ICES Journal of Marine Science (2011), 68(3), 528–536. doi:10.1093/icesjms/fsr002

A revision of daily egg production estimation methods, with application to Atlanto-Iberian sardine. 2. Spatially and environmentally explicit estimates of egg production

Miguel Bernal[1,*,‡], Yorgos Stratoudakis[2], Simon Wood[3], Leire Ibaibarriaga[4], Luis Valdés[5], and David Borchers[6]

[1]*Instituto Español de Oceanografía (IEO), Centro Oceanográfico de Cádiz, Puerto Pesquero, Muelle de Levante s/n, Apartado 2609, 11006 Cádiz, Spain*
[2]*Instituto Nacional de Recursos Biologicos (INRB/IPIMAR), Avenida de Brasília 1449-006, Lisboa, Portugal*
[3]*Mathematical Sciences, University of Bath, Bath BA2 7AY, UK*
[4]*AZTI-Tecnalia, Marine Research Unit, Txatxarramendi ugartea z/g, 48395 Sukarrieta, Basque Country, Spain*
[5]*Intergovernmental Oceanographic Commission of UNESCO, 1 rue Miollis, 75732 Paris cedex 15, France*
[6]*CREEM, University of St Andrews, The Observatory, Buchanan Gardens, St Andrews, Fife KY16 9LZ, Scotland, UK*

**Corresponding Author: tel: +34 932 309500; fax: +34 932 309555; e-mail: miguel.bernal@cd.ieo.es.*
‡Current address: Institute of Marine and Coastal Sciences, University of Rutgers, 71 Dudley Road, New Brunswick, New Jersey 08901, USA.

Bernal, M., Stratoudakis, Y., Wood, S., Ibaibarriaga, L., Uriarte, A., Valdés, L., and Borchers, D. 2011. A revision of daily egg production estimation methods, with application to Atlanto-Iberian sardine. 2. Spatially and environmentally explicit estimates of egg production. – ICES Journal of Marine Science, 68: 528–536.

Received 1 June 2010; accepted 2 January 2011.

A spatially and environmentally explicit egg production model is developed to accommodate a number of assumptions about the relationship between egg production and mortality and associated environmental variables. The general model was tested under different assumptions for Atlanto-Iberian sardine. It provides a flexible estimator of egg production, in which a range of assumptions and hypotheses can be tested in a structured manner within a well-defined statistical framework. Application of the model to Atlanto-Iberian sardine increased the precision of the egg production time-series, and allowed improvements to be made in understanding the spatio-temporal variability in egg production, as well as implications for ecology and stock assessment.

Keywords: egg production, GAM, mortality, *Sardina pilchardus*, sardine, spatially explicit models.

Introduction

The daily egg production method (DEPM; Lasker, 1985) estimates spawning biomass by comparing the observed daily egg production over a spawning area with the population's daily fecundity rate. Together with acoustics (Simmonds and MacLennan, 2005), the DEPM has been the preferred fishery-independent method of assessing spawning-stock biomass (SSB) of sardine (*Sardina pilchardus*) in Atlantic waters of the Iberian Peninsula. The DEPM has been applied to the fishery since the late 1980s (Cunha *et al.*, 1992; García *et al.*, 1992), but recent reviews (e.g. ICES, 2004; Stratoudakis *et al.*, 2006) have raised concerns regarding both daily egg production and population fecundity rate estimators. For egg production, the principal issues include bias and lack of precision of the mortality estimates, and the effect of the spatial structure of the population on the estimators. A revision of mortality estimates for Atlanto-Iberian sardine is presented in Bernal *et al.* (2011), and an extension of traditional daily egg production estimators to include spatial analysis, and its application to improve estimates of Atlanto-Iberian sardine egg production, are presented and discussed in this paper.

The basic model to estimate egg production was originally obtained from early differential equation models of population growth (e.g. Verhulst, 1839; Lotka, 1925), by assuming a constant rate of mortality from an initial number of eggs released (Lasker, 1985; Gunderson, 1993):

$$\frac{\partial N}{\partial t} = -m \rightarrow N_a = N_0\, e^{(-ma)}, \tag{1}$$

where N_a is the number of eggs of a given daily cohort, i.e. the eggs released on the same day, with mean age a, N_0 the daily rate of egg production, and m the daily mortality rate. Or, expressed as density (e.g. using effective surface area, Efarea, as the sampling unit; see ICES, 2004; Stratoudakis *et al.*, 2006; Bernal *et al.*, 2011):

$$D_a = D_0\, e^{(-ma)}, \tag{2}$$

where D_a is the density of eggs of a given daily cohort with mean age a and D_0 the daily egg production rate by unit area. Estimates of m and D_0 are traditionally obtained by fitting Equation (2), assuming that ages are known and eggs can be classified into daily cohorts without error, using a statistical distribution based on the residual structure (Picquelle and Stauffer, 1985).

Spatial variability has been incorporated into egg production estimation in different ways. The traditional way of dealing with the spatial structure of the different parameters involved in the DEPM is to post-stratify the data (Lasker, 1985). Post-stratification has been performed following different criteria (Lo *et al.*, 2005). For the case of Atlanto-Iberian sardine, post-stratification has been applied in early applications (e.g. García *et al.*, 1992), based mainly on differences in adult characteristics. Nevertheless, Stratoudakis and Fryer (2000) showed that if spatial correlations between the different parameters of the DEPM exist, then post-stratification of the DEPM estimation procedure can lead to large bias. Solutions proposed include the use of survey indices that allow effort allocation proportional to some biomass indicators, and therefore use survey stratification rather than post-stratification (Stratoudakis and Fryer, 2000), or the use of spatial models of the parameters involved in the DEPM.

Different spatial models of egg abundance have been used (e.g. Bez *et al.*, 1995; Borchers *et al.*, 1997; Augustin *et al.*, 1998). For species in which the first egg stage at sea lasts for more than a day, and therefore that enough samples can be obtained in an ichthyoplankton survey throughout that day, estimates of egg production have been derived by assuming a constant mortality rate across the first stage and extrapolating the abundance of the first daily cohort back to that at the spawning time (Borchers *et al.*, 1997). Nevertheless, such models have not been used for sardine, because the egg stages are short and develop fast, so not enough information can be extracted from a single stage for use in estimating egg production.

