Vol. 347: 275–284, 2007
doi: 10.3354/meps06984

MARINE ECOLOGY PROGRESS SERIES
Mar Ecol Prog Ser

Published October 11

OPEN ACCESS

Formulation and application of an efficient optimized biophysical model

David Brickman[1,2,*], Gudrun Marteinsdottir[2,3], Lorna Taylor[2]

[1]Dept. of Fisheries and Oceans, PO Box 1006, Dartmouth, Nova Scotia B2Y 4A2, Canada
[2]Marine Research Institute, Skulagata 4, 121 Reykjavik, Iceland
[3]Institute of Biology, University of Iceland, Sturlugata 7, 101 Reykjavik, Iceland

ABSTRACT: The formulation of an efficient optimized biophysical model is described, and the model is applied to the simulation of the climatological 0-group distribution of Icelandic cod *Gadus morhua* larvae. The method is based on representing the results from particle tracking as drift probability density functions describing the probability that particles released from a given spawning ground are found at a specific downstream grid location some time later. Spawning is considered to take place from 15 spawning grounds, and the model is used to determine 45 egg production model parameters as the solution of a bound constrained optimization problem that minimizes model-data misfits in abundance and age distributions. The problem is solved using a direct search minimization routine. Two cost functions are used. One penalizes misfits in the gridded abundance and age distributions (Model 1). The other directly penalizes the misfit in the spatial age gradient (Model 2). A simple age-based settlement module is tested to determine whether it improves the model fit. Results from Model 1 show a large error in the spatial age gradient. Model 2 achieves a 20-fold reduction in this error, with only a small degradation of the gridded abundance and age distributions. The settlement model does not improve the model fit. The results indicate that the addition of more processes to a model does not always improve model performance, while focusing on gradients in age instead of simple age distributions can lead to overall improved performance. The technique presented in the present paper allows quantitative evaluation of various model processes in a computationally efficient framework.

KEY WORDS: Optimized biophysical model · Probability density function · Icelandic cod

INTRODUCTION

Biophysical models (BPM) of fish larvae simulate the drift, development, growth, and mortality of released fish eggs (Heath & Gallego 1998, Brickman & Frank 2000, Hinrichsen et al. 2002). Typical components of such models are (1) a particle-tracking routine, which simulates egg/larval drift based on flow fields from a circulation model and information regarding spawning ground location(s), (2) an egg production model (EPM), which describes the space/time release of eggs, based on spawning stock data, and (3) a controlling program, which, using particle tracking, the EPM, and a mortality routine, computes the (time dependent) spatial distributions of eggs and larvae.

A characteristic of BPMs is that they contain a number of parameters that are poorly known, often bounded within a range of possible values. For example, the EPM simulation of egg and larval drift depends on peak spawning time, spawning duration, and the number of eggs spawned. Because flow fields vary in time, uncertainty in these parameters translates into uncertainty in the model-predicted spatial distributions of age and abundance. A way to solve for uncertain parameters is by finding values that minimize the mismatch between model predictions and observations. The model is thus said to be 'optimized'. This paper presents an efficient optimized BPM applied to the problem of simulating the climatological distribution of the pelagic 0-group survey data of Icelandic cod *Gadus morhua*.

*Email: brickmand@dfo-mpo.gc.ca

In a related paper, Brickman et al. (2007a) applied this model to data from 2002 and 2003, with the specific goal of understanding spawning stock and drift characteristics specific to these focus years of the METACOD project (a fifth framework research project granted by the European Commission). In addition to providing more details of the model and applying it to a different dataset, the present paper further explores the power of the optimization method to simulate specific characteristics of observed data and to assess the importance of biophysical processes in improving model performance. The climatological distribution of juvenile Icelandic cod is characterized by a negative spatial age gradient, with a fairly abrupt decrease in age near the northwest corner of Iceland (Begg & Marteinsdottir 2000, Marteinsdottir et al. 2000a, Brickman et al. 2007b). Brickman et al. (2007a) showed that the model had some ability to reproduce this feature, as it was represented in the 2002 data. Here, we show how the model cost function can be formulated to specifically focus on the gradient in age, leading to an improved model fit. The path toward settlement and beyond involves several biophysical processes that may or may not be explicitly modeled by the BPM. The Icelandic summer survey of the pelagic 0-group occurs at a time when it is possible that the settlement phase has begun. A simple age-based settlement module is added to the BPM, and the model is used to determine whether this leads to an improved fit to the data. In this way, we show how the optimization technique can be used to ascertain the degree to which a particular process is important to the modeling system.

