International Council for the Exploration of the Sea

Incorporation of External Factors in Marine Resource Surveys
CM 2000/K:18

Incorporating temporal information in ichthyoplankton surveys using a model-based approach: cod (*Gadus morhua* L.) in the Irish Sea

C.M. O'Brien, and C.J. Fox

CEFAS Lowestoft Laboratory
Pakefield Road
Lowestoft
Suffolk NR33 0HT
United Kingdom

fax: +44 (0) 1502 513865
e-mail: first initial.second initial.surname@cefas.co.uk

ABSTRACT

Ichthyoplankton surveys are increasingly being employed for the purpose of assessing the spawning stock biomass of species. To investigate trends in egg production generalized additive models (GAMs) can be used to model the spatial-temporal distribution of egg density. Time is treated as a continuous variable and egg production is modelled as a function of location, time, and oceanographic and environmental variables. Survey series often include large numbers of stations where eggs of a particular species are not observed. The application of two-stage models is recommended in these cases, where the presence/absence of eggs is firstly modelled as a binary process and a model-based surface subsequently fitted to egg production (conditional on their presence). By integrating under the predicted egg production surfaces, a cumulative production curve can be generated and a GAM-based estimate of annual egg production produced. The successful application of GAMs, however, requires a survey design with good coverage in both space and time. Ichthyoplankton samples were collected in 1995, over the period February to June, on a series of eleven cruises covering most of the Irish Sea. Cod (*Gadus morhua* L.) is used to illustrate the application of the model-based estimation of annual egg production and the calculation of its coefficient of variation. The methods will be applied to analyses of the year 2000 surveys recently conducted in the Irish Sea as part of a collaborative EU-funded project.

Keywords: generalized additive modelling, simulation, spawning

INTRODUCTION

Ichthyoplankton surveys are increasingly being employed for the purpose of assessing the spawning stock biomass (SSB) of species whose egg production can be determined in this way.

During 1995 large-scale fish egg sampling exercises were carried out in the Irish Sea (Anon., 1997). The aim of these surveys was to estimate the SSB of cod, plaice and sole using the Annual Egg Production Method (AEPM). This method is not based upon data from the commercial fisheries and therefore provides an independent estimate of SSB. The usual assessment method for these stocks is known as Virtual Population Analysis (VPA) and is based upon data from the commercial fisheries. As concerns have been raised about the quality of data from the commercial fisheries, the development of fishery independent methods of stock assessment can provide a useful addition to assessments generated via VPA. Surveys of Irish Sea cod and plaice were conducted again this year (2000) as a comparison to the surveys undertaken during 1995.

The rate of egg development is largely dependent upon the temperature of the sea water in which the eggs are incubated. During development, eggs pass through a number of recognizable stages. In order to estimate the numbers of eggs produced during the plankton surveys, an estimate of egg age is required. This is obtained by assigning the eggs collected in the plankton survey to a visual development stage. By comparing the stage and known water temperature at the time of collection with laboratory derived calibration curves, the age of the egg can be estimated. The 1995 AEPM employed six visual stages to determine egg age (IA, IB, II, III, IV and V).

This is a methodological paper and the papers by Anon.(1997), Armstrong *et al*.(2000) and Fox *et al*.(2000) should be consulted for further details of the sampling design, survey implementation and fish biology. In this paper, we model the stage IA egg production during the spawning season based on the 1995 surveys using generalized additive models (GAMs) as described in Fox *et al*.(2000). The approach is here extended to incorporate the estimation of the coefficient of variation of the estimate of annual egg production. Cod (*Gadus morhua* L.) is used for illustrative purposes.

MATERIAL AND METHODS

Ichthyoplankton samples were collected in 1995 during the months of February through to June on a series of eleven cruises covering most of the Irish Sea. Plankton was collected using Gulf VII/PRO-NET and MAFF/Guildline high-speed plankton samplers deployed in oblique hauls to within 2 m of the sea bed (Nash *et al*., 1998). Three different survey grids were used as the spawning season progressed based upon the expected distribution of eggs of cod, plaice and sole (Nichols *et al*., 1993).

Surveys of this type often include large numbers of stations where eggs of a particular species are not observed. In these cases, the application of two-stage models has been recommended (Pennington, 1983; Borchers *et al*., 1997).