Herein, a general spatially and environmentally explicit egg production model that allows information from all egg stages to be used for estimating egg production is derived from basic statistical assumptions. The model requires assumptions about the shape of the mortality curve that affects the numbers of eggs at different stages, and can use mortality estimates derived from sources different from the data used for the egg production estimation. Once the general model is derived, the results of recent analyses of the characteristics of Atlanto-Iberian sardine spawning behaviour and egg mortality (Bernal *et al.*, 2007, 2011) and a new method for assigning ages to sardine eggs (Bernal *et al.*, 2008) are used to provide the required assumptions to apply the general model to Atlanto-Iberic sardine DEPM data. The newly derived estimates of egg production are then compared with traditional ones, and the time-series of estimates of egg production is revised.

Material and methods

General spatially explicit egg production model

Instead of using the traditional egg production model derived from Equation (1), the distribution of eggs at sea can be modelled assuming a probability density function (pdf) that describes the statistical distribution of egg production and a survival probability that relates the observed number of eggs at given ages with the distribution of egg production. The expected number of eggs by cohort ($\mathrm{E}[N_a]$) can be modelled as a non-parametric generalized additive model (GAM; Hastie and Tibshirani, 1990; Wood, 2006; see the Supplementary material for the derivation of the model):

$$E[N_a] = g^{-1}(\text{offset} + s(\boldsymbol{x}_1,\ by = P_0) + s(\boldsymbol{x}_2,\ by = a)), \qquad (3)$$

in which the additive predictor is a sum of smooth functions of (i) a series of variables that affect egg production [$\boldsymbol{x}_1$, with the smooth function represented by $s(\boldsymbol{x}_1,\ by = P_0)$, noting that the term *by* in $s(x,\ by =)$ allows one to specify an interaction between the variable x and any other given variable (or factor)], and (ii) a series of variables that affect age (a) [$\boldsymbol{x}_2$, with the smooth function represented by $s(\boldsymbol{x}_2,\ by = a)$]. The second smoother represents the probability of egg survival (or the mortality curve) and is a generalization of the exponential mortality decay assumed in Equations (1) and (2). This equation follows the syntax of the implementation of a GAM in the statistical software and the computer language R (Ihaka and Gentleman, 1996), through the package mgcv (Wood, 2006). The implementation represents an extension of the original GAMs developed in Hastie and Tibshirani (1990), because each smoother can include more than one variable, and therefore each additive term can be multidimensional.

Equation (3) can be fitted to egg abundance by age assuming any discrete distribution from the exponential family (e.g. Poisson or negative binomial distribution, see Supplementary material). The selection of the appropriate smoothers can be carried out using general cross validation (GCV; Wood, 2006), and comparison between nested models to select the appropriate variables can be performed using a likelihood ratio test.

Using external mortality estimates

Although the general model described above can provide egg production and mortality estimates from data on abundance-at-age, there may not be enough contrast in the data from a single DEPM survey to obtain precise estimates of both egg production and mortality, as discussed in Bernal *et al.* (2011). Therefore, overparametrization and the lack of identifiability problems may arise in fitting the general model, and may lead to unrealistic estimates of the GLM or GAM parameters (Wood, 2006). To avoid this problem, a specific case of the general model that allows the use of estimates of mortality obtained independently from the fitting process for the egg production model is derived.

If both mortality rate and mean cohort age are assumed to be known, then the estimate of daily egg production using Equation (1) is simply the average of the observed egg densities corrected by the survival rate. If the survival probability is assumed to be an exponential function of age (see Supplementary material), then Equation (3) can be reformulated as

$$E[N_a] = g^{-1}(\text{offset} + s(\boldsymbol{x}_1, by = P_0)), \qquad (4)$$

where the offset now includes both the effective area and the correction attributable to survival:

$$\text{offset} = \log(\text{efarea}) - m\,a. \qquad (5)$$

Variance estimation

Variance estimation for the general model [Equation (3)] can be estimated by computer intensive methods (e.g. a bootstrap, as in

Borchers *et al.*, 1997), or by approximate Bayesian confidence intervals (Wood, 2006). For the case of the GAM production model with external mortality [Equation (4)], variance estimates can be obtained by resampling the offset of both age and mortality estimates [Equation (5)] and recreating the variance using any of the two options described above, in a procedure similar to that described by Buckland *et al.* (1993), as demonstrated below for Atlanto-Iberian sardine.

Application

Most of the Atlanto-Iberian sardine ichthyoplankton data used in this analysis were compiled within the European project SARDYN (Bernal *et al.*, 2007). In addition to these data, the results from a combined Spanish and Portuguese ichthyoplankton survey carried out in 2008 (ICES, 2009) were used. The geographic and temporal limits of the data used span the area between the Strait of Gibraltar in the south and the border of the Spanish and the French continental shelves, for the years 1985–2008 (Figure 1). Available data from the Armorican Shelf (southwest France) were not used, however, because they do not cover the temporal range of interest (1985–2008). All surveys from the database in which abundance by stage and temperature were available were used for the estimation of mortality in Bernal *et al.* (2011), but egg production here is only estimated for those years in which the survey was originally aimed to produce a DEPM estimate of Atlanto-Iberian sardine spawning biomass; 1988, 1997, 1999, 2002, 2005, and 2008. The same three spatial strata defined in Bernal *et al.* (2007, 2011) were used: (i) a north stratum that covers the Spanish coast in the northern part of the Iberian peninsula, (ii) a west stratum that covers the western Iberian area, from Cape Finesterre to Cape St Vicente, and (iii) a south stratum from Cape St Vicente to the Strait of Gibraltar (Figure 1).