OPTIMIZED BIOPHYSICAL MODEL

In this section we present the formulation of a computationally efficient BPM, and show how we can optimally determine model parameters by minimizing the mismatch between model predictions and data. The technique relies on particle-tracking results being stored to disk and on the results being converted to drift probability density functions (PDFs) as described in Brickman et al. (2007b). As such, we begin the section by reviewing the PDF approach. We then show how the PDF technique can be used in the formulation of a BPM and how, by minimizing the mismatch between model predictions and data, the problem of determining model parameters can be transformed into one of bound constrained optimization.

PDF representation of drift results

Particle tracking that includes a random component of drift (Lagrangian stochastic modeling) typically requires tens of thousands of particle releases in order to achieve stable drift statistics. These results can be characterized in a compact way by determining the probability that an ensemble of particles released from some region (i.e. a given spawning ground) at time $t0$ will be found in another region at $t1$. This is done by breaking the domain into a grid and counting the number of particles in each grid box at $t1$. The PDF for drift from a given spawning ground to a given grid cell is defined as:

$$P(i,j,t1;\text{SPG-}k,t0) = \frac{n(i,j,t1)}{N_k} \quad (1)$$

where SPG-k denotes the k-th spawning ground, (i, j) denotes grid cell, N_k is the total number of particles released from SPG-k, and $n(i, j, t1)$ is the number of particles found in grid cell (i, j) at time $t1$. This PDF is a type of 'transition probability matrix' for a Markov process. In general, as the velocity field is a function of time, $P(i, j, t1;$ SPG-$k, t0)$ is a function of time-of-release. These PDFs are calculated offline by a 'front end' routine. The effect is to translate a scatterplot of particle positions at a given time into a contour plot of drift probabilities computed on the chosen grid. Examples of drift PDFs around Iceland are presented in the 'Model application' section.

Model formulation

We are interested in the contribution from the batch of eggs released at time $t0$ from spawning ground 'k' (Fig. 1) to the abundance distribution at grid location (i, j) at $t1$ (ab$(i, j, t1)$). Assume the EPM release curve is Gaussian, characterized by parameters PS, σ, and F. Here, PS is the peak spawning time, σ is the standard deviation (SD) of the Gaussian curve (representing spawning duration), and F is the total number of eggs released at the spawning ground in question. If these parameters are known, then the number of eggs released at $t0$, $E(t0)$, is known.

The number of particles, or the concentration, at grid cell (i, j) at $t1$ is:

$$c(i,j,t1) = E(t0) \times P(i,j,t1;\text{SPG-}k,t0) \quad (2)$$

where P is known from particle tracking. This is in the absence of mortality. On their way to grid cell $(i, j, t1)$ these eggs hatch into larvae, which grow as they drift. The egg and larval phases are subject to (exponential) mortality. We take the egg mortality rate (M_E) to be constant in space and time and the larval mortality (rate) to be of the form (Houde 1997):

$$M_L(a) = \frac{b}{cL(a)^d} \quad (3)$$

where a is larval age and $L(a)$ is length (mm). The parameters M_E, b, c, and d are typically assumed, or can be estimated from data constraints. The total mortality at (i, j, $t1$) acting on the $E(t0)$ eggs released is:

$$M(\Delta t = t1 - t0) = \{e^{-M_E H}\}\{e^{-\int_H^{t1} M_L(a)da}\} \quad (4)$$

where H is the hatch time. Both H and M are computed during particle tracking and are carried as attributes for each of the particle tracks.

The contribution to ab(i, j, $t1$) is:

$$M_T(i,j,t0,t1) \times E(t0) \times P(i,j,t1; \text{SPG-}k,t0) \quad (5)$$

where $M_T(i, j, t0, t1)$ is the gridded mortality from the release at $t0$, calculated by the front end routine. In this particular example, M_T is just the average value of M for the particles found in grid cell (i, j) at $t1$. The total contribution from SPG-k is the sum over all release times. This computation is similar to a dot product, and can be efficiently coded. By summing the contributions from all spawning grounds (S) for all release times, an abundance and age distribution can be created for all grid cells.

