Predictive models of northern anchovy *Engraulis mordax* and Pacific sardine *Sardinops sagax* spawning habitat in the California Current

E. D. Weber*, S. McClatchie

NOAA Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, California 92037-1508, USA

ABSTRACT: We modeled presence of anchovy and sardine eggs in the California Current from 1984 to 2009 using predictors related to water mass properties and primary productivity. Spawning areas could be predicted with good discrimination for anchovy, as indicated by logistic models with an area under receiver-operating characteristic curve (AUC; i.e. probability of concordance) of 0.86, and acceptable discrimination for sardine, as indicated by an AUC of 0.79. Knowledge of 3-dimensional characteristics of the water column was particularly important for predicting anchovy spawning habitat, as indicated by 2 variables: depth at which maximum chlorophyll *a* (chl *a*) occurred and an index of geostrophic flow. These variables were not useful predictors in the sardine model. For both species, variables related to low productivity limited modeled spawning probability most frequently during the 1980s and 1990s, but low temperature was most frequently limiting during the 2000s. Egg abundance was extremely patchy for both species; models to predict numbers of eggs conditional on their presence had almost no additional predictive ability for either species. However, most eggs occurred in areas where predicted probabilities of egg presence were relatively high. About 80% of anchovy eggs were captured in 40% of the samples that had modeled capture probabilities ≥0.25. About 80% of sardine eggs were captured in 35% of the samples that had predicted probabilities of capture ≥0.25. These models may be useful for understanding interannual changes in populations of predators that depend on anchovies and sardine. They could also improve the precision of fisheries surveys if predictors were collected beforehand using unmanned gliders.

KEY WORDS: Spawning habitat · Anchovy · Sardine · California Current · Logistic regression · Generalized additive models

INTRODUCTION

Northern anchovy *Engraulis mordax* and Pacific sardine *Sardinops sagax* populations fluctuate greatly in the California Current of western North America but together form the most abundant group of coastal pelagic fish species (Baumgartner et al. 1992, Lluch-Belda et al. 1992). As in other highly productive coastal systems, sardine and anchovy create important trophic links between zooplankton and larger avian, fish, and mammalian predators (Cury et al. 2000). Both species occur from British Columbia to Baja California during some years, but their ranges are reduced when population sizes are small (MacCall 1990, McFarlane et al.

2005) or changes in water temperature create unfavorable conditions in portions of their ranges (Rodríguez-Sánchez et al. 2002, Emmett et al. 2005). Both species are multiple spawners that release free-floating eggs into the water column that typically hatch within 3 to 4 d (Zweifel & Lasker 1976). Hence, most eggs occur near the spawning grounds and in similar environmental conditions to those in which spawning occurred. Although anchovies are less migratory and more restricted to nearshore areas than sardines, both species exhibit relatively large interannual variability in locations of peak spawning (Brewer & Smith 1982, Lo et al. 2007).

It is important to understand environmental processes that affect the spawning locations of anchovy

and sardine for several reasons. First, the distribution of spawning anchovies and sardines affect predators that depend on coastal pelagic fish prey. For example, the breeding success of the previously endangered California brown pelican *Pelecanus occidentalis californicus* depends on the abundance of anchovies in the region of nesting sites (Anderson et al. 1982). Recruitment of brown pelicans has been reduced during years when El Niño events caused anchovy to move offshore (Lentz et al. 2006). The distribution of spawning sardines may also have important regional effects on trophic interactions. Sardines occur offshore of Washington and Oregon, USA, only when water temperatures are sufficiently warm (Lo et al. 2007). These fish provide forage for larger salmon and may reduce predation pressure on emigrating juvenile salmon (Emmett et al. 2005). Another reason to predict the spawning habitat of anchovy and sardine is that egg densities may be used in conjunction with other information to estimate the size of the spawning stock via the daily egg production method (Lasker 1985). The method is currently used as part of the stock assessment for the sardine commercial fishery (Hill et al. 2009). If the areas where spawning is most likely to occur could be identified before a research cruise, sampling allocation could potentially be improved to provide better estimates (i.e. lower variance) for a given sampling effort or survey cost.

Most published studies have modeled anchovy or sardine spawning habitat as a function of water temperature and a measure of primary or secondary productivity such as chlorophyll (Fiedler 1983, Reiss et al. 2008), upwelling rate (Lluch-Belda et al. 1991), or zooplankton abundance (as estimated by acoustic backscatter; Lynn 2003). Checkley et al. (2000) reported that spawning habitat of sardines was associated with water masses having characteristic temperatures of 13.5 to 15°C and salinities <33.3 PSU. These studies demonstrated important relations between predictors and egg abundance but generally exhibited enough unexplained variability to suggest that additional predictors may have important effects on spawning habitat. Recruitment success has been correlated with different rates of upwelling, probably because upwelling rates affect the size structure of zooplankton upon which they prey (Rykaczewski & Checkley 2008). Sardine larvae located in offshore mesoscale eddies in the California Current also exhibit elevated survival rates, probably as a result of increased concentration of prey and greater encounter rates with them (Loggerwell & Smith 2001). Anchovy and sardine undergo strong selective pressure to spawn near areas where larvae can survive because eggs develop into larvae within a few days. Thus, horizontal and vertical water movement may also be important in determining appropriate spawning grounds.

The goal of the present study was to model the presence and abundance of anchovy and sardine eggs as proxies of spawning habitat. We aimed to develop predictive models using a longer time series than has been possible in most previous studies and considering potentially important predictor variables related to vertical properties of the water column, horizontal water movement, primary productivity, temperature, and salinity. We had several additional objectives that were meant to develop biologically meaningful and statistically robust models. They were that (1) the models should include environmental variables that could be measured remotely by unmanned vehicles (e.g. underwater gliders; Rudnick et al. 2004) and, thus, potentially used for prediction; (2) predictors must have a hypothesized mechanistic relation with spawning habitat rather than just a correlation (i.e. distance from shore); and (3) the models considered must be relatively parsimonious. They must exhibit biologically meaningful response functions or simpler functions if the data did not support more parameterized curves.