Mortality

Egg mortality for Atlanto-Iberian sardine was estimated within the modelling process, using Equation (3), as well as from a mortality analysis carried out by Bernal *et al.* (2011) and incorporated into Equation (4). Mortality estimates using Equation (3) were obtained to illustrate the lack of significance problems stated in Bernal *et al.* (2011). For this purpose, an exponential mortality (e^{-ma}) curve, with m allowed to vary across spatial strata for a given year, was assumed to describe the survival rate in Equation (3). In this way, the functional form for mortality is the same as that used in traditional implementations (Lo, 1985; García *et al.*, 1992), and egg production is estimated as a non-parametric function of space (see the Supplementary material for further detail). External age and mortality estimates were obtained from the analysis of the complete Atlanto-Iberian sardine ichthyoplankton database (Bernal *et al.*, 2011). Criteria for excluding early (caused by incomplete recruitment to the sampler) and late stages (as a consequence of hatching) were the same as those adopted by Bernal *et al.* (2011).

Spatial model and general model structure

The spatial structure for the different models used was similar to that used by Bernal *et al.* (2007) to analyse Atlanto-Iberian

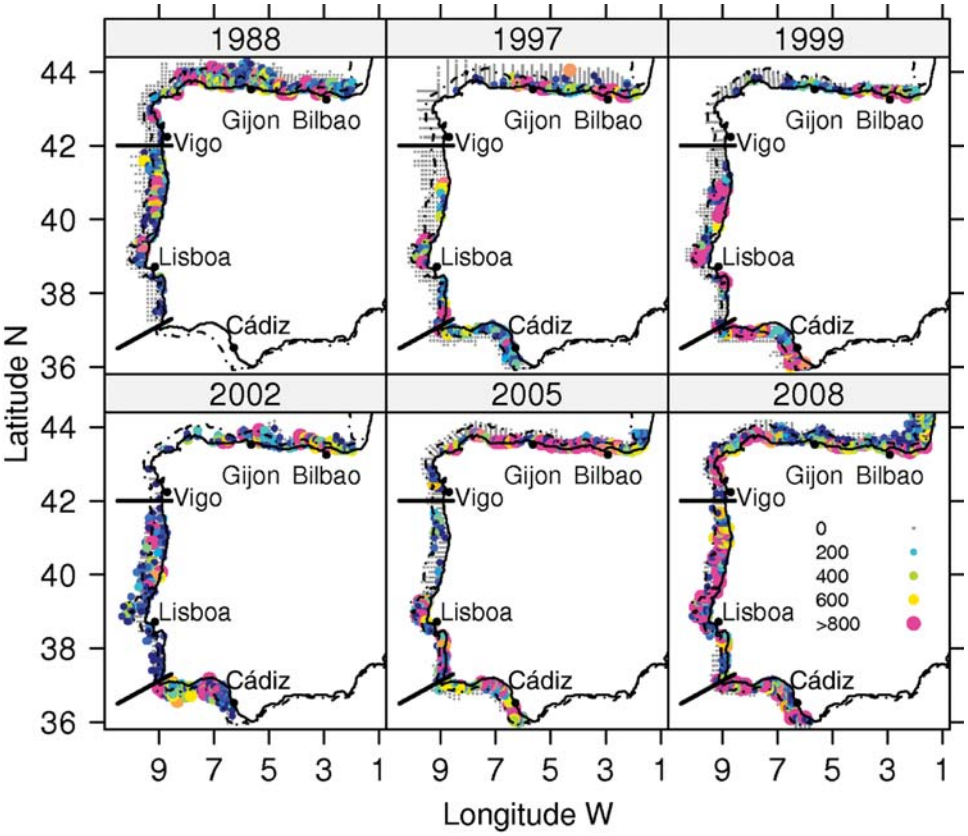

Figure 1. Observed egg densities (eggs m^{-2}) in the DEPM surveys off the Iberian Peninsula. Some existing data in the southern area in 1988 were not plotted because they were not used in the analysis. The abundance scale appears in the lower right panel.

sardine abundance data, and consisted of two bivariate smoothers, one including position (x, y) in Mercator projection, another including temperature (temp) and the logarithm of the bottom depth (logbot). Depth and temperature were included in the same smoother to account for the potential effect of hydrographic structures such as the slope current (Stratoudakis *et al.*, 2003; Bernal *et al.*, 2007). Variables in each smoother were standardized against each other to account for anisotropy, i.e. to allow each covariate within a smoother to use roughly the same degrees of freedom within its range (Wood, 2006).

Abundance of eggs for all cohorts within the imposed age limits, including those in which no eggs were observed, were used as the response variable. Error distribution was assumed to be Poisson with an estimated dispersion parameter, i.e. equivalent to a quasi-Poisson distribution (Wood, 2006). General cross validation (GCV) was used to select the appropriate degrees of freedom associated with each smoother.

The general model structure is

$$E[N_a] = g^{-1}(\text{offset} + s(x, y) + s(\text{logbot}, \text{temp})). \tag{6}$$

In one year, this model structure did not yield an adequate representation of the data, so for that year an alternative model structure was selected (see the Results section). The alternative simplified structure was based on a different geographic coordinate system, in which points in space are located by two variables; distance along the 100 m depth contour line (alongdist), and the distance perpendicular to the 100 m depth contour (perpdist; see Bernal *et al.*, 2007, for another use of this coordinate system).

Estimates of egg production by strata were obtained by spatial integration of the fitted values of egg production for each station within each stratum.

Variance estimation

For the models with external mortality estimates, variance estimates were obtained in the following way:

(i) Pseudosamples of the mortality estimate for each survey were obtained from the model fitted in Bernal *et al.* (2011), given the observed surface temperature at each station. A normal distribution of the mortality estimates was assumed, following the central limit theorem, and standard errors were estimated from the mortality fitting procedure (see Bernal *et al.*, 2011);

(ii) Pseudosamples of mean age by cohort for each station were obtained by sampling the posterior distribution of ages for each stage and temperature combination, following the multinomial egg development model and the age-determination procedure described in Bernal *et al.* (2008);

(iii) Pseudosamples of the offset used in Equation (6) were obtained by multiplying the pseudosamples of mortality and mean age by cohort, for all possible cohorts present at a given station, with an effective area of 1 m^2, to match units;

(iv) Iterative estimates of egg abundance for each station were obtained from the final model chosen to fit Equation (6), by resampling the posterior distribution of the fitted parameters (Wood, 2006) and using the original offset values;

(v) Egg abundances by cohort were raised to egg production using the pseudosamples of the offset obtained in steps (i)–(iii) above;

(vi) Coefficients of variation (CVs) and confidence intervals by strata were obtained using the vector of estimates for all cohorts for the stations within each stratum.

