	-1.26
3	0.00	-0.05	3	-0.11	-0.84
4	-0.09	-1.08	4	0.08	0.57
5	0.02	0.28	5	0.13	1.13
6	0.17	1.68	6	0.09	0.71
7	0.18	1.62	7	0.28	1.83
8	0.37	3.24	8	0.29	2.17
9	0.37	2.91	9	0.19	1.74
10	0.42	3.21	10	0.13	0.91
11	0.39	3.38	11	0.09	0.90
12	0.19	1.84	12	-0.02	-0.21

The cross-correlation between SST and the NAO is 0.26.

the latter approach. To limit text on the graphs, months are represented by numbers (1 refers to January, 12 to December). Results indicate roughly that the first common trend is important for the series 1–6 and 11–12 in both areas (except for February in area IVa). The second trend is related to months 10–12 in area IVa and 6–11 in area VIa. The third trend is important for months 7–10 in area IVa and 9–12 in area VIa. Besides factor loadings, correlations between the original series and the common trends can be calculated. In techniques such as discriminant analysis (Hair et al., 1998; Huberty, 1994) and canonical correspondence analysis (Ter Braak, 1986), these are called canonical correlations. For these data the canonical correlations and factor loadings showed similar information, and therefore canonical correlations were not presented. The fitted curves of the model are given in Fig. 5. Close inspection of these curves indicated that all CPUE series were fitted reasonably well, except for a few years with extremely high abundances (e.g. 1973 for month 2 in IVa).

The elements of the error covariance matrix R obtained by the model show the ‘misfits’ of the fitted model. Large diagonal elements mean that the corresponding time series are not fitted well, and large off-diagonal elements are an indication that the two corresponding time series have a joint pattern in common that is not fitted by the model. R can either be inspected or visualised by multidimensional