Stage IA cod egg abundance data (numbers per metre square) were converted to egg production (numbers per metre square per day) using a temperature-dependent stage duration model (Ryland and Nichols, 1975) with the parameter estimates reported in Fox *et al*.(2000).

Time was treated as a continuous variable and egg production was modelled as a function of location, depth and time. Other oceanographic and environmental variables are not well defined except at sampled points and times so were excluded from further consideration in the subsequent modelling. GAMs (Hastie and Tibshirani, 1990) are an extension of generalized linear models, GLMs (McCullagh and Nelder, 1989), but allow greater flexibility for modelling spatial and temporal trends than do GLMs.

Five steps were involved in model specification for both egg presence/absence data and egg production data.

<u>step 1</u>: Covariates were selected from those available, namely, survey grid; latitude and longitude (lat and long); date of sampling (date); time of sampling; mean bottom depth during haul (depth); depth-integrated sea-temperature; and salinity. The choice of covariates to include in a GLM was based on stepwise backward elimination.
<u>step 2</u>: A GAM was fitted using smoothing splines with 4 degrees of freedom (df).
<u>step 3</u>: Testing whether the GAM models the data significantly better than the GLM, using tests based upon change in deviance.
<u>step 4</u>: Interaction terms to be included in the model were selected by stepwise backward elimination of all pair-wise combinations of the covariates selected at the initial step 1.
<u>step 5</u>: Examination of residual plots for evidence of lack of model fit followed by increasing the df of the smoothing splines, if necessary. Improvements in model fit were evaluated on the basis of approximate F-tests (Hastie and Tibshirani, 1990).

In order to derive estimates of egg production over the study area at specific dates, prediction surfaces were generated over a regular grid based on ICES statistical rectangles (Figure 1), using a two-stage model as follows. The probability, *p*, of obtaining a zero observation was estimated first using a logit link with a binomial distribution (Cox, 1970; Nelder and Wedderburn, 1972); namely,

$$\log_e \{ p / (1 - p) \} = a + s_1(\text{latitude}, 4) + s_2(\text{longitude}, 4) + s_3(\text{date}, 4) + s_4(\text{depth}, 4) + s_5(\text{latitude.longitude}, 4)$$

followed by estimation of the mean number of eggs, *n*, given the presence of eggs, using a log link with a Gamma distribution; namely,

$$\log_e (E \{ n \mid \text{presence} \}) = b + s_1(\text{latitude}, 6) + s_2(\text{longitude}, 6) + s_3(\text{date}, 6) + s_4(\text{depth}, 6) + s_5(\text{latitude.longitude}, 6)$$

where E denotes expectation, a and b are intercept terms, s_i(variable-name, df) is a one-dimensional cubic spline smoother for the *i*th covariate variable-name based on degrees of freedom df. Total egg production (number of stage IA eggs per metre square per day) was estimated by integrating over the product of the estimated presence probability surface and the estimated egg density surface, given presence.

The variance of the stage IA egg production estimator can be estimated using a parametric bootstrap (c.f. Efron and Tibshirani, 1993; Augustin *et al*., 1998). This involves generating a large number (500, 1000, 5000, …) of pseudo-samples of the egg survey data using the fitted model, and re-fitting the GAM to each of these pseudo-samples. Integrating each re-fit over space and time yields a single bootstrap estimate corresponding to each of the pseudo-samples. The coefficient of variation (CV) of these bootstrap estimates is the estimate of the CV for the GAM estimate. The bootstrap procedure can generate variance estimates for egg production estimates at any spatial and/or temporal resolution. The limiting factor is computational time since the parametric bootstrap has two distinct components for each pseudo-sample – one generating presence/absence data, the other generating numbers given presence.

All modelling, parameter estimation and computer-based simulation were performed within the programming environment of S-PLUS (Statistical Sciences Inc., 1999). Computer scripts were written in order to semi-automate and optimize the numerical estimation and bootstrap procedures.