Table 1. An illustration of the mortality parameters by stratum (“south” includes Cádiz and Algarve, “west” includes all western Iberian coasts, and “north” includes the Cantabrian coast) estimated by the general model.

Year	South	West	North	Percentage deviance
1988		−0.014 [0.005]*	0.015 [0.003]**	44
1997	0.007 [0.017]	0.053 [0.011]**	−0.018 [0.006]**	38
1999	0.015 [0.009]	0.017 [0.007]*	−0.001 [0.007]	59
2002	−0.018 [0.008]*	0.029 [0.010]**	0.003 [0.005]	57
2005	−0.013 [0.006]*	0.019 [0.005]**	0.004 [0.004]	54
2008	0.028 [0.007]**	0.004 [0.005]	−0.004 [0.006]	52

Estimates of mortality and standard error (in square brackets) are presented by year and spatial stratum.

*Significance at $p < 0.05$.

**Significance at $p < 0.01$.

Estimates of the percentage of total deviance explained by the full model are presented for comparison with models with assumed mortality values (see Tables 2 and 3).

Table 2. Summary of the egg production models for the different surveys in the Atlanto-Iberian region.

Year	Initial d.f.	Associated d.f.		Percentage deviance
1988	49	$s(x,y,\ k = 21)$	$s(\text{Logbot,Fittemp},\ k = 23)$	42.91
1997	20	$s(\text{Alongdist,Perpdist},\ k = 17)$		37.31
1999	39	$s(x,y,\ k = 19)$	$s(\text{Logbot,Fittemp},\ k = 18)$	60.61
2002	45	$s(x,y,\ k = 19)$	$s(\text{Logbot,Fittemp},\ k = 19)$	38.02
2005	51	$s(x,y,\ k = 25)$	$s(\text{Logbot,Fittemp},\ k = 24)$	54.29
2008	63	$s(x,y,\ k = 30)$	$s(\text{Logbot,Fittemp},\ k = 29)$	49.67

Selected degrees of freedom (d.f.) are shown for the bivariate smoothers of each model. Variables x, Alongdist, and Fittemp are standardized against the respective covariate in the bivariate smoothers. See text for details on standardization and model structure.

Results

As shown by Bernal *et al.* (2007), the spatial distribution of egg abundance in samples changed over time, with a continuous presence of eggs throughout the northern Iberian shelf in 1988, and a more coastal distribution of eggs in the area in the other surveys (Figure 1). In the western area, the situation was more variable, some years having a greater presence of eggs (1988 and 2002), and other years showing concentrated patches of eggs (1997, 1999, and 2005).

Mortality estimates by year and region using Equation (3) were only significant and plausible, i.e. indicating a decreasing number of eggs with time, in 6 of 17 cases (Table 1), and four of the mortality estimates indicated significant negative mortality (i.e. “natality”, an increasing number of eggs with age). These estimates are therefore not plausible, so only the model with external mortality estimates [Equations (4) and (6)] is used to provide estimates of egg production by area.

Table 3. Estimates of egg production and *CV*s for all surveys by stratum ("south" includes Cádiz and Algarve, "west" includes all west Iberian coasts, and "north" includes the Cantabrian coast).

Model	South	West	North
1988	–	1.60 (0.09) [0.13]	2.90 (0.05) [0.08]
1997	1.08 (0.16) [0.23]	1.84 (0.10) [0.29]	1.38 (0.07) [0.14]
1999	6.97 (0.07) [0.13]	2.85 (0.09) [0.18]	0.62 (0.13) [0.14]
2002	2.49 (0.08) [0.19]	0.67 (0.14) [0.25]	1.38 (0.08) [0.11]
2005	2.11 (0.09) [0.14]	2.15 (0.08) [0.12]	2.08 (0.06) [0.09]
2008	5.51 (0.06) [0.18]	3.50 (0.07) [0.14]	2.50 (0.08) [0.11]

No data were available for the south in 1988. Egg production is in eggs × 10^{12} d^{-1}, the *CV* without variability in ages and mortality is provided in parenthesis, and the final *CV* including variability of ages and mortality in square brackets.

Deviance explained by the final models with fixed mortality [Equations (4) and (6)] ranged from 37% in 1997 to 61% in 1999 (Table 2). These values are larger than the deviance explained by the traditional mortality curve model (10–30%; ICES, 2009). The chosen model structure for 1997 was different from that for the other years. In 1997, the distribution of eggs in the north stratum was different from that in the rest of the time-series. In 1997, surface temperatures in the north were the highest of the years analysed, and in the same area, there were a few offshore stations with very great depth and medium-to-large egg densities (Figure 2). The combination of these two features has a large influence in the bivariate smoother of bottom depth and temperature (Figure 3b), in which the north and west strata show different characteristics. Also the bivariate smoother of x and y requires a large number of degrees of freedom to explain the gap in the NW corner of the Iberian Peninsula (Figure 3a), so the model is prone to overfitting. In comparison, the alternative alongdist–perpdist axis (Figure 3c) allows the northern Iberian area to be modelled independently of the west and south, because they have a different value on the alongdist axis. Also, the gap in southern Galician waters can be smoothly represented along the alongdist axis, using fewer degrees of freedom than in the previous model structure, but at the expense of a lesser percentage of deviance explained (results not shown).

Residual plots for the final models show a general pattern, with a few stations having large positive residuals and more stations with negative but smaller residuals (Figure 4). This is a common feature in egg production analysis, and may indicate a combination of spatial correlation and zero-inflated data not solved by the models. However, the residual patterns are less obvious than in a traditional analysis of the same data (results not shown), and previous analysis (e.g. ICES, 2004) has revealed egg production estimates to show robustness to residual patterns as well as to a small bias in mortality (McGarvey and Kinloch, 2001).