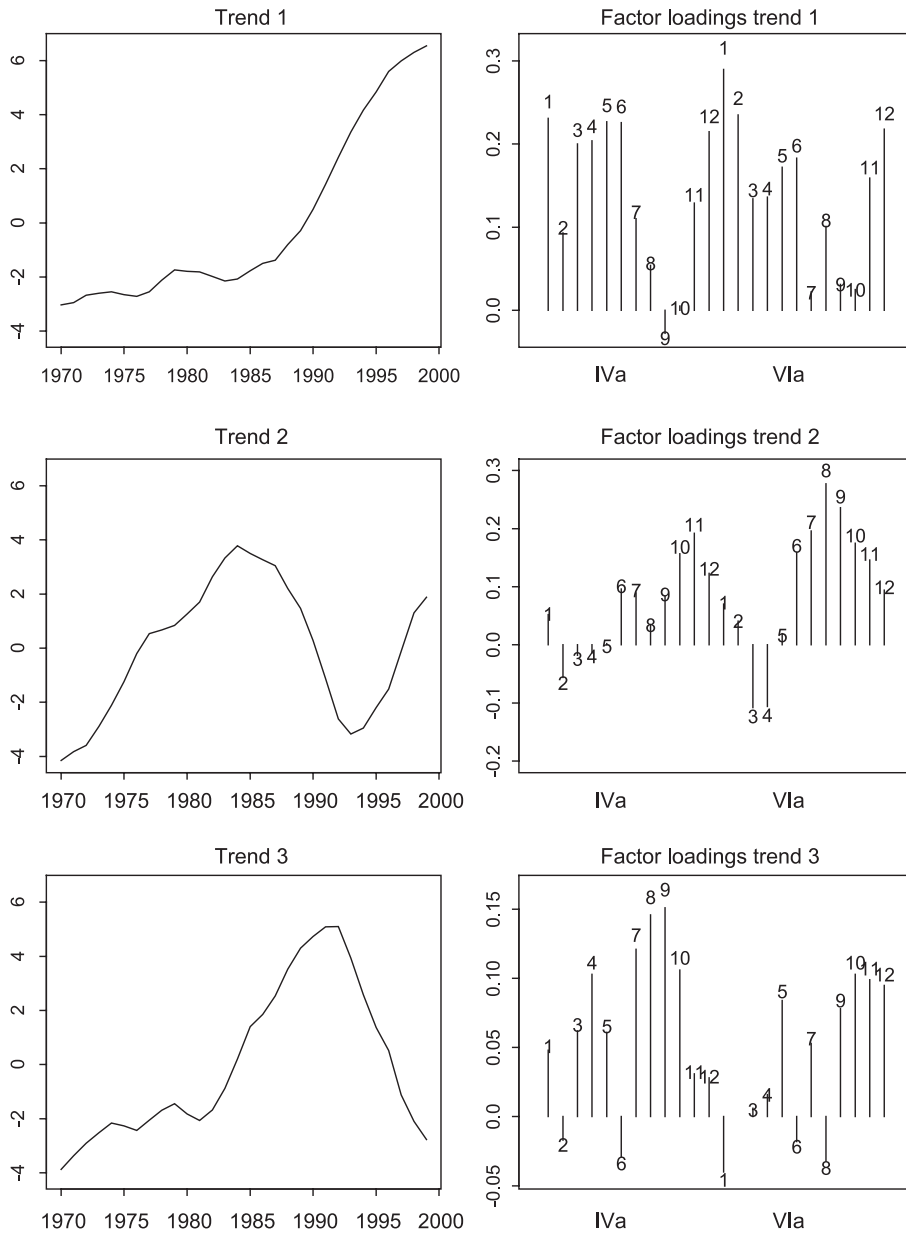


Fig. 4. Results for model 4j. The first row of figures contains the first common trend (left), factor loadings (right). The second and third rows contain the graphs for the second and third common trend, respectively. Months are represented by numbers (1 refers to January, 12 to December). The first 12 numbers refer to area IVa and the second twelve to area VIa. Because the time series are standardised, all the y-axes are unitless and the constant parameters are zero. Factor loadings are represented as vertical lines from 0 to the estimated value. The first vertical line corresponds to the factor loading of the January time series of IVa, the second to February, etc.

mensional scaling (Zuur et al., 2003a). Results are presented in Fig. 6 and suggest that there is a certain amount of information in the residuals that

cannot be explained with the SST, NAO, nor with the common trends. The first axis shows differences between months 9–12 and the other months, where-

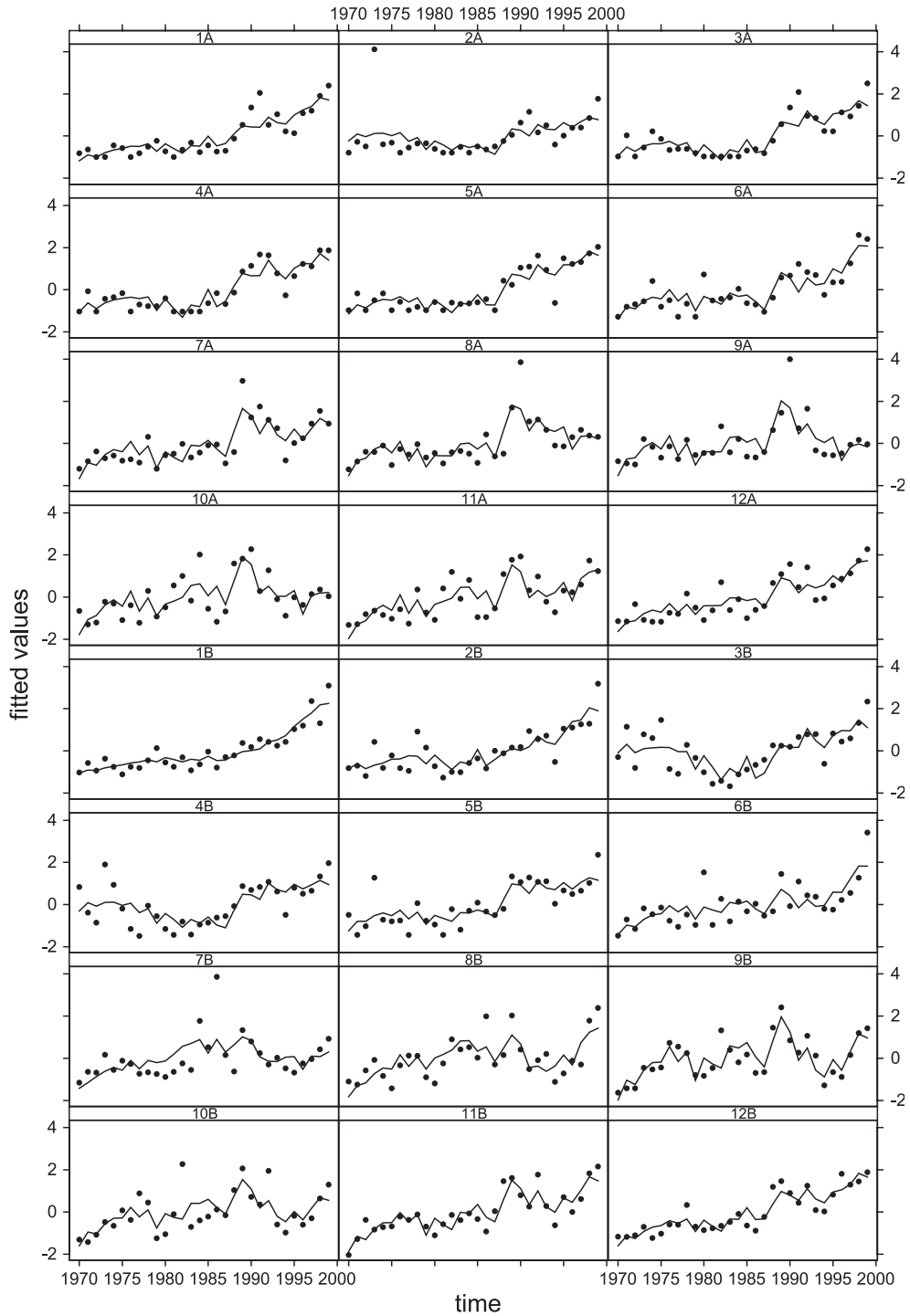


Fig. 5. Fitted values (line) and standardised observed values of CPUE Squid series. Months are represented by numbers (1 refers to January, 12 to December) and A and B refer to areas IVa and VIa, respectively. Because the time series are standardised, the y-axis is unitless.