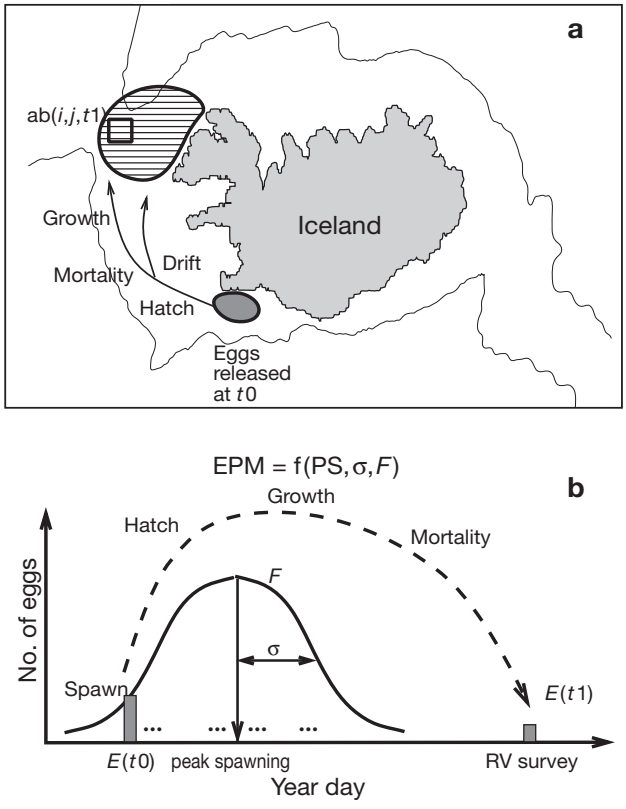

Fig. 1. Processes in the biophysical model showing (a) the progress of eggs released at $t0$ as they drift toward grid cell (black square) (i, j), arriving at $t1$ and (b) the egg production model release curve as a function of day-of-the-year. PS: peak spawning date; σ: the spread of the Gaussian release curve; F: the total number of eggs released. ab: abundances; $E(t0)$: eggs released at t0; $E(t1)$: larvae remaining from $E(t0)$ measured during a research vessel survey

In the above we have assumed that we know EPM(PS, σ, F)$_{spg = 1, S}$. In practice, this is not the case. We thus formulate the problem as follows: Given abundance and age data, determine the set of (PS, σ, F) that minimizes

$$\text{cost\_}fn = \sum_{\text{grid}} \{f([\text{ab(model)} - \text{ab(data)}]^2) + g([\text{age(model)} - \text{age(data)}]^2)\} \quad (6)$$

where the functions f and g on the right hand side will be called cost_ab and cost_age, respectively, and the summation is over all grid cells (i, j) for which there are data. For the Iceland problem under consideration, with 15 spawning grounds where EPM(PS, σ, F) are known within some bounds (see below), this is a 45 parameter problem in bound constrained form.

Solution details

Cost function

The cost function contains 2 components: one related to model-data discrepancies in abundance (cost_ab) and one related to model–data discrepancies in age (cost_age).

With respect to abundance, cost_ab was computed as ~$\sum\{\log[1 + \text{ab(model)}] - \log[1 + \text{ab(data)}]\}^2$. The main motivation for this was to de-emphasize data outliers as it was observed that the model could find unlikely optimal solutions that did a good job matching outliers but faired poorly on the majority of data.

The climatological age distribution of Icelandic 0-group cod (described below) is characterized by a spatial gradient in age concentrated near the northwest corner of the island (Begg & Marteinsdottir 2000). Two versions of cost_age were used. One was the simple squared difference between the model prediction and data (cost_age1), as in Brickman et al. (2007a). The other broke the area around Iceland into 4 boxes, from west to northeast and penalized the (sum of the) squared model-data differences in average age in each of the boxes. This version (cost_age2) directly penalizes errors in the spatial age gradient. Weights were used so that cost_ab and cost_age made roughly equal contributions to the overall cost function and so that the cost function was dimensionless.