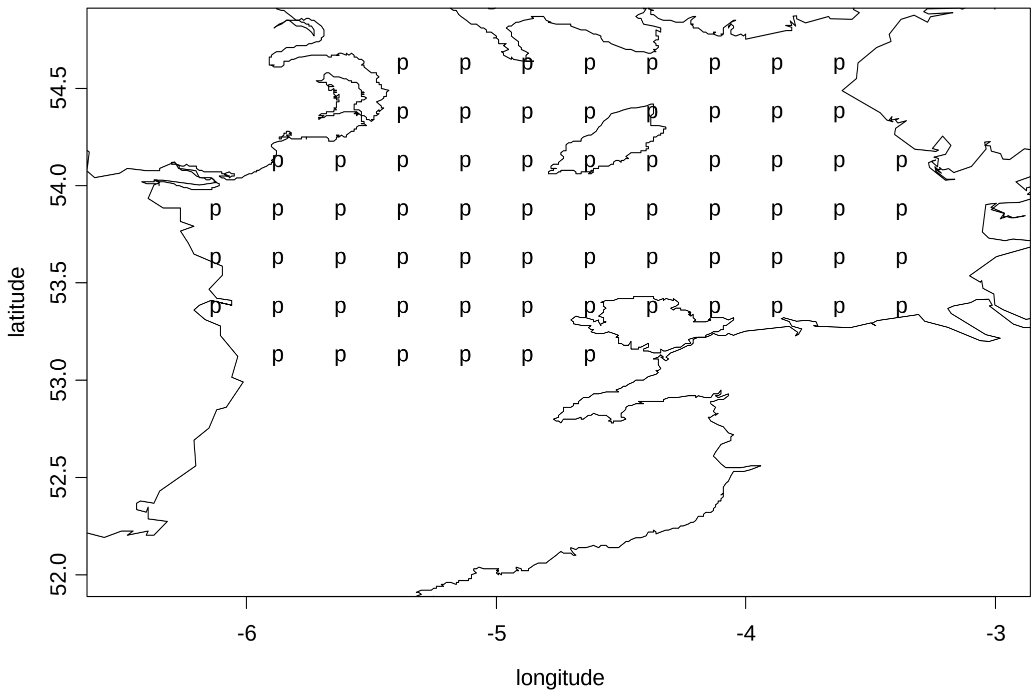

Figure 1. Chart showing the overall survey region within the Irish Sea. Each letter p identifies the mid-point of an ICES statistical rectangle that is used in the estimation of egg production for cod.

RESULTS

Survey coverage is described in Fox *et al*.(1997) but the general area of study is shown in Figure 1. Data collected south of latitude 53°N were excluded from the present analyses with GAMs because of incomplete survey coverage of this region.

Survey data

Figures 2, 3 and 4 show stage IA egg production for cod as pseudo-synoptic plots. Cod eggs were recorded at 481 out of the 972 stations sampled during the 1995 survey. Some spawning of cod was recorded during the first surveys in February and a general increase and decline in production during the sampling period is evident.