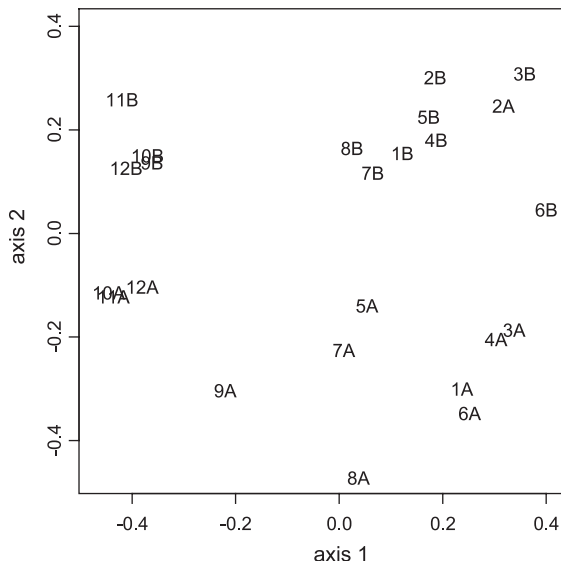


Fig. 6. Multidimensional scaling applied on the error covariance matrix. Months are represented by numbers (1 refers to January, 12 to December) and A and B refer to areas IVa and VIa, respectively.

as the second axis indicates a difference between the two areas.

To simplify the biological interpretation of the optimal model, results of models 4h (one common trend) and 4i (two common trends) are presented in Fig. 7. Models 4h and 4i use SST and the NAO index as explanatory variables and a non-diagonal matrix \mathbf{R} . The number of significant regression parameters for SST and the NAO is similar in models 4h, 4i and 4j. On the first row of Fig. 7, two graphs are presented, namely a common trend (left) and factor loadings (right). These graphs belong to model 4h (one common trend). The trend itself shows a general increase. Hence, besides the strong effect of SST, there is a general increase in most of the CPUE series, except for 7–10 in both areas (see factor loadings). The second and third rows of Fig. 7 contain similar information but now for model 4j. The second row contains the graphs for the first common trend, the last row for the second trend. The first common trend mirrors (roughly) the pattern of the trend in model 4h. It is also important for most CPUE series except for months 7–10. The second common trend shows a high peak around 1985 and a dip in the early 1990s. Based on factor loadings, this trend is important for

months (roughly) 9–11 for both areas. Looking at the original series, the high peak can indeed be detected in the CPUE series of months 9–11 in both areas. Fig. 8 shows five graphs. The first two graphs contain the SST at Millport and the NAO index plus the squid time series which were strongly related (Table 3) to these explanatory variables. All series were standardised. The remaining three graphs contain the common trends (obtained by model 4j) and squid time series with relatively large factor loadings for these trends. Again, all series were standardised. First of all, note that the SST series in Fig. 8 follows a pattern almost identical that of the CPUE series from 1987 onwards. This probably explains the large number of significant relationships between SST and the CPUE series. The first common trend reflects the general increase in squid series from 1985 onwards.

Using the information in Figs. 7 and 8, interpretation of the results of model 4j becomes slightly easier. SST is important for most series and explains the major variation in the series from 1987 onwards. The first trend is the most important (unexplained) common signal in the data, and shows a general increase. The second trend is mainly used to model high values for some of the series (months 7–10) around 1985. The third trend in the optimal model 4j is used to model high values for some of the CPUE series in the early 1990s. Hence, model 4j splits up the ‘months 7–10 behaviour’ of trend 2 in model 4i in slightly different patterns.

2.9. Discussion

Dynamic factor analysis was used to analyse multivariate non-stationary squid fishery CPUE time series. The dynamic factor model containing three common trends and SST and the NAO index as explanatory variables was the optimal model, as judged by the AIC. Results indicated that eight of the IVa series and eight of the VIa squid series were related to the SST series. Only four of the IVa series and one of the VIa series were related to the NAO index. However, differences between model 4j (SST and NAO as explanatory variables) and 4f (only SST as an explanatory variable) are large enough to justify using the NAO index in the model (Jones, 1993).

In DFA the order of the common trends is not related to their importance. However, the shape of the