MATERIALS AND METHODS

Survey data. Eggs and oceanographic data were collected during cruises conducted in March, April, and May as part of the California Cooperative Oceanic Fisheries Investigations (CalCOFI; cf. Bograd et al. 2003) from 1984 to 2009. (We note that the CalCOFI program has existed since 1951, but chlorophyll was not sampled at all stations before 1984). Samples were collected in an approximate grid pattern centered on the Southern California bight and rotated −30° off the meridian so that it was oriented with the shoreline (Fig. 1). A total of 66 core stations were sampled each year except where logistical problems resulted in missed samples. Additional sampling was conducted intermittently within the core area or at stations located to the north or south of it. A few samples were taken south of 30° N latitude during the study period, but were excluded from the data set. This was to avoid a bias caused by sampling fish from the southern stock of sardines, which have been shown to have very different spawning characteristics from those in the spawning area of the California bight (e.g. Lluch-Belda et al. 1991). In a few cases, 2 ships were used for the survey; one ship collected fish eggs and the other collected oceanographic data. These data were merged if samples were collected within 4 km of each other on the same day. We assumed that spawning occurred near sites where eggs were present because of their short hatching time.

Eggs were collected using 0.71 m diameter bridleless bongo nets with 0.505 mm mesh nylon. Nets were
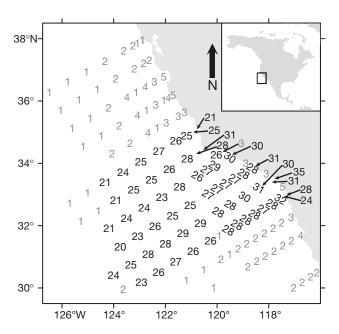
Fig. 1. CalCOFI sample sites used for model development. Numbers indicate the number of samples collected in March, April, or May 1984 to 2009 at their approximate geographical locations. Black numbers indicate core stations where sampling was attempted each year. Light gray area indicates the shoreline/land

towed obliquely at an angle of approximately 45° from 210 m depth to the surface as described by Smith & Richardson (1977) and Ohman & Smith (1995). Bongo net samples were treated as independent replicates because a previous study (Lo et al. 2001) reported that egg densities exhibited little or no spatial autocorrelation at distances greater than about 22 km, but most stations were about 74 km apart. Oceanographic data used to develop predictor variables for the model were chlorophyll *a* (chl *a*) concentration, dynamic height at 0/500 decibars, water temperature, and salinity. These data were measured or calculated based on bottle casts and conductivity temperature depth sensor (CTD) casts at each station. Variables were interpolated to the nearest 10 m for depths of 0 to 100 m, and at 125, 150, 200, 250, 300, 400, and 500 m. Detailed sampling protocols for the CalCOFI bottle and CTD samples are described by Lynn et al. (1982).

Model development. We used seven predictor variables to model presence and abundance of anchovy and sardine eggs. They were mean chl *a* concentration (µg l^{-1}) in the upper 50 m of the water column, depth (m) at which the maximum chl *a* concentration occurred, mean water temperature (°C) in the upper 50 m, mean salinity (PSU) in the upper 50 m, an index of geostrophic flow, day of the year, and an estimate of the species' stock size the previous year. Chlorophyll, temperature, and salinity were used as predictors of

food availability and physiological suitability of the habitat. These values were averaged over the upper 50 m because eggs and spawners of anchovy and sardine are typically distributed within this range (Curtis et al. 2007).

The 2 chlorophyll-related variables were selected to measure 2 related characteristics of the local environment. Mean chlorophyll concentration was a simple measure of the standing stock of primary producers that was potentially available for transmittal to higher trophic levels. The depth at which maximum chlorophyll occurred was used as an index of primary productivity (i.e. a rate). This is because nutrients are resuspended in the upper water column by upwelling and geostrophic flow during the spring transition in the California Current (Mantyla et al. 2008). High phytoplankton productivity occurs initially near the surface, where light is plentiful. As the spring season progresses, nutrients become limiting at the surface, and phytoplankton must adjust their depth to balance the needs for light and nutrients. This results in depths of maximum chlorophyll concentration that sink progressively through the season. Any other processes that cause shoaling of nutrients would affect the depth of maximum chlorophyll similarly. We hypothesized the relation between the probabilities of anchovy or sardine spawning in an area and maximum chlorophyll depth would be unimodal and skewed, with greatest probabilities occurring where maximum chlorophyll depth was near the surface. That is, that anchovy and sardine would select areas where high primary productivity had occurred for a period long enough to result in a concomitant increase in secondary productivity, but not so long that nutrient availability strongly limited productivity in the upper water column.

We included an index of geostrophic flow as a measure of horizontal flow, which could affect spawning and recruitment positively by increasing encounter rates with prey (Loggerwell & Smith 2001) or negatively by advecting eggs and larvae to inappropriate habitat (Lasker 1978). Geostrophic flow occurs perpendicular to the slope in dynamic height due to the Coriolis effect. Thus, the index was calculated based on a fitted surface in dynamic height for each year, which was estimated similarly to a method used to fit digital elevation maps to terrestrial slope data. First a surface was fit using the loess function (Cleveland & Grosse 1991) in the R programming environment, version 2.8.1 (R Development Core Team). The index of geostrophic flow was then calculated for each sample location as the slope of a line on the loess-estimated surface that extended for 10 km on each side of a sample location, where the line was oriented perpendicular to the direction of maximum slope. For points located on the outer edge of the surface, only the 10 km line that was

located within the bounds of the surface was used. Visual inspection of plots indicated the index matched contours in dynamic height well and, thus, provided a reasonable proxy for geostrophic flow.

We included 2 additional covariates to improve potential predictive ability of the models. Day of year was used to account for changes in spawning activity associated with the phase of the spawning season in which samples occurred. An estimate of the stock size during the previous year was also included as a blocking factor to estimate the degree to which spawning in otherwise suitable habitat was likely to be reduced simply because population sizes were low during some years. For sardines, we used the estimated spawning stock size for those aged 1 and greater, as reported by Hill et al. (2009). We used commercial landings as an index of anchovy population size (Pacific Fishery Management Council 2008) because no stock assessment for anchovy is currently conducted in the California Current.

Egg abundance could not be modeled directly as a function of the predictors for either anchovy or sardine because the number of zero catches was so great that they did not fit any exponential distribution. Instead, we modeled the probability of egg presence using logistic models, and then fit a second model for each species to estimate log-transformed egg abundance conditional on presence. This approach provided interpretable models to predict egg presence for each species (the logistic models) and the potential to calculate expected abundance as the product of the 2 models (i.e. a 2 stage estimate; estimated probability of occurrence × estimated abundance given that eggs are present; Welsh et al. 1996).