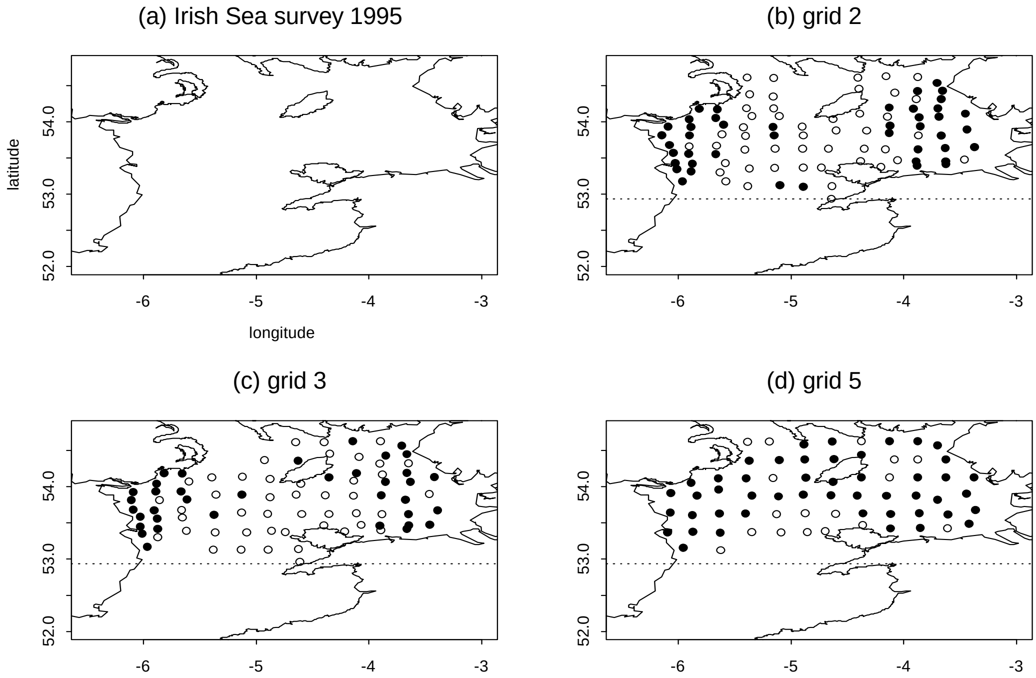

Figure 2. Charts showing the location of stations sampled on grids 2 (12-19 February), 3 (21-26 February) and 5 (8-14 March) during the 1995 spawning season; together with the presence/absence of stage IA cod eggs. Filled circles indicate the presence of stage IA cod eggs; unfilled circles indicate absence.

Modelled egg presence/absence

For cod it was found necessary to extend the GLM to a GAM (Table 1) to allow covariates in the model to take on non-parametric forms, but the degree of smoothing of the covariates required was low (4 degrees of freedom).

Figure 5 illustrates how the probability of egg presence developed over the survey period. Figure 5(b) corresponds to the date of the 9th March 1995 but as a similar plot is obtained for the 9th April 1995 only one of the dates is shown. Cod spawning occurred over most of the Irish Sea but with major concentrations off the Irish coast, between the Isle of Man and Cumbria and to a limited extent, off the south-west tip of the Isle of Man. The spatial structure of the probability of occurrence of cod eggs over the season indicates that all three sites build up and decline in phase and at similar rates.

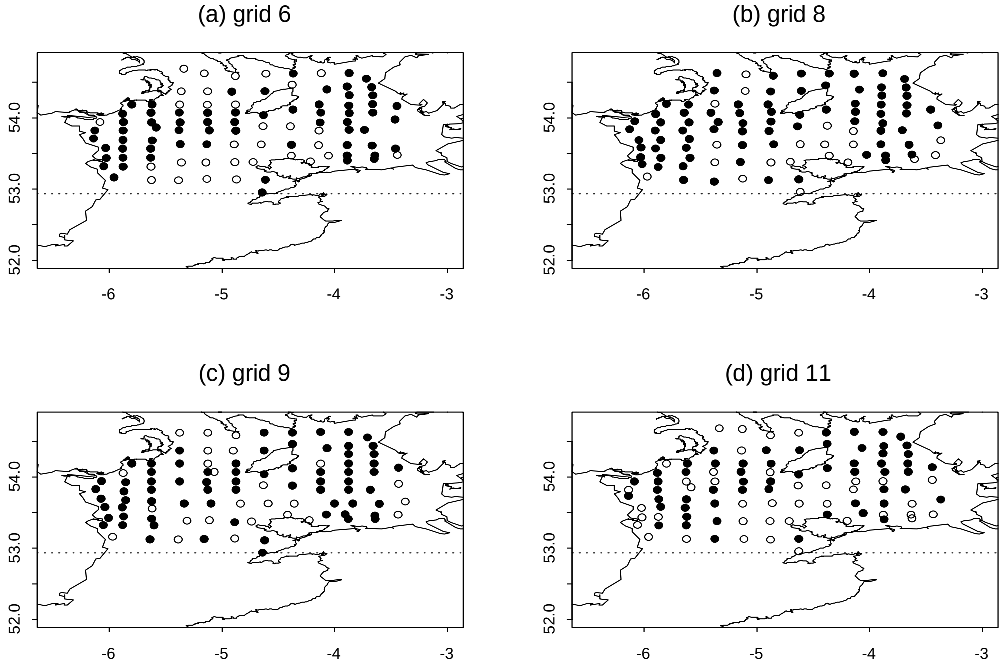

Figure 3. Charts showing the location of stations sampled on grids 6 (15-22 March), 8 (30 March – 6 April), 9 (10-21 April) and 11 (19-26 April) during the 1995 spawning season; together with the presence/absence of stage IA cod eggs. Filled circles indicate the presence of stage IA cod eggs; unfilled circles indicate absence.