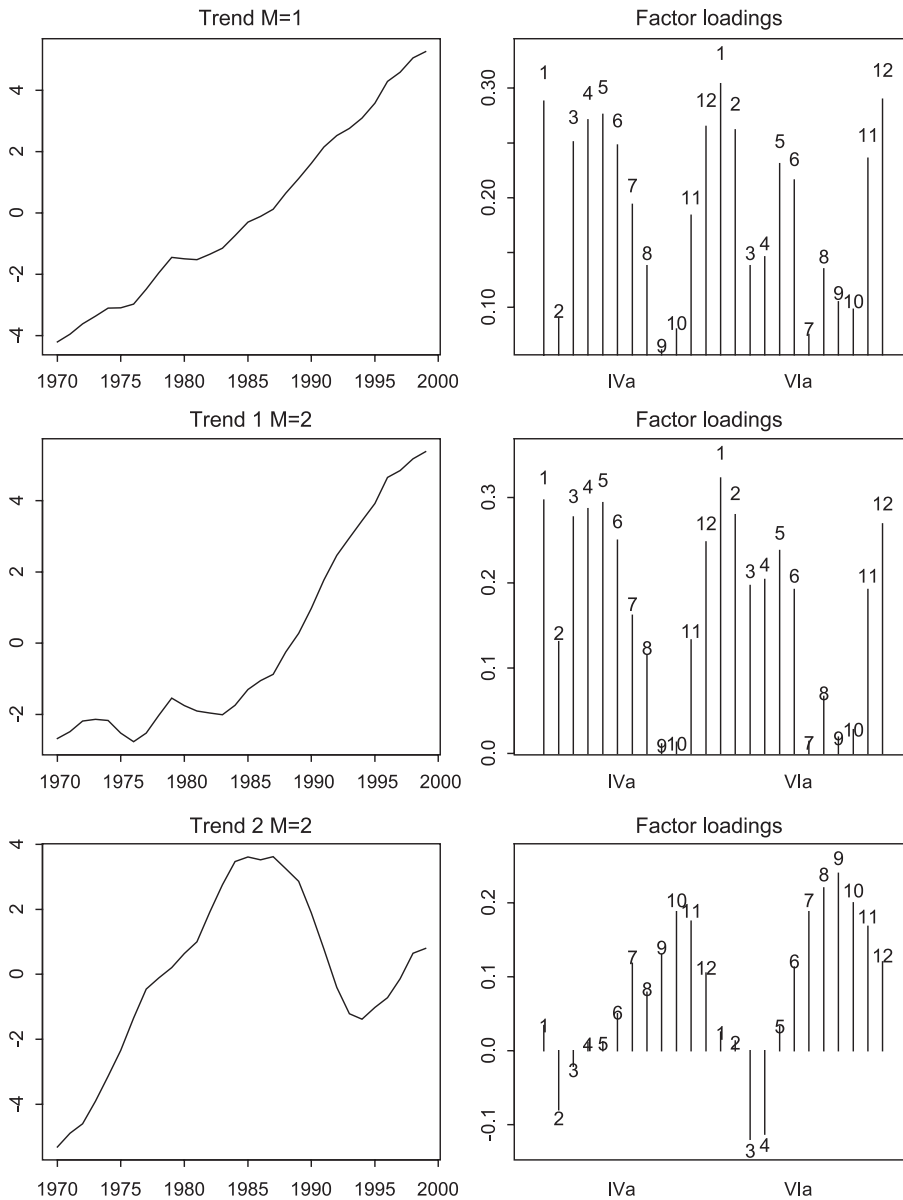


Fig. 7. Results for model 4h and 4i. The first row of figures contains the common trend (left) and factor loadings (right) of model 4h. The second and third rows contain the graphs for model 4i (two common trends). The second row shows the first common trend and corresponding factor loadings and canonical correlation. The third row shows information for the second trend. Months are represented by numbers (1 refers to January, 12 to December). The first 12 numbers refer to area IVa and the second twelve to area VIa. Because the time series are standardised, all the y-axes are unitless and the constant parameters are zero. Factor loadings are represented as vertical lines from 0 to the estimated value. The first vertical line corresponds to the factor loading of the January time series of IVa, the second to February, etc.

common trends of models 4h and 4i suggests that the first common trend in Fig. 4 is the most important one. The three common trends could clearly be linked

to groups of squid time series. The first common trend (an increase since 1984) was important for months (roughly) 1–7 and 11–12. Comparison between