Figure 5 shows the predicted estimates of egg production for the different surveys, obtained from the final models listed in Table 2. The general patterns are coherent with the distribution of total egg abundance (Figure 1), and differences in egg production can be compared in space. For example, although both 1988 and 2002 had a continuous distribution of eggs along the shelf in the west, egg production there was clearly greater in 1988 than in 2002 (Figure 6). Different patches of high egg

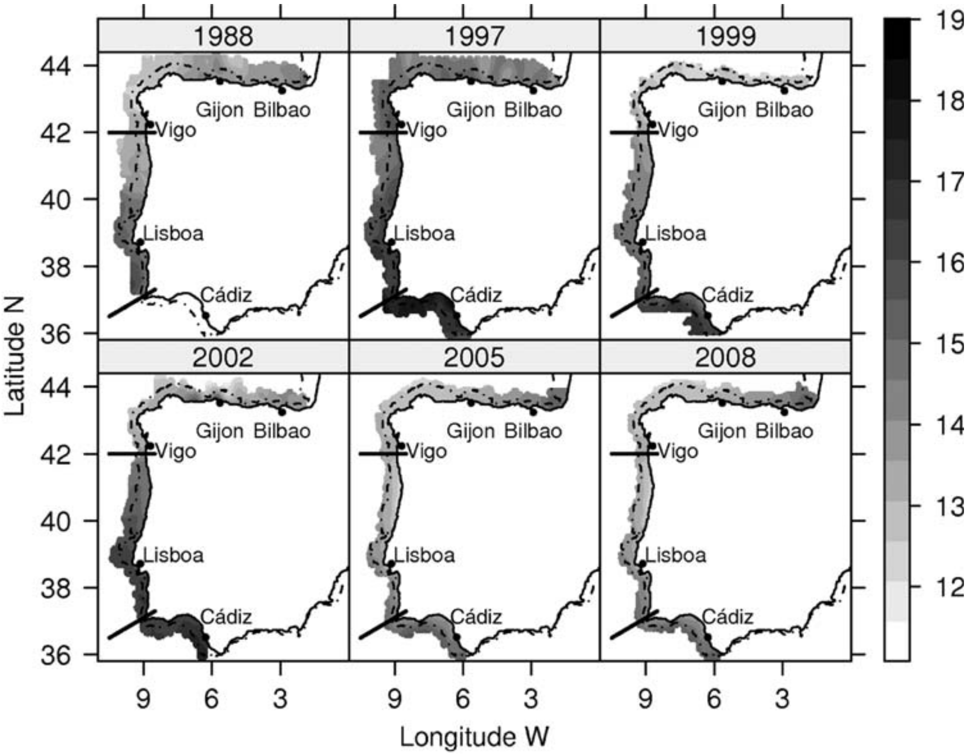

Figure 2. Temperature models along the Atlantic coast of the Iberian Peninsula for the years in which egg production is estimated. Note that surveys off the northern and western areas are not simultaneous, and for some years, not consecutive in time. The shading scale (in °C) is on the right.

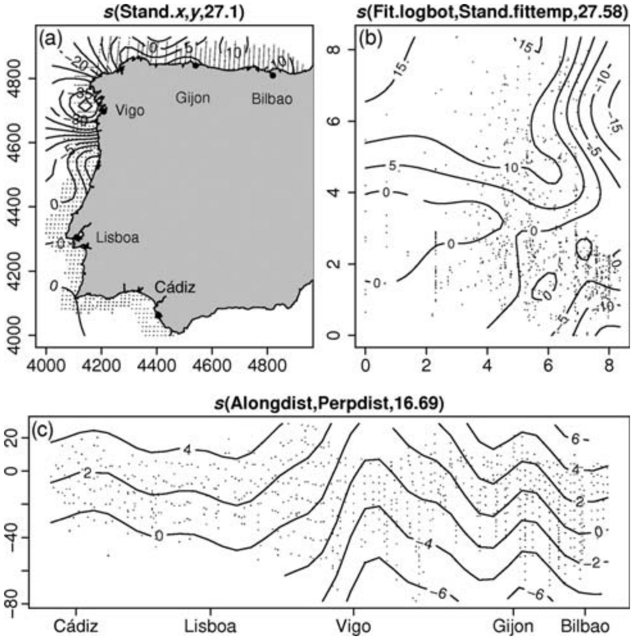

Figure 3. Partial effects of the different bivariate smoothers for the 1997 data, both in the original model structure (top panels) and the final one (bottom panel). In all panels, small dots indicate the observations, and the superimposed contour plot indicates the partial effect of each bivariate smoother. Numbers along the axis of the top panels do not have a direct meaning, because covariates are standardized within each smoother. Values on the *y*-axis of the bottom panel indicate distance to the 100 m depth contour, with positive values for greater depth. Labels on the *x*-axis indicate the relative position of some georeferences in the "Alongdist" scale (see text for further explanation).

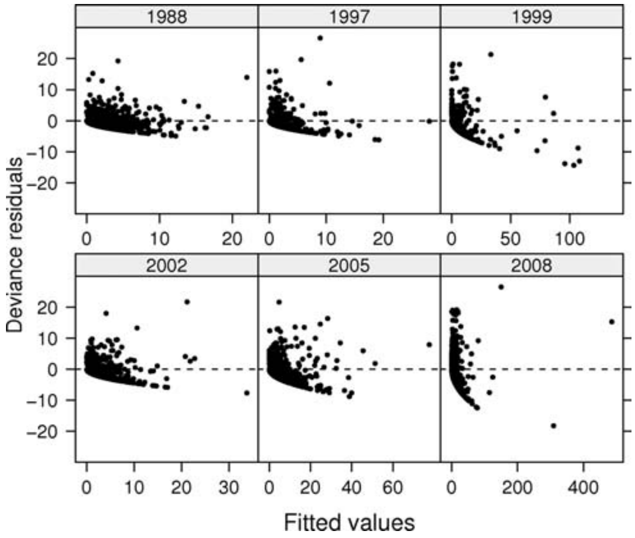

Figure 4. Deviance residuals of the egg production models listed in Table 2.

production are clear too in 1999 and 2005, in the west and south. Also, spawning near the coast in the north was high throughout the time-series, and differences in the degree of occupancy of the shelf clearly exist (Bernal *et al.*, 2007).