All models were fit semi-parametrically using the 'gam' function in the 'mgcv' package (Wood 2006) for R. The general form of the models was:

$$g(E(y)) = \beta_0 + \sum_k S_k(x_k) \qquad (1)$$

where $E(y)$ was the expected value of the response variable, $g(\cdot)$ was the link function defining the relation between the response and the linear model, β_0 was the intercept, $S_k(\cdot)$ was the smoothing function, and x_k was the value of the kth covariate. The response variable, y, was egg presence/absence for the logistic models or estimated density of eggs (number captured per 1000 m^{-3}) at sites where at least one egg was captured for the lognormal models. The link function, $g(\cdot)$, was a logistic link for the presence/absence models:

$$\log_e\left(\frac{\hat{y}}{1 - \hat{y}}\right) \qquad (2)$$

or a natural log transformation for the egg abundance models. The smoothing function was either a restricted cubic spline with shrinkage (the 'cs' curve in mgcv; cf. Wood 2006) or a parameter estimate if a term was entered as a simple linear predictor.

Several constraints were added to develop models that were parsimonious enough to prevent overfitting, which would result in poor predictive ability on other data, yet be flexible enough to be biologically realistic for a species' expected response along an environmental gradient (e.g. monotonic, unimodal, or skewed unimodal patterns). First, we limited the number of knots in the cubic splines to 3 for logistic models and 4 for the lognormal models. We note that a linear term in a model with a logistic link has a sigmoidal response curve after transformation back to probabilities, so only 3 knots are required to fit a skewed unimodal response pattern. The second constraint was that we increased the penalty per degree of freedom fit to each term by setting the 'gamma' option in the gam procedure to 1.4 to minimize potential overfitting (Wood 2006).

We performed model selection using the shrinkage features in the gam procedure rather than fitting a large set of potential candidate models (i.e. subsets of environmental variables fit with different amounts of flexibility for each term). The 'select' option was set to true for all models. This allowed coefficients with little or no predictive ability to be shrunken to zero, effectively dropping them from the model. In addition, we fit a final model for each species in which we replaced smoothed terms with their equivalent simpler term if the estimated degrees of freedom (df) in the initial gam indicated this was appropriate. That is, terms that were fit as splines in the initial gams were replaced by a spline with fewer df, or a linear term, if the substitution to the less parameterized model achieved the same fit. The stock size variable was initially entered as a linear term rather than a spline in the logistic models because a monotonically increasing response was the only biologically sensible response to increasing stock size.

The bootstrap (Efron & Tibshirani 1986) was used to estimate the variance of parameters and validate models. For each model, 1000 bootstrap resamples were conducted. Resamples were selected by sampling with replacement from the original data within each year, using a sample size equal to the number of points actually sampled for the year. These data were then aggregated to form a resample data set with the same total number of samples as the original data and which preserved its temporal structure. Resamples were selected randomly with replacement within years because there was little or no spatial autocorrelation, as described in 'Materials and methods' section 'Survey data'. Models were then fit using the same procedures described for the actual data. We calculated a variance/covariance matrix for each model based on the resamples, and estimated distribution-free standard errors for model para-

meters as the square root of the variance for each parameter. We also used the means of bootstrap resample values for measures of model specificity to provide better estimates of the true predictive ability of models on other data. Area under a receiver-operator characteristic curve (AUC) was the measure used for logistic models, and R^2 was used for the lognormal models. An AUC is a plot of the proportion of positive outcomes that were predicted correctly (true positive rate) as a function of the proportion of negative outcomes that were predicted incorrectly (false negative rate). Thus, the AUC corresponds to the probability of concordance and ranges from zero to one.

After fitting the logistic models, we conducted an exploratory analysis to determine which environmental variables were acting most strongly to limit predicted probabilities of capturing one or more eggs when all other variables were controlled. For each station sampled, we calculated the probability of capturing one or more eggs at the measured level of a variable and the median value of all other variables (i.e. the partial effects). The variable that resulted in the lowest predicted probability at the median of all others was judged to be most limiting at a station. We then calculated the proportion of stations at which each variable was limiting, by year and overall.

RESULTS

Anchovy spawning habitat

The logistic model to predict presence of anchovy eggs exhibited good predictive ability as indicated by a bootstrap-corrected AUC of 0.86. All variables except commercial landings (a proxy for stock size) were retained in the model (Table 1). The salinity and day-of-year variables were simplified to linear terms. The model had moderate variance as indicated by

standard errors for coefficients that generally were less than 30% of their absolute values.

Anchovies were most likely to spawn in highly productive areas, as indicated by estimates of partial effects for chl *a* concentration and depth of maximum chl *a* (Fig. 2A). The probability of spawning increased with increasing chl *a* concentrations in the range of 0 to 4 µg l^{-1} (about −3 to 1.4 on the log scale), where most of the observations occurred. The modeled probability of spawning remained high where phytoplankton blooms caused higher chlorophyll concentrations to occur. However, there were few measurements at higher concentrations, and confidence intervals were wide. The likelihood of capturing eggs also increased with the proximity to the surface of depth of maximum chlorophyll concentration. The probability of spawning decreased rapidly at water temperatures < ca. 14°C but was relatively stable between 14 and 18°C. Salinity was positively related to the probability of spawning throughout the modeled range; the linear term for salinity indicated that the data did not support a more parameterized model, such as a unimodal curve. The probability of spawning declined monotonically with day of year, indicating the peak spawning period for anchovy normally occurred before the spring study period. Spawning was most likely to occur in areas with relatively high geostrophic flow, which generally occurred near the shoreline.

The most commonly limiting single factor in the anchovy model was (low) water temperature, which affected the greatest percent of sites in 17 of 26 yr (Table 2). However, the total number of sites limited by either chlorophyll concentration or depth of maximum chlorophyll was greater than other single predictors in some of those years, indicating that some aspect of productivity was limiting. Productivity was most frequently limiting in all but one year before 1999. After 1999, temperature was most frequently limiting or tied every year except 2004. Productivity also tended to be

Table 1. Variable range, knot locations, and parameter estimates for logistic generalized additive models to predict presence of anchovy eggs. Variables with 2 coefficients were natural splines with 3 knots, resulting in 2 parameter estimates. The model was estimated using 1885 samples collected from 1984 to 2009. –: not applicable; note: predictor variables were scaled from zero to one before modeling (Wood 2006)