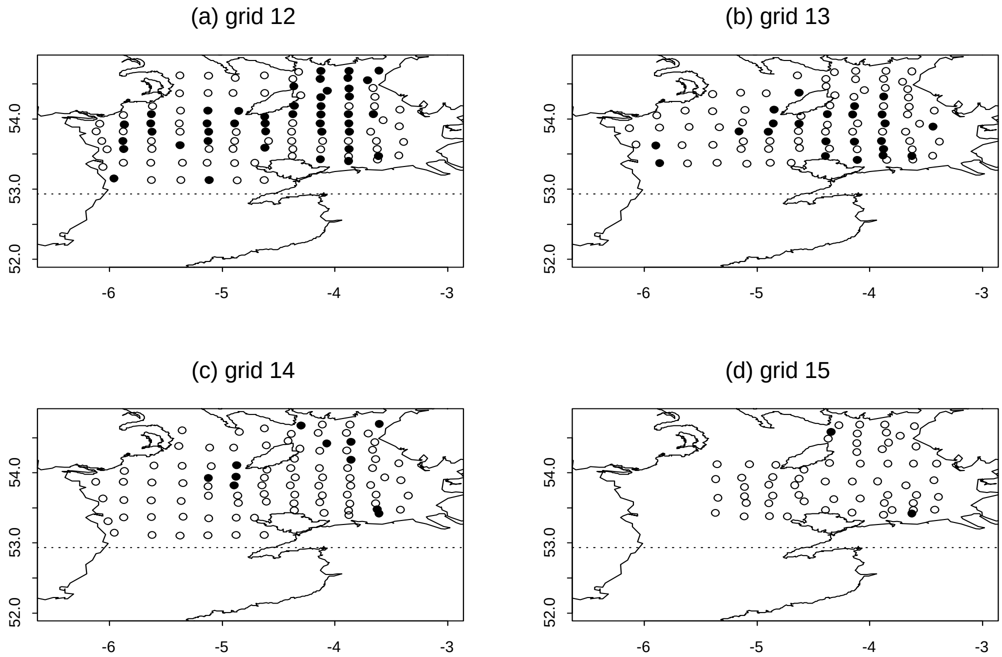

Figure 4. Charts showing the location of stations sampled on grids 12 (1-7 May), 13 (14-20 May), 14 (23-28 May) and 15 (5-14 June) during the 1995 spawning season; together with the presence/absence of stage IA cod eggs. Filled circles indicate the presence of stage IA cod eggs; unfilled circles indicate absence.