Computational method

The problem is solved using a version of a direct search algorithm (see Kolda et al. 2003, and references therein) coded by D. Brickman. To illustrate the method, consider the simple case of a 2-parameter system (Fig. 2). Starting at a first guess, take 4 steps (2 for each para-

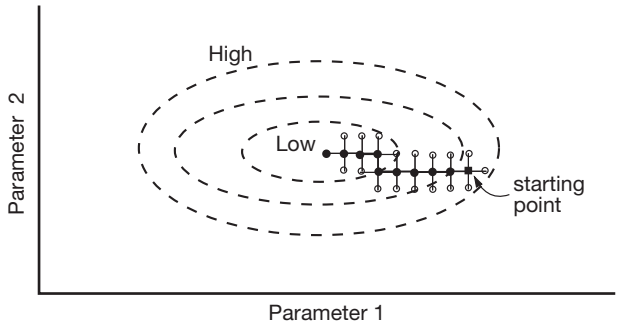

Fig. 2. Direct search minimization routine in 2 dimensions. Dashed lines: contours of the (unknown) cost function; small circles: the 4 trial steps per iteration. The path to the minimum follows the filled circles

meter while holding the other one fixed) and compute the cost function for each step. Take as the next starting point, the step that yielded the lowest cost function value. Repeat this procedure (4 steps plus 'decision') until arriving at the minimum value. (The method is often supplemented by using smaller steps as the minimum is approached.) It can be seen that this method is less efficient than a derivative-based method (e.g. steepest descent, Levenberg-Marquardt [Press et al. 1988]), but these considerations are less important now than they were historically. Also, the computations of numerical derivatives and subsequent matrix inversion have been shown to be the main sources of error and failure of derivative-based methods (Kolda et al. 2003). By comparison, the direct search method is easy to code and robust.

To get an idea of the computational efficiency of the BPM, define as 1 'model run' the computations: choose model parameters, compute the model prediction, calculate the cost function, and perform bookkeeping. On a Linux workstation (single AMD 3200 processor) 200 to 300 model runs s^{-1} are achieved. In terms of convergence of the direct search method, it was observed that about 50 to 60 direct search iterations (each consisting of ~90 model runs) were required to find the minimum, taking <30 s. In problems such as this (i.e. high-dimensional and non-linear), it is difficult to determine whether a global, versus local, minimum has been found. To find the global minimum, the direct search routine was run with 1000 different sets of initial parameter values, and the (final) parameter set associated with the absolute minimum cost function was recorded as the optimal set.

MODEL APPLICATION

The optimized BPM is applied to the problem of simulating the climatological 0-group survey data for Icelandic cod. We start by describing the survey data, then present data needed by the EPM. The circulation model developed for the METACOD project is presented, followed by representative output from the particle tracking algorithm designed to provide an overview of the circulation around Iceland. The settlement module is described.

0-group data

The data come from 29 yr of summer pelagic 0-group surveys (1970 to 1998), trawled at 20 to 50 m depth, and containing >150 stations yr^{-1}. To adjust for non-synopticity of the survey, abundance (number-per-km-towed) is adjusted to the mean survey date (Day 230 of the year) using a mortality rate of 0.03 d^{-1}, consistent with late larval mortality (Houde 1997). Length data (L, mm) are adjusted to the model mean survey date using a growth rate of 0.65 mm d^{-1} (Marteinsdottir et al. 2000a). Length is converted to total age (A, days) using the relation (Begg & Marteinsdottir 2000):

$$A = (L - 15.8723 + 2.5401 \times T)/(0.4256 + 0.0307 \times T) + 16 \quad (7)$$

where T is temperature (Celsius), and the offset (16) is the mean hatch time. For the range of lengths encountered (85% between 30 and 60 mm) and typical temperature range (6 to 10°C), using a constant temperature of 8°C gave a maximum error of <2 d (~2%), so that the simpler formula was considered acceptable. (Note that this relation does not constitute an age–temperature growth relation but, rather, expresses a length–temperature association for the older larvae collected by the survey.)

Climatological distributions of abundance and age were created by averaging the data on a 0.25 × 0.25° grid. The choice of grid size was based on preserving the inherent fine-scale structure of the data while avoiding graininess due to spatial oversampling. Furthermore, the data were decimated to include only those grid cells for which there were at least 6 tows in the 29 yr period. This captures the main features of the distribution while stopping the model from pursuing areas where confidence in the data may be questionable.

There are 3 characteristic features to the climatological 0-group distribution (Begg & Marteinsdottir 2000): (1) the majority of juveniles are found along the northern shelf, (2) there is an inshore/offshore gradient in abundance along the northern shelf with the majority of juveniles found inshore, and (3) a spatial gradient in age exists with age decreasing sharply in a clockwise direction around the northwest corner of Iceland.