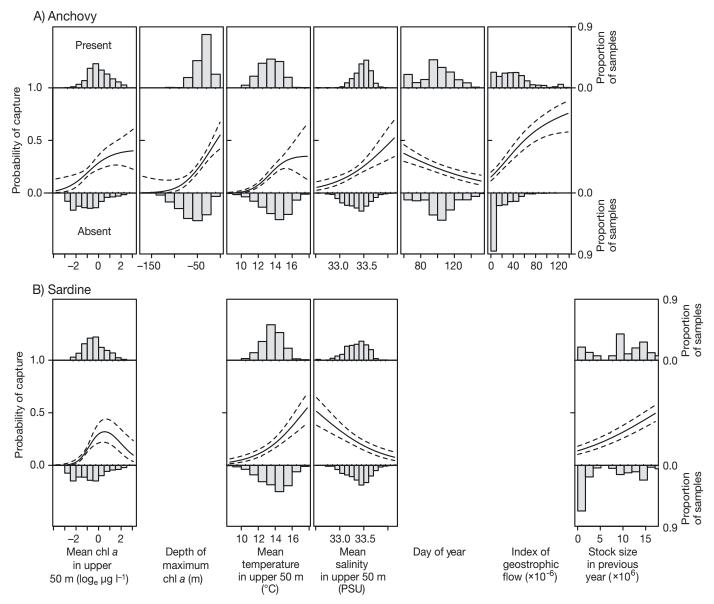| Variable	Range		Coefficient 1 (SE)		Coefficient 2 (SE)	
Intercept	–	–	−61.25	(15.22)	–	–
Log$_e$ chl *a* (µg l^{-1})	−3.79	3.08	2.36	(0.93)	1.19	(0.38)
Depth of max chl *a* (m)	−169	0	4.24	(2.22)	2.76	(0.64)
Water temperature (°C)	8.65	17.79	4.14	(0.63)	1.19	(0.58)
Salinity (PSU)	32.52	34.13	1.83	(0.46)	–	–
Day of year	60	152	−0.02	(0.00)	–	–
Index of geostrophic flow	6.27×10^{-11}	1.39×10^{-5}	−0.81	(0.11)	2.13	(0.40)
Landings[a]	1124	5720	–	–	–	–
[a]Effect was shrunk to zero						

Fig. 2. Partial effects of predictor variables on logistic models to predict presence of anchovy and sardine eggs. Solid lines indicate the predicted probability of capturing one or more eggs at a given level of the predictor variable and the median value of all other variables in the model. Broken lines indicate 95% confidence limits on the estimates. Histograms above and below the line graphs depict the sample distribution of egg presences (above) and absences (below) in relation to each predictor variable. The effects of the index of depth of maximum chl *a*, day of year, and geostrophic flow for sardines and stock size for anchovy are not plotted because they were not in the final models (they were estimated to be zero)

the limiting factor at a large proportion of sites even when temperature was most frequently limiting for a year overall, and vice-versa. Salinity, day of year, and geostrophic flow were important at some sites but were less frequently the strongest factors that limited modeled spawning probabilities. The mean annual probability of capturing eggs (i.e. mean of the response variable adjusted for all variables) ranged from 18 to 36% but did not exhibit a consistent pattern among years.

The lognormal model to predict densities of anchovy eggs conditional on their presence had bootstrap estimated R^2 of 0.24. However, the bootstrap estimated variance of the model was high. The mean relative standard error on predicted values was 0.56, indicating that many confidence intervals on the second stage included zero. We did not calculate expected densities for each site based on a 2-stage model because the 2-stage variance estimate is a function of the variances of the logistic and lognormal models. Given the moderate

Table 2. Annual results and percentage of sites sampled where modeled factors produced the smallest predicted probability of capturing anchovy eggs at the median values of all other factors (i.e. the limiting factor). **Bold** lettering indicates the factor that was limiting the greatest proportion of the time within a row (year or overall). The final column is the sum of the chl *a* and depth of maximum chlorophyll variables. It is listed to demonstrate when some aspect of productivity was limiting

Year	Landings (mt)[a]	Percent of sites with eggs (n)	Mean predicted probability of capturing eggs	Percent of sites where factor was limiting						Depth of max. Chl *a* + Chl *a*
				Temp- erature	Depth of max. chl *a*	Chl *a*	Salinity	Day of year	Geostrophic flow	
1984	4427	20 (193)	0.19	**31**	23	24	7	11	3	**47**
1985	2889	36 (61)	0.22	**48**	15	18	0	20	0	33
1986	1626	27 (59)	0.18	31	**32**	7	2	27	2	**39**
1987	1535	33 (128)	0.27	20	23	**27**	11	17	3	**50**
1988	1390	35 (51)	0.29	31	14	16	0	35	4	30
1989	1478	13 (63)	0.25	**43**	6	33	2	8	8	**39**
1990	2449	19 (106)	0.27	**38**	20	20	0	6	17	**40**
1991	3208	32 (41)	0.39	**32**	**32**	**32**	5	0	0	**64**
1992	4014	23 (65)	0.27	15	18	**32**	8	20	6	**50**
1993	1124	18 (65)	0.19	28	**29**	25	17	0	2	**54**
1994	1959	16 (64)	0.22	20	**42**	19	19	0	0	**61**
1995	1789	28 (53)	0.22	26	**43**	9	21	0	0	**52**
1996	1886	08 (59)	0.24	42	29	17	2	3	7	46
1997	4419	25 (61)	0.28	33	**34**	11	10	3	8	45
1998	5720	23 (104)	0.26	16	**49**	7	15	13	0	**56**
1999	1481	18 (61)	0.24	**70**	26	2	2	0	0	28
2000	5214	25 (65)	0.29	**42**	18	14	6	8	12	32
2001	11753	22 (65)	0.22	**46**	35	11	0	0	8	**46**
2002	19277	22 (63)	0.24	**56**	13	11	11	0	10	24
2003	4650	13 (64)	0.27	**44**	13	3	41	0	0	16
2004	1676	40 (60)	0.25	30	28	3	**38**	0	0	31
2005	6793	51 (72)	0.30	**38**	26	7	25	4	0	33
2006	11182	51 (72)	0.36	**43**	17	8	26	0	6	25
2007	12791	35 (54)	0.36	**41**	22	13	0	6	19	35
2008	10390	23 (66)	0.18	**65**	6	2	17	0	11	8
2009	10390	14 (70)	0.27	**40**	11	13	14	0	21	24
All Data		25 (1885)	0.25	**36**	24	16	11	8	6	**40**

[a]Commercial fishery landings reported by the Pacific Fishery Management Council (2008)

variance associated with the logistic model and the high variance associated with the lognormal model, the 2-stage estimate would have very poor predictive ability. However, most eggs occurred in areas where predicted probabilities of egg presence were relatively high. That is, egg densities were extremely patchy within the areas that they were likely to occur, but the logistic model could successfully predict a subset of sites that had a high probability of spawning and also contained most of the eggs (Fig. 3A). The range of predicted probabilities for the data used to develop the model was <0.01 to 0.97. In these samples, about 80 % of anchovy eggs were captured in about 40 % of the samples that had modeled capture probabilities ≥0.25.