Estimates of egg production by region and year are similar to those obtained in the traditional analysis (ICES, 2009) for all years and spatial strata, except for 1997 in the south and 2002 in the north and the west (Table 3, Figure 6). Egg production *CV*s without variability of age and mortality range from 6 to 16% (Table 3), but including age and mortality variability increases these percentages from a fractional increase (from 13 to 14%) to nearly triple the original value (from 10 to 29%). The increase in *CV*s when age and mortality variability are integrated was higher in the south and the west than in the north. This is related to higher temperatures and associated greater estimates of mortality in those strata (Figure 2; see Bernal *et al.*, 2011). Total egg production by stratum ranges from 0.4×10^{12} (west in 2002) to 7×10^{12} eggs d^{-1} (south in 1999).

The temporal variation in all strata is similar to that depicted by traditional analysis (Figure 6), although differences in the south for 1999 change the perception of the trend there. Traditional estimates suggest similar egg production there for 1997 and 1999, and a slight decline in 2002 and 2005 (although confidence intervals are large and the former level slightly overlaps the latter), and a high egg production in 2008. Nevertheless, the new estimates suggest two pulses of egg production in the south in 1999 and 2008, and lower levels in other years. In general, the trend in the stock is dominated by the total egg production in the south, because that is the area that shows the greatest variability and the highest peaks in egg production.

Discussion

Spatial variability and environmental control in egg production and daily fecundity has been reported for several stocks (Somarakis *et al.*, 2004; Barange *et al.*, 2009), including the Atlanto-Iberian sardine stock analysed here (ICES, 2004; Bernal *et al.*, 2007; Silva, 2007). Incorporation of spatial and environmental variables in the Atlanto-Iberian sardine egg production model derived here yields more-precise estimates of egg production than previously reported, but still similar to traditional estimates for the stock. Also, the increase in precision and the spatially explicit results provide more insight into the trends of egg production by area, which shows great spatial variability and some degree of independence in spawning intensity across areas.

The results of the Atlanto-Iberian sardine egg production analysis are in accord with current knowledge of the spatial dynamics of the sardine along the European Atlantic coast, in that there are several population nuclei with some degree of interconnectivity but also a degree of independence (Silva, 2007; Bernal, 2008). Pulses of egg production in the north and south appear during different years. Some, like the one in 1999, may affect various areas, in that case the south and the west, whereas others affect just a single area. In terms of absolute levels of egg production, peaks in the south are higher than in the rest of the stock, whereas average levels without taking into account the peaks are generally larger in the west. Levels of egg production in the north are similar to those achieved in the west, although in some years they are slightly lower. Pulses are also obvious in the Atlanto-Iberian sardine recruitment time-series (ICES, 2006), without a clear temporal pattern but showing a generally low-frequency signal. Recruitment is also asynchronous for the different spawning nuclei of the stock, with nursery areas restricted to the west and the south, and no recruitment in the north (Silva, 2007). In terms of the overall stock, the indication is therefore that egg production in the area is wasted or produces recruits into

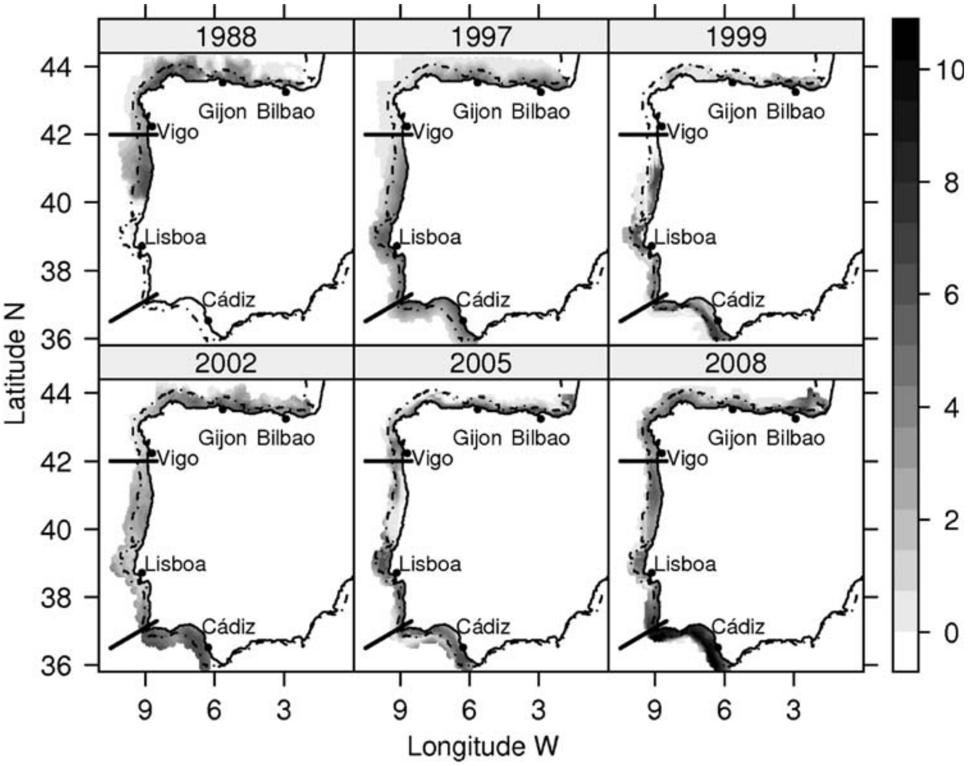

Figure 5. Predicted fields of egg production off the Iberian Peninsula for the different DEPM surveys. The abundance (log) scale (eggs m^{-2} d^{-1}) is on the right. Solid black lines represent the limits between the northern, western, and southern strata used in the analysis, but no data from the southern stratum were available in 1988.

different areas, which renders the population of some areas dependent upon inputs from the rest of the stock (Silva, 2007).