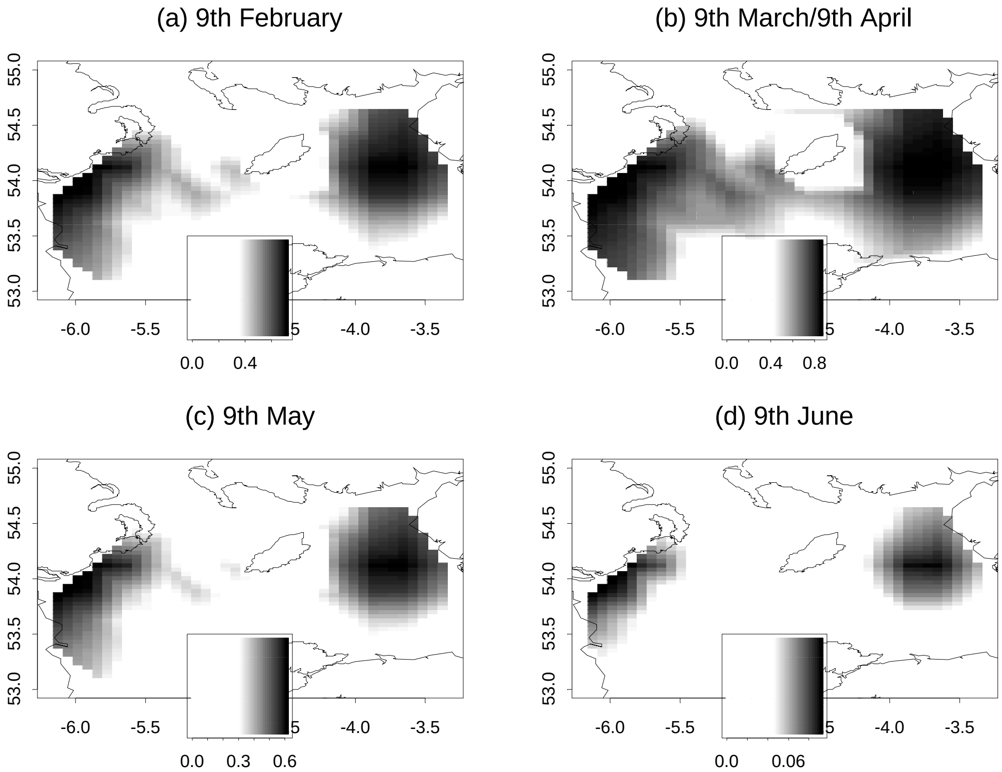

Figure 5. Changes in the probability of occurrence of stage IA cod eggs on five approximately equally spaced days through the 1995 spawning season. The small overlaid rectangle at the bottom of each plot indicates the grayscale used to display the probabilities on each date.

Modelled egg production

Initial model fits using splines with 4 degrees of freedom resulted in fitted egg densities that were lower than those observed in the ichthyoplankton samples. This tendency was rectified by increasing the flexibility of the non-parametric components in the model. 6 degrees of freedom for all covariates was found to be appropriate based purely on a desire not to over-fit the data. The GAM produced a good overall fit in terms of deviance (Table 2).

The seasonal and cumulative egg production curves obtained from the GAM for cod are shown in Figures 6(a) and 6(b), respectively. Cod spawning peaked during the last week of March and spawning by cod was effectively completed by May. The main spawning areas were predicted to be off the Irish coast, between the Isle of Man and Cumbria, and to the south-west of the Isle of Man.

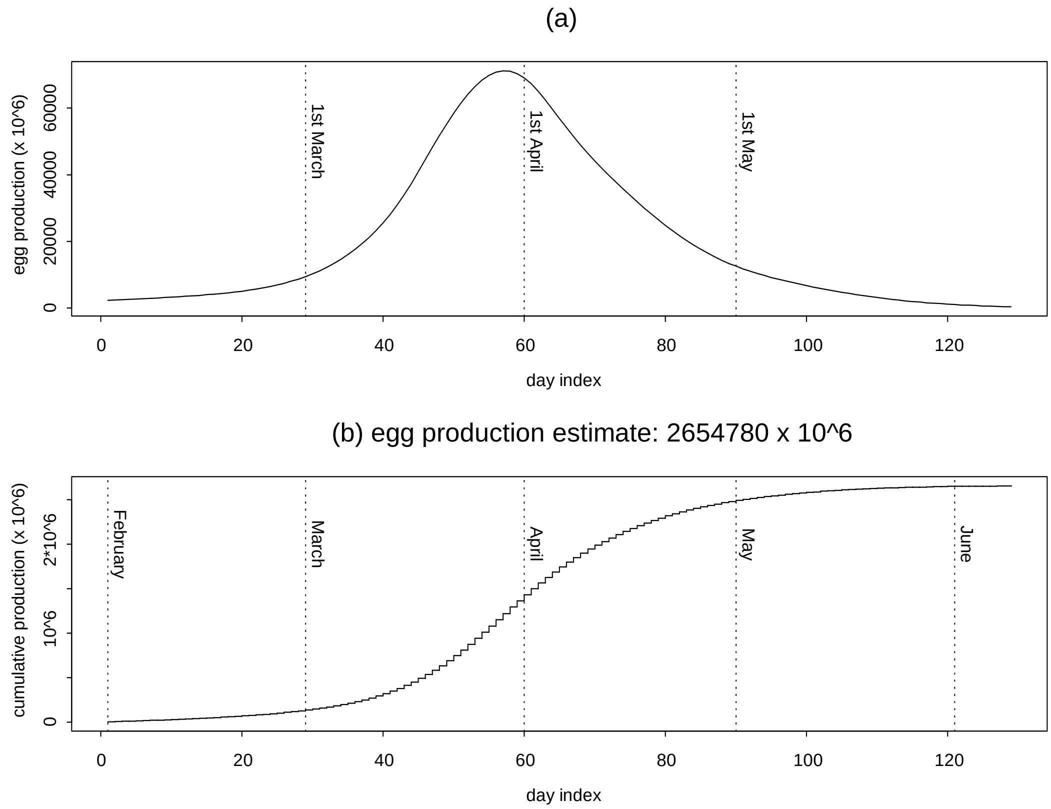

Figure 6. Modelled: (a) seasonal stage IA egg production, and (b) cumulative stage IA egg production for cod in the Irish Sea during the 1995 spawning season.