Sardine spawning habitat

The logistic model to predict presence of sardine eggs exhibited acceptable predictive ability as indi-

cated by bootstrap-corrected AUC of 0.77. The final model contained chlorophyll concentration as a spline and temperature, salinity, and previous stock size as linear terms (Table 3). Depth of maximum chlorophyll, day of year, and geostrophic flow were dropped from the model. The model had moderate variance as indicated by standard errors for coefficients that generally were <25 % of their absolute values.

Estimates of partial effects indicated that sardine were likely to spawn in areas of intermediate primary productivity (Fig. 2B). The probability of capturing eggs increased with chl *a* concentration in the range of 0 to 4 µg l^{-1}, similarly to anchovy. However, eggs were never captured at higher chl *a* concentrations caused by phytoplankton blooms. Eggs were captured with greatest probability at higher temperatures and lower salinities. Similar to the anchovy model, sparse data near the ranges of the temperature and salinity data may have contributed to the simpler linear terms being selected rather than more parameterized splines. Eggs
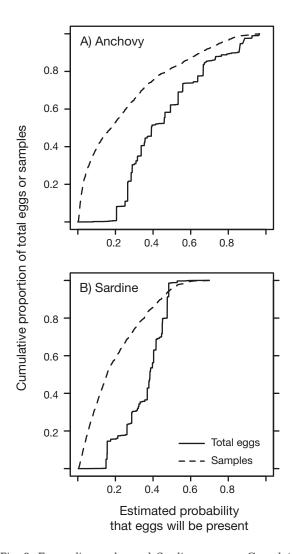
Fig. 3. *Engraulis mordax* and *Sardinops sagax*. Cumulative distributions of total eggs captured in CalCOFI samples (solid line) by probabilities of capturing one or more eggs, as predicted by logistic models for (A) anchovy and (B) sardine. Dashed lines indicate the cumulative proportion of stations sampled

were more likely to be captured at all sites during years when the estimated stock size was large the previous year.

Productivity, as indicated by mean chl *a* concentration, was the most frequently limiting factor modeled for sardine in 15 of 26 yr (Table $). When chl *a* concentration was limiting, it was usually because it was low (Fig. 2B). Similar to the anchovy model, productivity was most limiting during the 1980s and 1990s (14 of 16 yr sampled), and temperature was most frequently limiting thereafter (9 of 10 yr sampled). Temperature tended to be the limiting factor at many sites even when productivity was most limiting for the year on average, and vice-versa. Stock size was also a commonly limiting factor from 1984 until the mid 1990s, indicating that otherwise suitable habitat was not occupied during the period. Salinity was limiting at a large fraction of sites only during 1996 to 1998 and 2007, and thus appeared to be important during El Niño events. The mean annual predicted probability of capturing eggs ranged from 7 to 37% and exhibited a generally increasing trend over time. This was primarily due to the increase in stock size during the study period; when the means were recalculated using the median stock size for each year in the model, they did not exhibit a consistent trend.

The lognormal model to predict densities of sardine eggs conditional on their presence was similar to the corresponding anchovy model in that variance was so great that the model had little predictive ability. The bootstrap estimated R^2 for the model was 0.22. The mean relative standard error on predicted values was 1.2, indicating that many confidence intervals included zero. We did not calculate expected densities for each site using a 2-stage model based on this result. Despite the poor predictive ability of the least-squares model, most eggs occurred in areas where predicted probabilities of egg presence were relatively high for sardines (Fig. 3B). The range of predicted probabilities for the

Table 3. Variable range, knot locations, and parameter estimates for logistic generalized additive models to predict presence of sardine eggs. Variables with 2 coefficients were natural splines with 3 knots, resulting in 2 parameter estimates. The model was estimated using 1885 samples collected from 1984 to 2009. –: not applicable; note: predictor variables were scaled from zero to one before modeling (Wood 2006)

Variable	Range		Coefficient 1 (SE)		Coefficient 2 (SE)	
Intercept	–	–	43.96	(10.20)	–	–
Log$_e$ chl *a* (µg l^{-1})	−3.79	3.08	5.78	(1.16)	0.17	(0.54)
Depth of max chl *a* (m)[a]	−169	0	–	–	–	–
Water temperature (°C)	8.65	17.79	0.40	(0.07)	–	–
Salinity (PSU)	32.52	34.13	−1.56	(0.30)	–	–
Day of year[a]	60	152	–	–	–	–
Index of geostrophic flow[a]	6.27×10^{-11}	1.39×10^{-5}	–	–	–	–
Estimated stock size (mt)[b]	2904	1002330	1.80×10^{-6}	(1.68×10^{-7})	–	–

[a]Effect was shrunk to zero. [b]Estimated size of the spawning stock the previous year, aged 1 and greater (Hill et al. 2009)

Table 4. Annual results and percentage of sites sampled where modeled factors produced the smallest predicted probability of capturing sardine eggs at the median values of all other factors in the model (i.e. the limiting factor). **Bold** lettering indicates the factor that was limiting the greatest proportion of the time within a row (year or overall). Rows may not add up to 100% due to rounding error

Year	Estimated stock size[a]	Percent of sites with eggs (n)	Mean predicted probability of capturing eggs	Mean predicted probability of capturing eggs at median stock size[b]	Chl a	Temperature	Stock size	Salinity
					——— Percent of sites where factor was limiting ———			
1984	2904	3 (193)	0.09	0.14	**44**	29	27	0
1985	5292	7 (61)	0.07	0.11	**34**	**34**	28	3
1986	5919	3 (59)	0.09	0.13	**47**	29	24	0
1987	9029	5 (128)	0.10	0.15	**51**	7	42	0
1988	19674	14 (51)	0.10	0.15	29	25	**45**	0
1989	42191	8 (63)	0.09	0.12	**41**	24	35	0
1990	70887	4 (106)	0.09	0.13	**39**	33	28	0
1991	88376	34 (41)	0.12	0.16	**51**	12	37	0
1992	117160	8 (65)	0.18	0.22	**48**	14	38	0
1993	170236	12 (65)	0.12	0.13	**57**	28	15	0
1994	170178	023 (64)	0.11	0.13	**66**	19	16	0
1995	271031	30 (53)	0.16	0.16	**55**	38	8	0
1996	437942	27 (59)	0.16	0.12	**49**	41	0	10
1997	531859	44 (61)	0.21	0.15	**51**	28	0	21
1998	559613	43 (104)	0.25	0.18	**55**	23	0	22
1999	589564	36 (61)	0.20	0.12	21	**77**	0	2
2000	887809	17 (65)	0.31	0.14	**38**	**38**	0	23
2001	1002330	25 (65)	0.31	0.12	38	**57**	0	5
2002	878841	16 (63)	0.28	0.12	43	**54**	0	3
2003	785200	25 (64)	0.35	0.19	45	**55**	0	0
2004	610683	2 (60)	0.33	0.22	38	**62**	0	0
2005	730489	39 (72)	0.36	0.21	43	**51**	0	6
2006	847585	31 (72)	0.37	0.18	42	**57**	0	1
2007	949717	50 (54)	0.37	0.16	30	28	0	**43**
2008	867100	33 (66)	0.32	0.14	21	**74**	0	5
2009	662886	31 (70)	0.29	0.17	47	**53**	0	0
	All Data	20 (1885)	0.20	0.15	**44**	37	15	5

[a]Reported in Hill et al. (2009)
[b]Stock size was held at its median value but measured values of all other predictors were entered

data used to develop the model was <0.01 to 0.70. About 80% of sardine eggs were captured in about 35% of the samples that had predicted probabilities of capture ≥0.25.