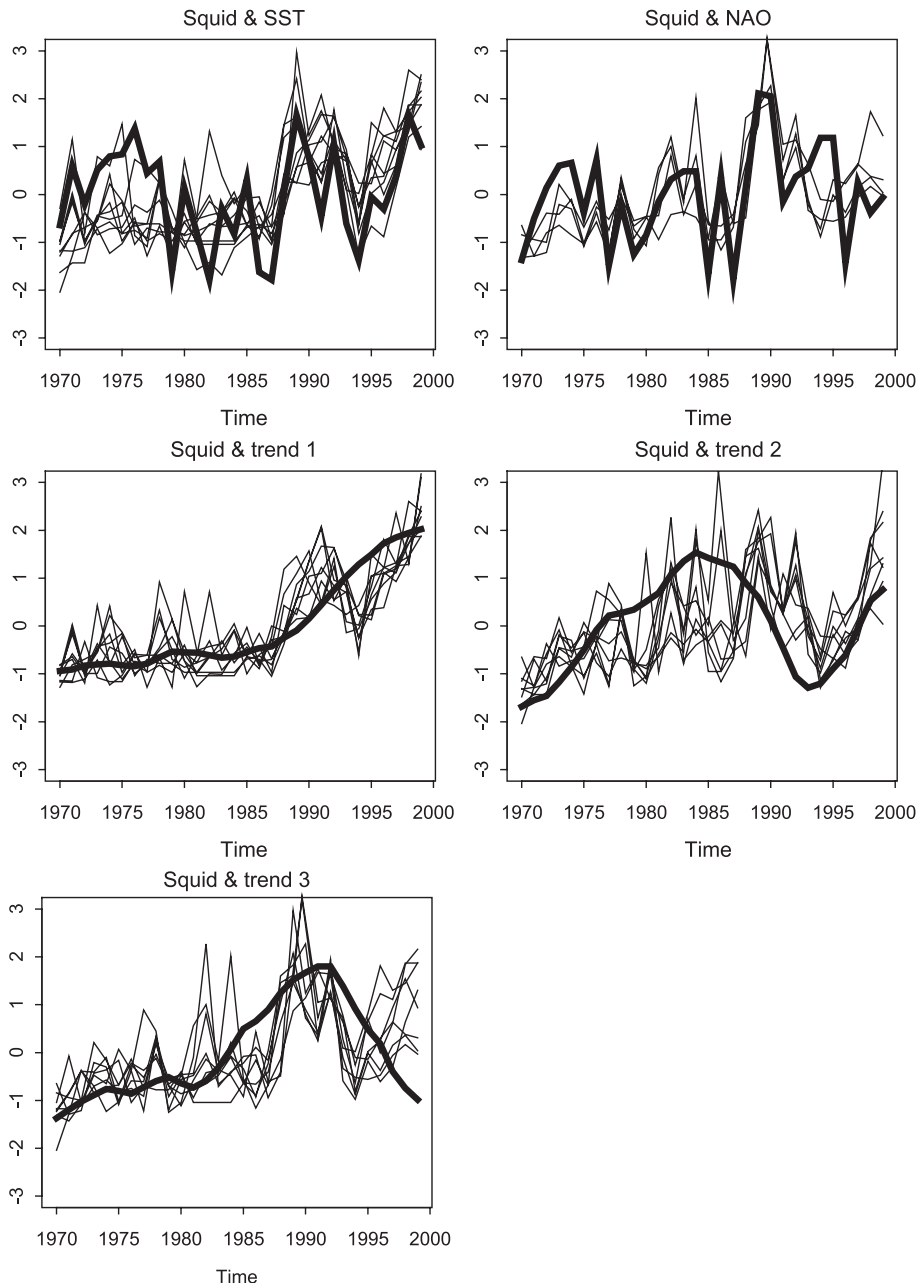


Fig. 8. Selected squid time series and explanatory variables and common trends. Upper left: Millport SST and squid series of months 3–7 from IVa and 3, 9, 11–12 from VIa. Upper right: NAO and squid series of months 8–11 from IVa. Middle left: Trend 1 and squid series of months 1, 3–6, 12 from IVa and 1–2, 12 from VIa. Middle right: Trend 2 and squid series of months 10–11 from IVa and 6–11 from VIa. Lower left: Trend 3 and squid series of months 4, 7–10 from IVa and 10–12 from VIa. All series are standardised and therefore the y-axes are unitless.

trends and factor loadings of model 4i and 4j indicated that 4i used a second trend which was mainly related to months 7–11 in both areas. In model 4j, which allows for more flexibility through an extra trend, these months are split up. Hence, trend 1 seems to represent the winter-early-summer patterns and trend 2 the autumn patterns. The third trend gives additional information on the autumn trends. The SST series probably explains most of the variation from 1987 onwards.

All models using a non-diagonal error matrix **R** had lower AIC values than corresponding models with a diagonal **R**. This indicated that there is a certain amount of information present in the residuals which could not be explained with the SST and NAO index, nor with the common trends. Visualisation of **R** suggests that the unexplained information is related to differences between (i) months 9–12 and the other months and (ii) between the two areas.

The significant contribution of SST as an explanatory variable is common to many analyses carried out on distribution and/or abundance data on *Loligo forbesi*, and indeed for other squid species around the world (e.g. Coelho, 1985; Roberts and Sauer, 1994; Brodziak and Hendrickson, 1999; Robin and Denis, 1999; Waluda et al., 1999) and for various fish species in UK waters (e.g. Bailey and Steele, 1991; Heessen and Daan, 1994; Planque and Frédou, 1999).

The link between temperature and squid catches is most likely to reflect effects on growth. Higher temperatures favour faster growth either directly or through improved food supply (O'Dor et al., 1982; Forsythe, 1993). By growing faster, more squid avoid predation during the early part of their life and the survivors are also larger. Increased catch biomass may reflect both factors. A final point is that increased SST may be reflected in increased sea bottom temperature (SBT), especially in shallow waters such as in the North Sea. Poikilothermic demersal animals such as squid might be expected to be more active at higher SBT, which could increase their availability to fishing gear—although it might also improve their ability to escape capture.

The results suggest that the importance of SST as a predictor increases after 1987. As noted by Pierce and Boyle (2003), the period 1989–1994, which approximately corresponds to the major peak in squid abundance in Scottish coastal waters, was character-

ised by the mildest winter climate seen in the North Sea for at least 50 years (Becker and Pauly, 1996). Thus it may be that when lower temperatures prevail other factors are more important in controlling abundance, with effects of temperature dominating during warm periods.