In addition to the results obtained for the case study presented here, the methods developed can in principle be applied to any other stock in which the DEPM is currently applied. Different statistical distributions from the exponential family, such as Poisson, quasi-Poisson, or negative binomial, can be used in the model (see Supplementary material). Overdispersed Poisson or negative binomial distributions allow one to accommodate different levels of skewness in the observations, so can better describe the variability in egg abundance observed at sea (Picquelle and Stauffer, 1985). This flexibility, however, comes at a cost in the model-fitting process, in the form of extra parameters and also in the appearance of identifiable problems, resulting from the confounding effects of variability as a result of the covariates used and the variability caused by the extra dispersion in the underlying distribution (Wood, 2006). For Atlanto-Iberian sardine, a compromise solution was taken, allowing for complex spatial structure and assuming a Poisson distribution, but allowing for overdispersion (similar to quasi-Poisson; Venables and Ripley, 2002). The resulting models explained a large percentage of the observed deviance and produced a reduction in the estimated residual dispersion parameter compared with traditional egg production estimators (∼4–8 residual dispersion here, vs. ∼20–170 in traditional models; results not shown). The final dispersion parameters indicate that overdispersion is being included in the model (Wood, 2006) without the extra parameters required for a negative binomial distribution. Even with this compromise, finding a common model structure for different surveys within a study system may not be possible, as in 1997 in this study. This situation is not unusual when complex spatial or environmental models are employed, underscoring the importance of a detailed validation of the models used to represent a given survey.

Another decision required to implement the egg production models presented here is how to model egg mortality. The general model derived allows mortality to be modelled as any parametric or non-parametric function of age and any other spatial or environmental covariates. This represents a generalization of the traditional mortality function used to fit data from a single survey, which assumes an exponential decay model with constant mortality (Lasker, 1985). However, for Atlanto-Iberian sardine, the analysis presented in Bernal *et al.* (2011) and the results here prove that the information from a single survey is not sufficient to obtain reliable estimates of mortality, even when a simplified traditional exponential mortality curve is used. This agrees with the results of Smith (1973) and McGarvey and Kinloch (2001), who suggested that the number of observations required to obtain significant mortality estimates is high. The approach taken here is to develop a method that allows mortality estimates obtained independently from the egg production methods to be integrated, although other approaches, such as using yolk-sac larvae to increase the age range of the data (Lo *et al.*, 2005), can also be implemented with the models developed here.

To conclude, the methods presented here provide a modern statistical framework for estimating egg production. The flexible model structure, and the possibility of estimating both mortality and egg production, or importing mortality from an external analysis, allows the models presented to be applicable to different populations for which the required data are available, including (but not restricted to) all stocks currently assessed by DEPM.

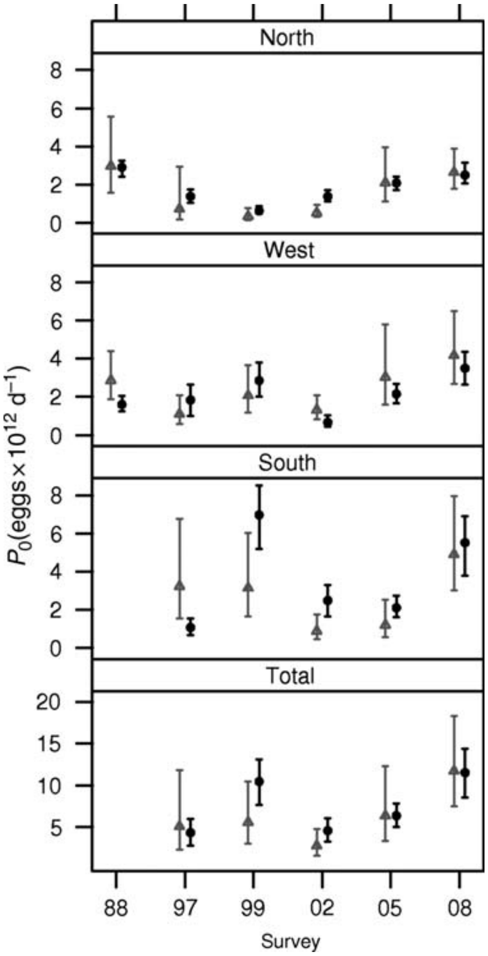

Figure 6. Egg production for the three spatial strata and for the total stock area for all surveys. Dots and lines indicate the GAM-based estimates of egg production and their confidence intervals, and grey triangles and lines the traditional estimates and their confidence intervals. The *y*-scale is the same for all strata, and different for the total stock area.

The application of such models provides more-precise estimates of egg production, by modelling the spatially and environmentally driven variability in the data, therefore also providing spatial representation of the results. The use of these models for Atlanto-Iberian sardine allows a consistent and more-precise revision of the time-series of egg production, which in turn provides improved ecological/biological information on the spawning behaviour of sardine and better understanding of spatio-temporal variability in egg production, so is expected to provide more reliable estimates of SSB for assessment purposes.

Supplementary material

The derivation of a general statistical egg production model is provided as Supplementary material at *ICESJMS* online. The model is derived from basic statistical assumptions and provides the statistical background for the generalized additive models fitted herein.

Acknowledgements

We acknowledge with thanks the work carried out in the different research institutes (IEO, IPIMAR, and AZTI) providing data for the analysis carried out here, as well as the crew of the research vessels used to collect the data. Early surveys used in this paper were funded by each institute and national research funding bodies, whereas most of the recent surveys were funded through the EU Data Collection Regulation. Data collection has been under the auspices of the ICES study and working groups SGSBSA and WGACEGG, and the EU project SARDYN. The manuscript benefited from discussion generated in EC Project 99/080 (using environmental variables with improved DEPM methods to consolidate the series of sardine and anchovy estimates), within the Research Unit for Wildlife Population Assessment (RUWPA) at St Andrews University, and within SGSBSA and WGACEGG, and we thank all participants of those groups for their support and ideas. Pierre Pepin and an anonymous referee provided valued comments on earlier versions of the manuscript, but the work is part of a doctoral dissertation by the first author, partially funded by an EU Marie Curie research grant. Finally, MB acknowledges IEO for allowing him the time to dedicate himself fully to finalizing his doctorate and the manuscripts produced therein.

References

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

Barange, M., Bernal, M., Cercole, M. C., Cubillos, L., Cunningham, C. L., Daskalov, G. M., De Oliveira, J. A. A., *et al.* 2009. Current trends in the assessment and management of small pelagic fish stocks. *In* Climate Change and Small Pelagic Fish, pp. 191–255. Ed. by D. M. Checkley, C. Roy, J. Alheit, and Y. Oozeki. Cambridge University Press, Cambridge, UK. xx + 372 pp.