The climatological inshore/offshore abundance gradient along the north shelf—defined as the ratio of the average abundance per kilometer inside versus outside the 100 m isobath (Fig. 3b)—is 5.1. The spatial age

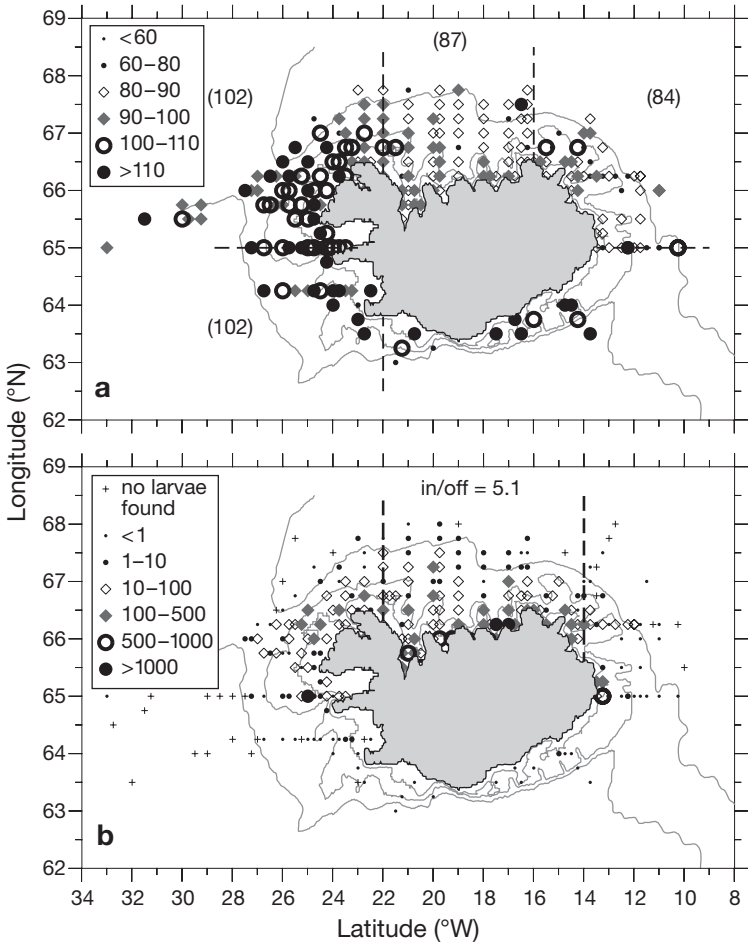

Fig. 3. Gridded climatological spatial (a) age and (b) abundance distributions derived from research vessel survey data, decimated to ≥6 tows cell^{-1}. Areas used to compute bulk measures of spatial age and inshore(in)/offshore(off) abundance gradients are delineated by dashed lines in the 2 panels, respectively, as well as the computed values. The 100, 200, and 600 m isobaths are represented in this and subsequent figures. Units are days for age, and number-per-km-towed for abundance

gradient—defined as the average age in boxes from SW to NE (Fig. 3a)—is 102, 102, 87, and 84. These boxes are used in cost_age2. The length-frequency distribution (not shown) is (roughly) Gaussian, with a peak at ~50 mm and a SD of ~10 mm. Less than 1% of pelagic juveniles caught historically were larger than 7 cm.

Spawning grounds and egg production estimates

The main spawning grounds for Icelandic cod are in the southwest, but some spawning is thought to occur in most of Iceland's bays and fjords. Thirteen locations to release particles were chosen, based on logbook and spawning survey data (Marteinsdottir & Bjornsson 1999, Marteinsdottir et al. 2000a,b). From genetic (Jonsdottir et al. 2002), otolith (Petursdottir et al. 2005), length-frequency, and abundance considerations, the main spawning area was subdivided into 3, making a total of 15 separate spawning grounds (Fig. 4), referred to as SPG-1(a, b, and c) to SPG-13.