The effect of the NAO is, in some respects, difficult to disentangle from the effect of SST. Waluda and Pierce (1998) suggested, by analogy with studies on fish, that a positive correlation between winter SST and local abundance of *L. forbesi* in the North Sea was due to the association between higher SST and incursions of Atlantic water, the latter favouring movements of squid into the North Sea. As discussed by Pierce and Boyle (2003), it is likely that a strong coastal current flow along the West Coast of Scotland both increases nutrient supplies and draws more squid into Scottish waters from further south.

Thus a tentative conclusion, to account for the improved fit arising from including both NAO and SST in the model, is that both the inflow of Atlantic water (with associated nutrients, prey organisms and squid) and favourable growth conditions are important in determining abundance as measured by CPUE.

Also of interest is the identification of the three common trends in the data. Broadly speaking, the months July to November (months 7–11) represent the main period of recruitment to the coastal fishery (Collins et al., 1997), so the biological interpretation of the second common trend in model 4i would appear to be a decrease in autumn recruitment between 1985 and 1993. Intuitively, environmental effects on recruitment might be expected, as discussed above. However, because the model contains SST and NAO as explanatory variables, these common trends refer to variation that is not related to either SST or NAO.

The first common trend relates mainly to CPUE in the remaining months of the year, which correspond to the (extended) breeding season and post-breeding period. Thus the first trend describes a general increase in the winter breeding population, apparently unrelated to NAO and SST. Furthermore, this trend continues upwards during the 1985–1993 period of low autumn recruitment strength (second common trend).

The life cycle of *L. forbesi* in Scottish waters may be more complex than is apparent at first sight. The

mis-match in abundance trends for autumn recruits and winter breeders in Scottish waters could indicate that part of the adult spawning population migrates into Scottish waters from elsewhere in winter rather than arriving as immature animals in the autumn. Another feature of the annual cycle is an additional pulse of recruitment in April and May (Boyle et al., 1995; Collins et al., 1997, 1999). By analogy with Holme (1974), who identified summer and winter spawning populations of *L. forbesi* in the English Channel, it is possible that the spring recruits seen in Scottish waters are summer spawners. However, recent biological studies have found little evidence of summer spawning in Scottish waters and the spring recruits may thus spawn elsewhere.

Although the interpretation of the links between the common trends in the CPUE data and the life cycle must be speculative, an interesting possibility is that we are seeing shifts in the relative contribution of ‘spring recruits’ and ‘autumn recruits’ and/or a shift in the relative importance of ‘summer breeders’ and ‘winter breeders’. Recent genetic studies found *L. forbesi* on the continental shelf of northern Europe to be essentially a single population, although animals taken offshore at Rockall and Faroe showed genetic differences (Shaw et al., 1999). It would be interesting to determine whether genetic differences exist between small immature squid from the two main recruitment periods.

Because the time series are standardised, all constant parameters are zero. It is also possible to analyse the data on the original scale. The constant parameters will then represent mean values per month. For these data, it is also possible to add a nominal variable for area, indicating whether mean values per month differ by the two areas. However, interpretation of factor loadings for unstandardised data is more complicated. Because our main interest was to estimate common trends, this analysis was not carried out here.

The current approach to deal with seasonality assumes that there is no shift in seasonal maxima and minima. If the seasonal maxima and minima do shift (in a consistent way), then there might be a problem with the current approach, as well as for most other time-series analysis methods (e.g. de-seasonalising by subtracting monthly averages). Aggregating by quarters might solve this, though the decision which months to aggregate will be arbitrary.

Since the results in this paper indicate that squid abundance is partly temperature driven, one would expect that there is no shift in seasonal maxima and minima.

The DFA methodology identified patterns in the squid time series not revealed by traditional approaches and the new results have raised interesting biological questions. The wider application of this methodology in analysis of biological time series is encouraged.

Acknowledgements

We are grateful to the University Station Millport (Millport) for providing the sea surface temperature data, and to Barbara Buckett (FRS Marine Laboratory) for creating Fig. 1. We would like to thank Ian Jolliffe (University of Aberdeen) and Rob Fryer (FRS Marine Laboratory), and two anonymous referees for their helpful suggestions.

Appendix A

The mathematical formulation of the univariate structural time series model in Eq. (1) is given by

$$y_t = z_t + \mathbf{b}\mathbf{x}_t + e_t$$

$$z_t = z_{t-1} + f_t$$

where y_t contains the value of the response variable at time t , \mathbf{x}_t contains the values of L explanatory variables, \mathbf{b} is a vector of dimension $1 \times L$ containing the regression coefficients, e_t and f_t are noise components and z_t represent the trend. It is assumed that e_t and f_t are independent of each other and that $e_t \sim N(0,r)$ and $f_t \sim N(0,q)$. Hence, the trend at time t is equal to the trend at time $t-1$ plus a contribution of the noise component. If q is relatively small, the contribution of the error component is likely to be small for all t , and the trend will be a smooth curve. If q is relatively large, the trend will show more variation. In smoothing techniques like LOESS or cubic splines (Hastie and Tibshirani, 1990), similar parameters exist (e.g. the span width). The value of q is estimated within the statistical algorithm or can be set to a chosen value.