Bernal, M. 2008. Modelling the abundance, distribution and production of sardine (*Sardina pilchardus*, Walb.) eggs in the Iberian Peninsula. PhD thesis, Universidad de Cádiz. 240 pp.

Bernal, M., Ibaibarriaga, L., Lago de Lanzós, A., Lonergan, M., Hernández, C., Franco, C., Rasines, I., *et al.* 2008. Using multinomial models to analyse data from sardine egg incubation experiments; a review of advances in fish egg incubation analysis techniques. ICES Journal of Marine Science, 65: 51–59.

Bernal, M., Stratoudakis, Y., Coombs, S., Angelico, M., Lago de Lanzós, A., Porteiro, C., Sagarminaga, Y., *et al.* 2007. Sardine spawning off the European Atlantic coast: characterization of and spatio-temporal variability in spawning habitat. Progress in Oceanography, 74: 210–227.

Bernal, M., Stratoudakis, Y., Wood, S., Ibaibarriaga, L., Uriarte, A., Valdés, L., and Borchers, D. 2011. A revision of daily egg production estimation methods, with application to Atlanto-Iberian sardine. 1. Daily spawning synchronicity and estimates of egg mortality. ICES Journal of Marine Science, 68: 519–527.

Bez, N., Rivoirard, J., and Walsh, M. 1995. On the relation between fish density and sea surface temperature, application to mackerel egg density. ICES Document CM 1995/D: 06. 17 pp.

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

Buckland, S. T., Breiwick, J. M., Cattanach, K. L., and Laake, J. L. 1993. Estimated population size of the California gray whale. Marine Mammal Science, 9: 235–249.

Cunha, M. E., Figueiredo, I., Farinha, A., and Santos, M. 1992. Estimation of sardine spawning biomass off Portugal, by the daily egg production method. Boletín del Instituto Español de Oceanografía, 8: 139–153.

García, A., Pérez, N., Lo, N. C. H., Lago de Lanzós, A., and Solá, A. 1992. The egg production method applied to the spawning biomass estimation of sardine, *Sardina pilchardus* (Walb.), on the North Atlantic Spanish coast. Boletín del Instituto Español de Oceanografía, 8: 123–131.

Gunderson, D. R. 1993. Survey of Fisheries Resources. John Wiley, New York. 256 pp.

Hastie, T., and Tibshirani, R. J. 1990. Generalized Additive Models. Chapman and Hall, London. 352 pp.

ICES. 2004. The DEPM estimation of spawning-stock biomass for sardine and anchovy. ICES Cooperative Research Report, 268. 91 pp.

ICES. 2006. Report of the Working Group on Acoustic and Egg Surveys for Sardine and Anchovy in ICES Areas VIII and IX (WGACEGG). ICES Document CM 2006/LRC: 01. 122 pp.

ICES. 2009. Report of the Working Group on Acoustic and Egg Surveys for Sardine and Anchovy in ICES Areas VIII and IX (WGACEGG). ICES Document CM 2009/LRC: 20. 177 pp.

Ihaka, R., and Gentleman, R. 1996. R: a language for data analysis and graphics. Journal of Computational and Graphical Statistics, 5: 299–314.

Lasker, R. (Ed). 1985. An egg production method for estimating spawning biomass of pelagic fish: application to the northern anchovy (*Engraulis mordax*). NOAA Technical Report, NMFS 36. 99 pp.

Lo, N. C. H. 1985. A model for temperature-dependent northern anchovy egg development and an automated procedure for the assignment of age to staged eggs. NOAA Technical Report, NMFS, 36: 43–50.

Lo, N. C. H., Macewicz, B. J., and Griffith, D. A. 2005. Spawning biomass of Pacific sardine (*Sardinops sagax*), from 1994–2004 off California. CalCOFI Reports, 46: 93–112.

Lotka, A. J. 1925. Elements of Physical Biology. Williams and Wilkins, Baltimore, MD.

McGarvey, R., and Kinloch, M. A. 2001. An analysis of the sensitivity of stock biomass estimates derived from the daily egg production method (DEPM) to uncertainty in egg mortality rates. Fisheries Research, 49: 303–307.

Picquelle, S. J., and Stauffer, G. 1985. Parameter estimation for an egg production method of anchovy biomass assessment. NOAA Technical Report, NMFS 36: 7–16.

Silva, A. 2007. Geographic variation in sardine population traits: implications for stock assessment. PhD thesis, Universidad do Algarve. 235 pp.

Simmonds, E. J., and MacLennan, D. N. 2005. Fisheries Acoustics: Theory and Practice, 2nd edn. Blackwell Science, Oxford, UK. 429 pp.

Smith, P. E. 1973. The mortality and dispersal of sardine eggs and larvae. Rapports et Procès-verbaux des Réunions du Conseil International pour l'Exploration de la Mer, 164: 282–292.

Somarakis, S., Palomera, I., Garcia, A., Quintanilla, L., Koutsikopoulos, C., Uriarte, A., and Motos, L. 2004. Daily egg production of anchovy in European waters. ICES Journal of Marine Science, 61: 944–958.

Stratoudakis, Y., Bernal, M., Borchers, D. L., and Borges, M. F. 2003. Changes in the distribution of sardine eggs and larvae off Portugal, 1985–2000. Fisheries Oceanography, 12: 49–60.

Stratoudakis, Y., Bernal, M., Ganias, K., and Uriarte, A. 2006. The daily egg production methods: recent advances, current applications and future challenges. Fish and Fisheries, 7: 35–57.

Stratoudakis, Y., and Fryer, R. 2000. Adult survey design and implications for sardine (*Sardina pilchardus*) DEPM estimation. ICES Document CM 2001/ACFM: 06.

Venables, W. N., and Ripley, B. D. 2002. Modern Applied Statistics with S, 4th edn. Statistics and Computing Series, Springer, Dordrecht. 495 pp.

Verhulst, P. F. 1839. Notice sur la loi que la population suit dans son accroissement. Correspondence in Mathematics and Physics, 10: 113–121.

Wood, S. N. 2006. Generalized Additive Models: an Introduction with R. CRC/Chapman and Hall, Boca Raton, FA. 416 pp.