Also annotated on Fig. 4 (and Table 1) are estimates of the fraction of total eggs released for each spawning ground and the peak spawning time, based on data presented in Brickman et al. (2007a) and other local knowledge. We see that the main spawning grounds (SPG-1, ca. –22°W, 63.5°N) provide 30 to 50% of the total eggs spawned, with significant contributions from SPG-12 and SPG-3. The total contribution from all northern (and eastern) spawning grounds is estimated to be <5%. Total eggs spawned per year is computed to be 1.5×10^{14}. Peak spawning (PS) day of the year is generally accepted to increase in a clockwise direction starting from ~Day 105 (SPG-1, well established) to about Day 140 to 150 for spawning grounds along the north coast (more uncertain) (Begg & Marteinsdottir 2000). The peak spawning time for other south coast spawning grounds is considered to be later than that of SPG-1 (Fig. 4). Spawning duration (the SD of the Gaussian curve) is about 5 to 10 d for southern spawning grounds, and likely longer for northern ones. Uncertainties in these estimates provide the ranges used to constrain the optimized BPM (Table 1).

Circulation model and drift algorithm

As part of the METACOD project a circulation model was developed that covered the Atlantic and Arctic Oceans, with a focus on Icelandic waters. For details of this model, see Logemann & Harms (2006) and Brickman et al. (2007b). The model domain contains a hierarchy of nested subdomains, being reduced to 1.2 km resolution in Icelandic waters. To estimate the climatological circulation, the model is forced with the wind stress from the Ocean Model Intercomparison Project dataset (Röske 2006). This dataset, based on re-analysis of 15 yr of ECMWF atmospheric data, describes a 360 d cyclic, stationary, climatological year that includes the passage of storms (i.e. it is a 'storm climatology'). One tidal component (M2) is included. The result is a time-varying climatological circulation, with output as daily averaged fields for use in the particle-tracking routine.

Particles are advected by space and time-interpolated currents, plus they have an additional (horizontal)

Table 1. Egg production model input parameter ranges and output values from the optimized biophysical model (Model 2). SPG: spawning ground; PS: peak spawning time (day of the year); σ: standard deviation of Gaussian egg release curve (d); egg-frac.: fraction of total eggs released. The summation of egg-frac must total 1, so the model can achieve values outside of the input range

SPG	Input parameters			Output parameters		
	PS range	σ range	Egg-frac. range	PS	σ	Egg-frac.
1a	90–120	5–10	0.05–0.15	90	5	0.071
1b	90–120	5–10	0.05–0.15	90	10	0.107
1c	90–120	5–10	0.05–0.15	110	10	0.107
2	115–140	5–10	0.05–0.15	140	10	0.036
3	120–150	5–15	0.1–0.4	145	5	0.286
4	140–150	5–15	0.0–0.15	150	10	0.107
5	140–150	5–15	0.0–0.05	145	10	0.036
6	150–160	5–15	0.0–0.05	150	10	0.036
7	150–160	5–15	0.0–0.05	150	10	0.036
8	150–160	5–15	0.0–0.05	150	10	0.036
9	150–160	5–15	0.0–0.05	155	5	0.000
10	150–160	5–15	0.0–0.05	155	5	0.036
11	115–115	10–10	0.0–0.02	115	10	0.000
12	100–130	5–10	0.05–0.3	100	5	0.036
13	105–125	5–10	0.05–0.15	125	10	0.071

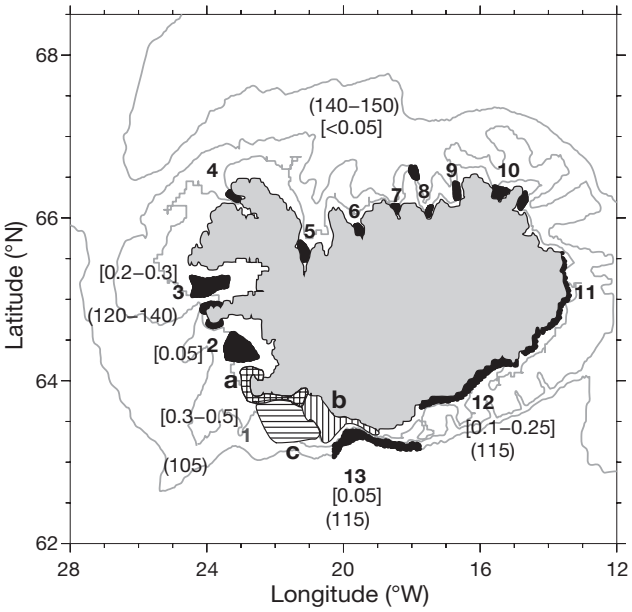

Fig. 4. The 15 spawning grounds and estimates of egg production model parameters. Square brackets enclose fraction of eggs spawned. For SPG-1a, b, and c this is [0.1–0.4], [0.05–0.15], and [0.05 to 0.20], respectively. The total for SPG-4 to 11 is <0.05. Round brackets enclose the day(s)-of-the-year of peak spawning