The mathematical formulation of the dynamic factor model is given by:

$$\mathbf{y}_t = \mathbf{A} \mathbf{z}_t + \mathbf{B} \mathbf{x}_t + \mathbf{e}_t$$

$$\mathbf{z}_t = \mathbf{z}_{t-1} + \mathbf{f}_t$$

The vector \mathbf{y}_t contains the values of the N variables at time t and the $N \times M$ matrix \mathbf{A} contains the factor loadings. The matrix \mathbf{B} contains the regression coefficients and is of dimension $N \times L$. For identification purposes, it is assumed that $\mathbf{f}_t \sim N(\mathbf{0}, \mathbf{I})$. Further details are given in Zuur et al. (2003a).

References

- Bailey, R.S., Steele, J.H., 1991. North Sea herring fluctuation. In: Glantz, M.H. (Ed.), Climatic Variability, Climatic Change and Fisheries. Cambridge University Press, Cambridge, pp. 213–230.
- Bellido, J.M., Pierce, G.J., Wang, J., 2001. Application of generalised additive models to reveal spatial relationships between environmental variables and squid abundance in Scottish waters. *Fish. Res.* 52, 23–40.
- Becker, G.A., Pauly, M., 1996. Sea surface temperature changes in the North Sea and their causes. *ICES J. Mar. Sci.* 53, 887–898.
- Boyle, P.R., Pierce, G.J., 1994. Fishery biology of Northeast Atlantic squid: an overview. *Fish. Res.* 21, 1–15.
- Boyle, P.R., Pierce, G.J., Hastie, L.C., 1995. Flexible reproductive strategies in the squid *Loligo forbesi*. *Mar. Biol.* 121, 501–508.
- Brockwell, P.J., Davis, R.A., 1996. Introduction to Time Series and Forecasting. Springer-Verlag, New York.
- Brodziak, J., Hendrickson, L., 1999. An analysis of environmental effects on survey catches of squids *Loligo pealei* and *Illex illecebrosus* in the northwest Atlantic. *Fish. Bull.* 97, 9–24.
- Chatfield, C., 1989. The Analysis of Time Series. An Introduction Chapman and Hall, London.
- Coelho, M.L., 1985. Review of the influence of oceanographic factors on cephalopod distribution and life cycles. *NAFO Sci. Counc. Stud.* 9, 47–57.
- Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I., 1990. STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *J. Official Statistics* 6, 3–73.
- Collins, M.A., Pierce, G.J., Boyle, P.R., 1997. Population indices of reproduction and recruitment in *Loligo forbesi* (Cephalopoda: Loliginidae) in Scottish and Irish waters. *J. Appl. Ecol.* 34, 778–786.
- Collins, M.A., Boyle, P.R., Pierce, G.J., Key, L.N., Hughes, S.E., Murphy, J., 1999. Resolution of multiple cohorts in the *Loligo forbesi* population from the west of Scotland. *ICES J. Mar. Sci.* 56, 500–509.
- Fogarty, M.J., 1989. Forecasting yield and abundance in exploited invertebrates. In: Caddy, J.F. (Ed.), Marine Invertebrate Fisheries: Their Assessment and Management. John Wiley and Sons, New York, pp. 701–724.
- Forsythe, J.W., 1993. A working hypothesis of how seasonal temperature change may impact the field growth of young cephalopods. In: Okutani, T., O'Dor, R., Kubodera, T. (Eds.), Recent Advances in Cephalopod Fisheries Biology. Tokai University Press, Tokyo, pp. 133–143.
- Georgakarakos, S., Haralabous, J., Valavanis, V., Koutsoubas, D., 2002. Loliginid and ommastrephid stock prediction in Greek waters using time series analysis techniques. *Bull. Mar. Sci.* 71, 269–288.
- Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C., 1998. Multivariate Data Analysis, Fifth Edition. Prentice Hall, New Jersey.
- Harvey, A.C., 1989. Forecasting, Structural Time Series Models and The Kalman Filter. Cambridge University Press, Cambridge.
- Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. Chapman and Hall, London.
- Heessen, H., Daan, N., 1994. Cod distribution and temperature in the North Sea. *ICES Mar. Sci. Symp.* 198, 244–253.
- Holme, N.A., 1974. The biology of *Loligo forbesi* Steenstrup (Mollusca: Cephalopoda) in the Plymouth area. *J. Mar. Biol. Ass. UK* 54, 481–503.
- Huberty, C.J., 1994. Applied Discriminant Analysis. John Wiley and Sons, Inc., New York.
- Hurrell, J.W., 1995. Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science* 269, 676–679.
- Johnston, J., Dinardo, J., 1997. Econometric Methods. McGraw-Hill Companies, Inc., Singapore.
- Jones, R.H., 1993. Longitudinal Data with Serial Correlation: A State-space Approach. Chapman and Hall, London.
- Ljung, L., 1987. System Identification: Theory for the User. Prentice-Hall, New York.
- Mendelssohn, R., Schwing, F.B., 1997. Application of state-space models to ocean climate variability in the northeast Pacific Ocean. In: Aoki, M., Havenner, A.M. (Eds.), Application of Computer-Aided Time Series Modeling, Lecture Notes in Statistics, vol. 119. Springer, New York, pp. 255–280.
- Mendelssohn, R., Schwing, F.B., 2002. Common and uncommon trends in SST and Wind stress in the California and Peru-Chile Current Systems. *Prog. Oceanogr.* 53, 141–162.
- O'Dor, R.K., Balch, N., Foy, E.A., Hirtle, R.W.M., Johnston, D.A., 1982. Embryonic development of the squid, *Illex illecebrosus*, and effect of temperature on development rates. *J. Northw. Atl. Fish. Sci.* 3, 41–45.
- Pajuelo, J.G., Lorenzo, J.M., 1995. Analysis and forecasting of the demersal fishery of the Canary Islands using an ARIMA model. *Sci. Mar.* 59, 155–164.
- Park, H.-H., 1998. Analysis and prediction of walleye pollock (*Theragra chalogramma*) landings in Korea by time series analysis. *Fish. Res.* 38, 1–7.
- Pierce, G.J., Boyle, P.R., 2003. Empirical modelling of interannual trends in abundance of squid (*Loligo forbesi*) in Scottish waters. *Fish. Res.* 59, 305–326.
- Pierce, G.J., Bailey, N., Stratoudakis, Y., Newton, A., 1998. Distribution and abundance of the fished population of *Loligo forbesi*