Bootstrapped egg production

The estimated annual production of stage IA cod eggs in the Irish Sea is reported by Armstrong *et al*.(2000) as 2.844 x 10^{12} (with a coefficient of variation of 0.144) based upon a traditional estimator (Pope and Woolner, 1984). This compares favourably with the GAM-based estimate of 2.655 x 10^{12} but not with its percentage coefficient of variation (%CV) of ~ 8%. The CV of the GAM-based estimate is close to a half of the CV of the traditional method. This increase in precision is partly to be expected given the substantial spatial and temporal trend in observed cod egg densities within the survey region of the Irish Sea. In contrast to the traditional method, the GAM-based method is able to model this trend relatively parsimoniously using spline smoothers.

DISCUSSION

Ichthyoplankton samples were collected in 1995, over the period February to June, on a series of eleven cruises covering most of the Irish Sea. GAMs have been used to model the spatial-temporal distribution of egg density. Time is treated as a continuous variable and egg production is modelled as a function of location, time, and oceanographic variables. The survey series includes a large number of stations where eggs of a particular species were not observed. The application of two-stage models has been recommended in these cases, where the presence/absence of eggs is firstly modelled as a binary process and a model-based surface subsequently fitted to egg production (conditional on their presence). By integrating under the predicted egg production surfaces, a cumulative production curve has been generated and a GAM-based estimate of annual egg production produced. Cod (*Gadus morhua* L.) has been used to illustrate the application of the model-based estimation of annual egg production for stage IA eggs and the calculation of its coefficient of variation.

The GAM-based estimate of cod stage IA egg production yields a much more precise estimate than that obtained by the traditional method. In addition to improving precision, the GAM method provides information on the relationship between the response and covariates. The most important relationship is between egg abundance and date. The GAM is able to model complex trends in egg abundance with respect to space, time and other covariates. The method provides information on the nature of these trends at a resolution which provides useful insights into the underlying mechanisms driving spawning distribution (Fox *et al*., 2000). In the traditional method only substantial changes in the distribution with time are visible.

Further developments of the AEPM methods will be applied to analyses of the year 2000 surveys of Irish Sea cod and plaice recently undertaken as part of a collaborative project (EU contract DG XIV 98/090) and related MAFF-funded studies. Other egg stages may be modelled with the GAM and spatially disaggregated estimates of production (or mortality) may be obtained, to mention but two further extensions.

ACKNOWLEDGEMENTS

The data used in this paper were collected under EU contract AIR3-CT94-2263 (Estimation of fish biomass in the Irish Sea by means of the annual egg production method) within DG XIV. The following are thanked for the collection and provision of data: M.J. Armstrong (DARD, Belfast), P. Connolly (Marine Institute, Dublin), R.D.M. Nash (Port Erin Marine Laboratory, Isle of Man) and M.G. Pawson (CEFAS, Lowestoft). The ship crews and laboratory staff of all participating institutes are thanked for their work on this project. The authors would like to thank Mark Dickey-Collas for his comments on the initial draft of this paper. The preparation of this paper was funded by the Ministry of Agriculture, Fisheries and Food under contracts MF0316 (Evaluation of fishery management procedures and development of stock assessment methods) and MF0423 (Development of models of plaice, *Pleuronectes platessa* L., population dynamics incorporating biological processes for use in risk assessment of management options).

REFERENCES

Anon. (1997). Estimation of fish biomass in the Irish Sea by means of the annual egg production method. Queen’s University, *Contract Report AIR3-CT94-2263*, Belfast.