DISCUSSION

The logistic models had good ability to predict spawning habitat of anchovies and sardine but exhibited important differences between species. The sardine model was less parameterized than the anchovy model. It was similar to previously reported studies (e.g. Checkley et al. 2000, Lynn 2003), in that only temperature, salinity, and chl a concentration were needed to predict spawning habitat within a year (the stock-size term was a blocking variable among years). This result suggests that understanding 3-dimensional character-

istics of the water is less important to successfully predict spawning habitat for sardine than it is for anchovies. Although the variables in the sardine model were averages over the upper 50 m of the water column, these averages are normally highly correlated with their corresponding measures at the surface. Two of the predictors in the anchovy model required detailed knowledge of the water column: the depth at which maximum chl a concentration occurs and the index of geostrophic flow. These additional explanatory variables partly explain the much greater predictive power that the anchovy model reported here had compared to a recently published model that used remotely sensed satellite data only (Reiss et al. 2008). The need for 3-dimensionally resolved data to adequately characterize spawning habitat for anchovies creates a potential problem for prediction because such data are not currently measured across the potential spawning area

before research cruises. It is likely that 3-dimensional models such as those reported here will become more useful as unmanned glider technology becomes less expensive and more available (Rudnick et al. 2004).

The differences in variables selected for each model were reasonable given the differences in life history between the 2 species. The depth of maximum chlorophyll variable likely was a better predictor variable for anchovies than sardines because strong shoaling of nutrients supports larger zooplankton prey items. Anchovies have larger gill rakers than sardines and thus require larger prey (Rykaczewski & Checkley 2008). This is also consistent with the pattern observed worldwide that anchovies are more closely associated with areas of high primary productivity than are sardines, probably because anchovies are smaller and less migratory (Bakun & Broad 2003). The mean chlorophyll variable may have been a sufficient measure of productivity in the sardine model because sardines can take advantage of smaller prey items associated with less strong shoaling of nutrients, and may move farther from areas of high productivity simply because of their greater mobility.

The association between stronger geostrophic flow and spawning for anchovy, but not sardine, may have been for similar reasons. Anchovy took advantage of inshore areas where the greater flow in the spring created highly productive but relatively stable conditions. Anchovy eggs were most frequently captured in the southeasterly current that develops along the coast as part of the spring transition in the Southern California bight each year (Lynn et al. 2003). This inshore current brings productive water into an area where vertical stability and other environmental conditions are conducive to spawning. Wind-induced turbulence in the area creates vertical mixing that is sufficient to suspend nutrients in the upper water column but not so great that feeding of larvae is disrupted (Husby & Nelson 1982). Offshore transport is also weak, creating relatively optimal conditions for survival of larval anchovies. This may explain the difference in results from those reported by Twatwa et al. (2005). They reported anchovies were more likely to spawn in moderate current and wind in the Benguela current of South Africa and were less associated with strong currents than sardines. The different results likely occurred because greater current strength was associated with greater wind speed, greater advection offshore, and reduced vertical stability in the Benguela current, but not so in the Southern California bight.

We chose predictor variables that had hypothesized mechanistic relations with spawning habitat and which were not highly correlated with each other. Nevertheless, these results should not be interpreted as evidence that mechanistic relations existed because the study was correlative. One obvious link to other variables is that high geostrophic flow and chl *a* concentrations are associated with inshore areas where anchovy commonly occur. Some other factors associated with distance from shore may be more important than these 2 in controlling anchovy spawning. Despite this limitation, bootstrap simulations indicated the models were robust. One explanation for this is that we used variables that were not highly correlated. Coefficients of determination between pairs of predictors were all less than 0.50. This suggests that even if the predictors used in the models did not affect selection of spawning areas directly, they must have been related to different environmental effects.

In the anchovy model, the day of year variable declined monotonically reflecting the fact that peak spawning usually occurs just before the study period, in late January to April (Hunter & Macewicz 1980). We note that no data were available for the peak spawning period because CalCOFI cruises are conducted quarterly. The extent to which anchovies might change behavior during the spawning season, and thereby cause a bias in the model, is unknown. Day of the year was occasionally the most limiting variable at a sample location for anchovy (Table 2). This indicates probabilities of capture should be adjusted for time of sampling in similar models. Day of year probably was not a good predictor in the sardine model because the CalCOFI spring cruise occurs near the peak spawning period, and it is relatively long.

The sardine model also indicated that it is important to adjust for the fact that spawning may not occur in suitable habitat simply because there were not enough spawners to fully occupy it (sensu MacCall 1990). This appeared to be the case for sardine at a large fraction of sites during the early part of the study, 1984 to 1995 (Table 4), when the population was recovering from a period of very low numbers. We cannot explain why the effect was not a good predictor for anchovies. We speculate that smaller numbers of anchovies tend to be more evenly distributed over suitable spawning habitat than sardines, even when their populations are relatively small. Anchovies are less mobile than sardines and thus may be comparatively less able to seek out and congregate at the most optimal spawning habitat over larger geographic areas.

The probabilities of capturing eggs estimated by our models are assumed to be directly related to realized spawning habitat, the area where spawning actually occurs (Planque et al. 2007). However, our estimated probabilities will be lower by an unknown amount than the probability that spawning occurs. This is because there is a chance that no eggs will be captured in the nets even if they are present (Mangel & Smith 1990). The area of water sampled by the nets is smaller

than the localized area with similar environmental conditions that represents a patch of suitable spawning habitat. Within a localized area of suitable habitat, spawners and eggs are likely to be patchily distributed due to schooling and other biological interactions. We assumed that the probability of capturing eggs where they were present was constant among all sample locations and thus did not create a bias in our modeled probabilities. The blocking variables, day of year and previous stock size, were important to include in the models to adjust for the fact that habitat could be suitable for spawning (i.e. theoretical spawning habitat; Planque et al. 2007), but spawning might not occur if fish were not available at the time of sampling. The models implicitly assumed that some spawners could locate appropriate habitat anywhere within the CalCOFI area if they were present because no adjustment was made for the spatial distribution of spawners before annual surveys, which was unknown.