stochastic component to represent the effects of turbulence. The latter depends on the local (time dependent) diffusivity derived from the circulation model and is enacted in the form of a 'random displacement model' (Rodean 1996, Brickman & Smith 2002). A total of 30 000 particles were released from the 15 spawning grounds, distributed roughly in proportion to the areas of the individual spawning grounds. The method of Brickman & Smith (2002) was used to determine that this number of particles gave satisfactory drift statistics. The vertical component of turbulence was not included, because, as shown by Brickman & Smith (2002), this can lead to erroneous particle tracks if the circulation model turbulence field is not sufficiently smooth.

The circulation model flow fields are time dependent, which means that drift patterns depend on the day of the year of release. For practical reasons, particles were released at 5 d intervals starting on Day 80 and ending on Day 175, resulting in a form of discretization of the EPM. This interval was chosen to safely bracket the expected egg release days from the various spawning grounds in Icelandic waters (Marteinsdottir & Bjornsson 1999). Each release was tracked until the climatological mean survey date (Day 230 of the year).

Icelandic cod eggs are released near the bottom and rise to the surface after spawning. Field data for gadoid larvae indicate that they exhibit an ontological increase in mean depth that increases as they age (Werner et al. 1993, Brickman et al. 2001). Unfortunately, there are not sufficient egg and larval data to deduce such a relationship for Icelandic cod. Brickman et al. (2007a), using the optimized BPM to study the 2002 and 2003 0-group data, considered algorithms that started particles at 5, 10, and 15 m depth, respectively, until they hatched, after which time they sank at a rate that put them at about 35 m, the mean trawl depth, after 100 larval days. They found that the 5 m initial depth algorithm gave the best fit to the data. We use that result in the present paper.

A number of particle attributes were recorded along the drift tracks. Of these, the hatch time (H) was determined using the temperature-dependent relation $H = 46.1e^{-0.17T}$ from Pepin et al. (1997). As there is no temperature-dependent growth relation available for Icelandic cod larvae, a linear growth relationship, consistent with the data of Marteinsdottir et al. (2000a), was used in which larvae grew at a rate of 0.35 mm d^{-1} for the first 30 d after hatching and thereafter at 0.65 mm d^{-1}. Thus, the total mortality along the drift path (M, Eq. 4) is an analytic function of age, hatch time, and the larval mortality constants b, c, and d. We found that b = 2.3, c = 0.7, and d = 1.35, plus M_E = 0.23 d^{-1}, gave a reasonable number of survivors compared to the data. These attributes, plus others (for example, the mean larval drift temperature) are calculated in grid coordinates by the front end routine.