- in Scottish waters: analysis of research cruise data. ICES J. Mar. Sci. 55, 14–33.
- Pierce, G.J., Wang, J., Zheng, X., Bellido, J.M., Boyle, P.R., Denis, V., Robin, J.-P., 2001. The cephalopod fishery GIS for the Northeast Atlantic: development and application. Int. J. Geogr. Inform. Sci. 15, 763–784.
- Planque, B., Frédou, T., 1999. Temperature and recruitment of Atlantic cod (*Gadus morhua*). Can. J. Fish. Aquat. Sci. 56, 2069–2077.
- Quinn, G.P., Keough, M.J., 2002. Experimental Design and Data Analysis for Biologists. Cambridge University Press, Cambridge.
- Roberts, M.J., Sauer, W.H.H., 1994. Environment: the key to understanding the South African chokka squid (*Loligo vulgaris reynaudii*) life-cycle and fishery? Antarct. Sci. 6, 249–258.
- Robin, J.P., Denis, V., 1999. Squid stock fluctuations and water temperature: temporal analysis of English Channel Loliginidae. J. Appl. Ecol. 36, 101–110.
- Shaw, P.W., Pierce, G.J., Boyle, P.R., 1999. Subtle population structuring within a highly vagile marine invertebrate, the veined squid *Loligo forbesi*, demonstrated with microsatellite DNA markers. Mol. Ecol. 8, 407–417.
- Shumway, R.H., 2000. Dynamic Mixed Models for Irregularly Observed Time Series. Resenhas IME-USP 4, 433–456.
- Sims, D.W., Genner, M.J., Southward, A.J., Hawkins, S.J., 2001. Timing of squid migration reflects North Atlantic climate variability. Proc. R. Soc. Lond. B 268, 2607–2611.
- Stergiou, K.I., 1989. Modelling and forecasting the fishery for pilchard (*Sardina pilchardus*) in Greek waters using ARIMA time-series models. J. Cons. Int. Explor. Mer 46, 16–23.
- Stergiou, K.I., 1991. Short-term fisheries forecasting—comparison of smoothing, ARIMA and regression techniques. J. Appl. Ichthyol. 7, 193–204.
- Stergiou, K.I., Christou, E.D., 1996. Modelling and forecasting annual fisheries catches: Comparison of regression, univariate and multivariate time series methods. Fish. Res. 25, 105–138.
- Stergiou, K.I., Christou, E.D., Petrakis, G., 1997. Modelling and forecasting monthly fisheries catches: comparison of regression, univariate and multivariate time series methods. Fish. Res. 29, 55–95.
- Ter Braak, C.J.F., 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 67, 1167–1179.
- Waluda, C.M., Pierce, G.J., 1998. Temporal and spatial patterns in the distribution of squid (*Loligo* spp.) in UK waters. S. Afr. J. Mar. Sci. 20, 323–336.
- Waluda, C.M., Trathan, P.N., Rodhouse, P.G., 1999. Influence of oceanographic variability on recruitment in the *Illex argentinus* (Cephalopoda: Ommastrephidae) fishery in the South Atlantic. Mar. Ecol. Prog. Ser. 183, 159–167.
- Zuur, A.F., Fryer, R.J., Jolliffe, I.T., Dekker, R., Beukema, J.J., 2003a. Estimating common trends in multivariate time series using dynamic factor analysis. Environmetrics 14, 665–685.
- Zuur, A.F., Tuck, I.D., Bailey, N., 2003b. Dynamic factor analysis to estimate common trends in fisheries time series. Can. J. Fish. Aquat. Sci. 60, 542–552.