Variables related to productivity tended to limit the spawning areas of both species until about the year 2000. Temperature tended to be most limiting thereafter. The mean chl *a* concentration in the upper 50 m across all sample sites was about 1.0 µg l^{-1} before 2000 and 1.3 µg l^{-1} after 2000. The corresponding temperature in the upper 50 m decreased from about 14.0 to 13.5°C. These changes were primarily the result of long-term fluctuations in ocean conditions (e.g. Mantua et al. 1997). However, the overall modeled probability of spawning did not exhibit a consistent pattern for anchovies or sardines after controlling for stock size. This indicated the trade-off between warmer water and more productive but cooler water transported by the California Current was about balanced in terms of spawning probabilities. Although the change in conditions did not have a consistent effect on likelihood of spawning, the different conditions may have affected the ultimate survival of larvae differently (i.e. successful spawning habitat; Planque et al. 2007). The extent to which the different conditions affected later recruitment success is unknown.

The partial effects reported here differed from previously reported species' response curves for predictor variables because we used relatively parsimonious models. For example, the temperature and salinity effects for both species were monotonic rather than exhibiting peaks followed by declines at higher levels, as reported in previous literature (e.g. Lluch-Belda et al. 1991, Checkley et al. 2000). Although our models were based on large sample sizes, the data often were sparse at the extreme ranges of an environmental gradient, where the conditions measured seldom occurred. Thus it was more difficult to estimate shapes of response curves at the tails. If we had fit more parameterized models (i.e. using more knots), the functions

would have had more flexibility to fit the tails but would have had the undesirable effect of fitting random variability in the data set. In our experience, the constraints we used were a good compromise between allowing for realistic biological responses and avoiding overfitting of the data. They should not be interpreted as physiological response curves that would be observed under more controlled conditions.

The ability to predict presence or absence of eggs was important because egg densities were extremely patchy for both species, but patches were concentrated within relatively optimal spawning habitat. Thus, areas in which most eggs occurred could still be estimated, despite poor ability of the lognormal models to predict numbers of eggs at a sample location (i.e. the second stage of a 2-stage abundance model). The extreme variability in egg densities reported here was likely the result of several factors. First, biological interactions create patchy distributions of spawners within localized areas of suitable habitat, as described above. Second, patches of eggs generally are larger than the area sampled in a particular net tow. This means that a few eggs could be collected from the edge of a large patch, or a very large number from the center of the patch, depending on a small difference in the location of a net tow. Previous studies of coastal pelagic species have also indicated that modeling presence/absence rather than abundance can provide nearly the same amount of information for a given level of variance when fish distributions are very patchy and contain many zero catches (Mangel & Smith 1990). This is probably the case for most coastal pelagic species worldwide.

We did not include interaction terms in our models because preliminary analysis indicated they did not provide interpretable models. The introduction of an interaction term generally resulted in one of the main effects being shrunk to zero. We could think of no biologically plausible reason to include interaction terms in the absence of main effects.

Several variables that were collected as part of the CalCOFI program had some value for predicting spawning habitat but were not entered into the models because they were highly correlated with other predictors used. For example, depth at which maximum chlorophyll concentration occurred was highly correlated with nitricline depth (r = 0.81) and mixed-layer depth (r = 0.84). In these cases, we used the variable that was trophically closer to anchovy and sardine, or which had a more direct hypothesized mechanism for influencing spawning. We also found that simpler models could be constructed which had less predictive ability than those reported but which may be easier to calculate. The depth at which maximum oxygen concentration occurred had predictive value for both spe-

cies, but was correlated with depth of maximum chlorophyll concentration (r = −0.83), water temperature (r = −0.79), and the index of geostrophic flow (r = 0.63). This variable alone could be used to provide a crude estimate of spawning habitat for both species if oxygen profiles were the only data available. It is also possible that more refined models of spawning habitat could be constructed using physical predictors and egg data that are more finely resolved. For example, it may be fruitful to examine egg data collected using the Continuous Underway Fish Egg Sampler (Checkley et al. 2000) if physical variables such as productivity were collected in conjunction and a sufficiently long time series were available.

LITERATURE CITED

Anderson DW, Gress F, Mais KF (1982) Brown pelicans: influence of food supply on reproduction. Oikos 39:23–31

Bakun A, Broad K (2003) Environmental 'loopholes' and fish population dynamics: comparative pattern recognition with focus on El Nino effects in the Pacific. Fish Oceanogr 12:458–473

Baumgartner TR, Soutar A, Ferreira-Bartrina V (1992) Reconstruction of the history of pacific sardine and northern anchovy populations over the past 2 millennia from sediments of the Santa-Barbara basin, California. CalCOFI Repts 33:24–40

Bograd SJ, Checkley DA Jr., Wooster WS (2003) CalCOFI: a half century of physical, chemical, and biological research in the California Current System. Deep-Sea Res II 50: 2349–2353

Brewer GD, Smith PE (1982) Northern anchovy and Pacific sardine spawning off southern of the nearshore coastal region California during 1978–80: preliminary observations on the importance of the nearshore coastal region. CalCOFI Repts 23:160–171

Checkley DM Jr., Dotson RC, Griffith DA (2000) Continuous, underway sampling of eggs of Pacific sardine (*Sardinops sagax*) and northern anchovy (*Engraulis mordax*) in spring 1996 and 1997 off southern and central California. Deep-Sea Res II 47:1139–1155

Cleveland WS, Grosse E (1991) Computational methods for local regression. Stat Comput 1:47–62

Curtis KA, Checkley DM Jr., Pepin P (2007) Predicting the vertical pro?les of anchovy (*Engraulis mordax*) and sardine (*Sardinops sagax*) eggs in the California Current system. Fish Oceanogr 16:68–84

Cury P, Bakun A, Crawford RJM, Jarre A, Quiñones RA, Shannon LJ, Verheye HM (2000) Small pelagics in upwelling systems: patterns of interaction and structural changes in 'wasp-waist' ecosystems. ICES J Mar Sci 57:603–618

Efron B, Tibshirani R (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci 1:54–77