Armstrong, M.J., Connolly, P., Nash, R.D.M., Pawson, M.G., Alesworth, E., Coulahan, P.J., Dickey-Collas, M., Milligan, S.P., O’Neill, M.F., Witthames, P.R. and Woolner, L. (2000). An application of the annual egg production method to estimate the spawning biomass of cod (*Gadus morhua* L.), plaice (*Pleuronectes platessa* L.) and sole (*Solea solea* L.) in the Irish Sea. *ICES Journal of Marine Science* (accepted).

Augustin, N.H., Borchers, D.L., Clarke, E.D., Buckland, S.T. and Walsh, M. (1998). Spatio-temporal modelling for the annual egg production method of stock assessment using generalized additive models. *Canadian Journal of Fisheries and Aquatic Sciences*, **55**, 2608-2621.

Borchers, D.L., Buckland, S.T., Priede, I.G. and Ahmadi, S. (1997). Improving the precision of the daily egg production method using generalized additive models. *Canadian Journal of Fisheries and Aquatic Sciences*, **54**, 2727-2742.

Cox, D.R. (1970). *The Analysis of Binary Data*. London: Methuen.

Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. London: Chapman and Hall.

Fox, C.J., Dickey-Collas, M. and Winpenny, A.J. (1997). Spring plankton surveys of the Irish Sea in 1995: the distribution of fish eggs and larvae. Centre for Environment, Fisheries and Aquaculture Science (CEFAS), *Science Series Technical Report 104*, Lowestoft, Suffolk, UK.

Fox, C.J., O'Brien, C.M., Dickey-Collas, M. and Nash, R.D.M. (2000). Patterns in the spawning of cod (*Gadus morhua* L.), sole (*Solea solea* L.) and plaice (*Pleuronectes platessa* L.) in the Irish Sea as determined by generalized additive modelling. *Fisheries Oceanography*, **9**, 33-49.

Hastie, T.J. and Tibshirani, R.J. (1990). *Generalized Additive Models*. London: Chapman and Hall.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*. London: Chapman and Hall.

Nash, R.D.M., Dickey-Collas, M. and Milligan, S.P. (1998). Descriptions of the Gulf VII/PRO-NET and MAFF/Guildline unencassed high-speed plankton samplers. *Journal of Plankton Research*, **20**, 1915-1926.

Nelder, J.A. and Wedderburn, R.W.N. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, **135A**, 370-384.

Nichols, J.H., Haynes, G.M., Fox, C.J., Milligan, S.P., Brander, K.M. and Chapman, R.J. (1993). Spring plankton surveys of the Irish Sea in 1982, 1985, 1987, 1988 and 1989: hydrography and the distribution of fish eggs and larvae. *Fisheries Research Technical Report*, MAFF Directorate of Fisheries Research, Lowestoft, (95): 111pp.

Pennington, M. (1983). Efficient estimators of abundance for fish and plankton surveys. *Biometrics*, **39**, 281-286.

Pope, J.G. and Woolner, L. (1984). An investigation of the precision of the 1983 western mackerel egg survey. ICES CM 1984/H:70.

Ryland, J.S. and Nichols, J.H. (1975). Effect of temperature on the embryonic development of the plaice, *Pleuronectes platessa* L. (Teleostei). *Journal of Experimental Marine Biology and Ecology*, **18**, 121-137.

Statistical Sciences Inc. (1999). *S-PLUS 2000 for Windows*. Seattle: MathSoft Inc.

Table 1. Comparison of logistic GLM and logistic GAM fitted to stage IA egg presence/absence data.

Model		residual df	residual deviance	test	df	deviance	pr(Chi)
M1:	1 + lat + long + date + depth + lat*long	966.000	1202.622				
M2:	1 + s(lat,4) + s(long,4) + s(date,4) + s(depth,4) + s(lat*long,4)	951.406	961.202	M1 vs. M2	14.594	241.42	< 0.001

Dispersion parameter for binomial family taken to be 1. Mean deviance (M2) estimate 1.01 (p-value not significant).

Table 2. Comparison of GLM and GAM fitted to stage IA egg production data conditional upon presence.

Model		residual df	residual deviance	test	df	deviance	F-value	pr(F-value)
M3:	1 + lat + long + date + depth + lat*long	475.000	1027.601					
M4:	1 + s(lat,6) + s(long,6) + s(date,6) + s(depth,6) + s(lat*long,6)	450.001	620.602	M3 vs. M4	24.999	406.999	9.959	< 0.001