Emmett RL, Brodeur RD, Miller TW, Pool SS, Krutzikowsky GK, Bentley PJ, McCrae J (2005) Pacific sardine (*Sardinops sagax*) abundance, distribution, and ecological relationships in the Pacific northwest. CalCOFI Repts 46:122–143

Fiedler PC (1983) Satellite remote sensing of the habitat of spawning anchovy in the southern California bight. CalCOFI Repts 26:202–209

Hill KT, Dorval E, Lo NCH, Macewicz BJ, Show C, Felix-Uraga R (2009) Assessment of the Pacific sardine resource in 2009 for US management in 2010. NOAA Tech Rep NOAA-TM-NMFS-SWFSC-452. US Department of Commerce, La Jolla, CA

Hunter JD, Macewicz BJ (1980) Sexual maturity, batch fecundity, spawning frequency, and temporal pattern of spawning for the northern anchovy, *Engraulis mordax*, during the 1979 spawning season. CalCOFI Repts 21:139–149

Husby DM, Nelson CS (1982) Turbulence and vertical stability in the California Current. CalCOFI Repts 23:113–129

Lasker R (1978) The relationship between oceanographic conditions and larval anchovy food in the California Current: identification of factors contributing to recruitment failure. Cons Int Explor Mer 173:212–230

Lasker R (ed) (1985) An egg production method for estimating spawning biomass of pelagic fish: application to the northern anchovy (*Engraulis mordax*). NOAA Tech Rep NMFS 36. US Dept of Commerce, La Jolla, CA

Lentz JE, Gaede P, DesJardin D (2006) Introduction to birds of the southern California coast. University of California Press, Berkeley, CA

Lluch-Belda D, Lluch-Cota DB, Hernandez-Vazquez S, Salinas-Zavala CA, Schwartzlose RA (1991) Sardine and anchovy spawning as related to temperature and upwelling in the California Current System. CalCOFI Repts 32:105–111

Lluch-Belda D, Schwartzlose RA, Serra R, Parrish R, Kawasaki T, Hedgecock D, Crawford RJM (1992) Sardine and anchovy regime fluctuations in four regions of the world oceans: a workshop report. Fish Oceanogr 1:339–347

Lo NCH, Hunter JR, Charter R (2001) Use of a continuous egg sampler for icthyoplankton surveys: application to the estimation of daily egg production of Pacific sardine (*Sardinops sagax*) off California. Fish Bull 99:554–571

Lo NCH, Macewicz BJ, Griffith DA, Charter RL (2007) Spawning biomass of Pacific sardine (*Sardinop sagax*) off U.S. and Canada in 2006. NOAA Tech Memo NMFS-SWFSC- 401. US Dept of Commerce, La Jolla, CA

Loggerwell EA, Smith PE (2001) Mesoscale eddies and survival of late stage pacific sardine (*Sardinops sagax*) larvae. Fish Oceanogr 10:13–25

Lynn RJ (2003) Variability in the spawning habitat of Pacific sardine (*Sardinops sagax*) off southern and central California. Fish Oceanogr 12:541–553

Lynn RJ, Bliss KA, Eber LE (1982) Vertical and horizontal distributions of seasonal mean temperature, salinity, sigma-t, stability, dynamic height, oxygen, and oxygen saturation in the California Current, 1950–1978. CalCOFI Atlas No 30. La Jolla, CA

Lynn RJ, Bograd SJ, Chereskin TK, Huyer A (2003) Seasonal renewal of the California Current: the spring transition of California. J Geophys Res, 108, C8, doi:10.1029/2003JC001787

MacCall AD (1990) Dynamic geography of marine fish populations. Washington University Press, Seattle, WA

Mangel M, Smith PE (1990) Presence-absence sampling for fisheries management. Can J Fish Aquat Sci 47:1875–1887

Mantua NJ, Hare SR, Zhang Y, Wallace JM, Francis RC (1997) A Pacific decadal climate oscillation with impacts on salmon. Bull Am Meteorol Soc 78:1069–1079

Mantyla AW, Bograd SJ, Venrick EL (2008) Patterns and controls of chlorophyll-*a* and primary productivity cycles in the Southern California bight. J Mar Syst 73:48–60

McFarlane G, Schweigert J, Macdougall L, Hrabok C (2005) Distribution and biology of Pacific sardines (*Sardinops sagax*) off British Columbia, Canada. CalCOFI Repts 46: 144–160

Ohman MD, Smith PE (1995) A comparison of zooplankton sampling methods in the CalCOFI time series. CalCOFI Repts 36:153–158

Pacific Fishery Management Council (2008) Status of the Pacific coast coastal pelagic species fishery and acceptable biological catches; stock assessment and fishery evaluation—2008. Pacific Fishery Management Council. Portland, OR

Planque B, Bellier E, Lazure P (2007) Modelling potential spawning habitat of sardine (*Sardina pilchardus*) and anchovy (*Engraulis encrasicolus*) in the Bay of Biscay. Fish Oceanogr 16:16–30

R Development Core Team (2008) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Reiss CS, Checkley DM Jr., Bograd SJ (2008) Remotely sensed spawning habitat of Pacific sardine (*Sardinops sagax*) and northern anchovy (*Engraulis mordax*) within the California Current. Fish Oceanogr 17:126–136

Rodríguez-Sánchez R, Lluch-Belda D, Villalobos H, Ortega-García S (2002) Dynamic geography of small pelagic fish populations in the California Current system on the regime time scale (1931–1997). Can J Aquat Sci 59: 1980–1988

Rudnick DL, Davis RE, Eriksen CC, Fratantoni DM, Perry MJ (2004) Underwater gliders for ocean research. Mar Technol Soc J 38:73–84

Rykaczewski RR, Checkley DM Jr. (2008) Influence of ocean winds on the pelagic ecosystem in upwelling regions. Proc Natl Acad Sci USA 105:1965–1970

Smith PE, Richardson SL (1977) Standard techniques for pelagic fish egg and larva surveys. FAO Fish Tech Pap 175: l–100

Twatwa NM, Van Der Lingen CD, Drapeau L, Moloney CL, Field JG (2005) Characterising and comparing the spawning habitats of anchovy *Engraulis encrasicolus* and sardine *Sardinops sagax* in the southern Benguela upwelling ecosystem. Afr J Mar Sci 27:487–499

Welsh AH, Cunningham RB, Donnelly CF, Lindenmayer DB (1996) Modelling the abundance of rare species: statistical models for counts with extra zeros. Ecol Model 88: 297–308

Wood SN (2006) Generalized additive models: an introduction with R. Chapman & Hall/CRC, Boca Raton, FL

Zweifel JR, Lasker R (1976) Prehatch and posthatch growth of fishes: a general model. Fish Bull 74:601–621