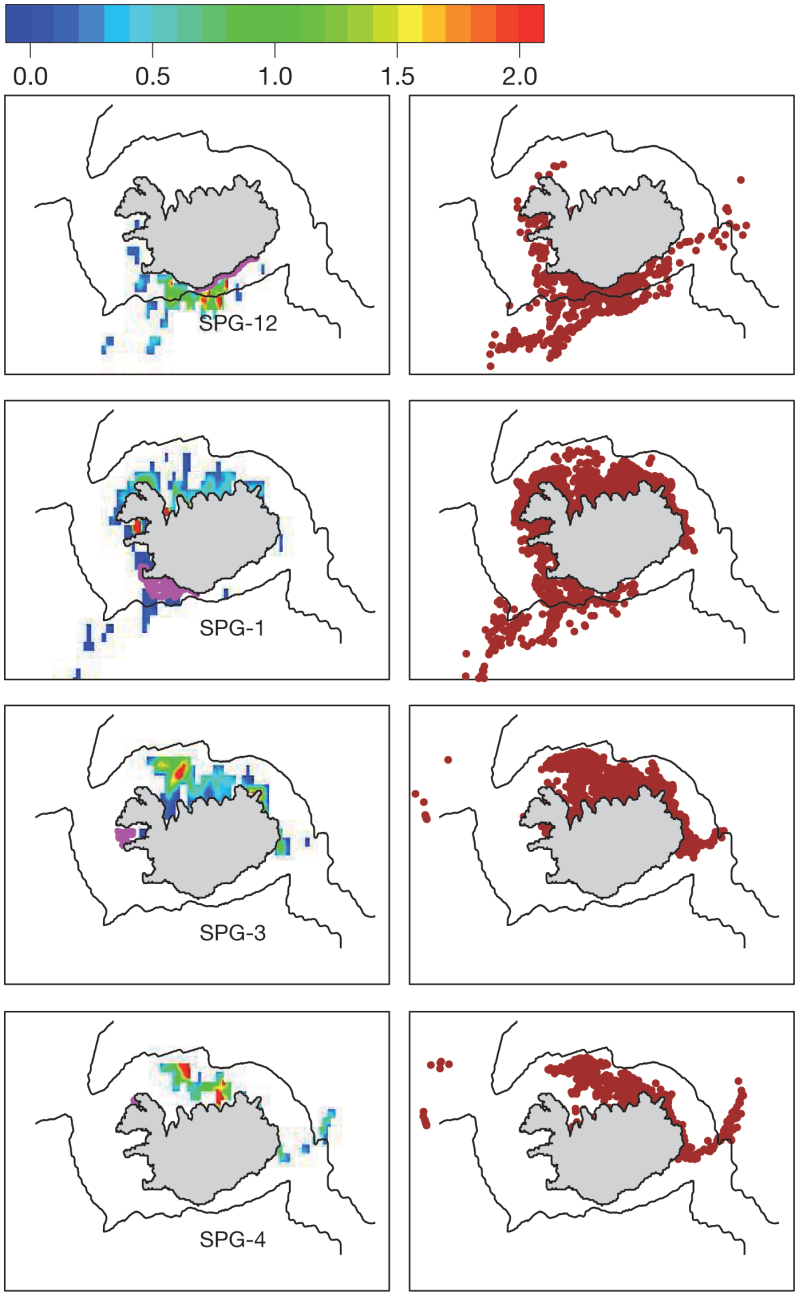

Fig. 5. Example of drift probability density functions for drift in the climatological circulation. Left panels: drift probability in percent. Spawning ground (SPG) is shown in magenta. Right panels: raw final particle positions (at Day 230). Release time is Day 110 of the year. The 100 and 200 m isobaths are omitted for clarity

Representative drift PDF plots for the climatological circulation (release time = Day 110 of the year), with spawning ground of release proceeding in a clockwise fashion from top to bottom are presented (Fig. 5). The general clockwise circulation around Iceland is apparent. Drift from SPG-12 is partly downstream, partly to the southwest (along the Reykjanes Ridge), with significant retention near the release site. For later release times, this spawning ground exhibits increasing drift towards the southeast (Brickman et al. 2007b). SPG-1, -3, and -4 mainly supply the north coast (SPG-4 is an in-fjord spawning ground in the northwest peninsula). SPG-3 and -4 show some drift toward Greenland for this release time.

Settlement module

Little is known about the settlement phase of juvenile Icelandic cod. As mentioned above, virtually no juveniles larger than 7 cm are found by the August pelagic trawl survey. Based on the length–age equation and a 16 d hatch time, a 7 cm juvenile would be spawned on roughly Day 105, about the peak spawning time for the main spawning grounds. By-catch data from the offshore demersal trawl shrimp survey in July and August indicates the presence of 5 to 7 cm cod (peak at 6 cm), although only about 250 data points have been recorded historically (Marine Research Institute [MRI], Iceland unpubl. data). The October adult survey finds settled cod with an average size of 8 to 9 cm (MRI unpubl. data).

The above indicates that the summer 0-group survey is likely sampling, in some regions, a population part of which is in the process of entering the demersal phase. To allow for this possibility, a simple settlement module will be included that settles a fraction of juveniles based on a minimum size (L_s) and a time scale (τ), i.e. fraction settled ~$e^{(L-L_s)/\tau}$. This introduces 2 more parameters into the model (L_s, τ). We will investigate whether the addition of this 'age–settlement' model improves the fit to the data. Note that the larval growth relationship makes length and age interchangeable.

MODEL RESULTS

The model that directly penalized discrepancies in the bulk age gradient (Model 2: cost_fn = cost_ab + cost_age2, no-settlement model; Fig. 6) did a reasonable job with the abundance distribution and the north coast inshore/offshore abundance gradient (cf. Fig. 3),

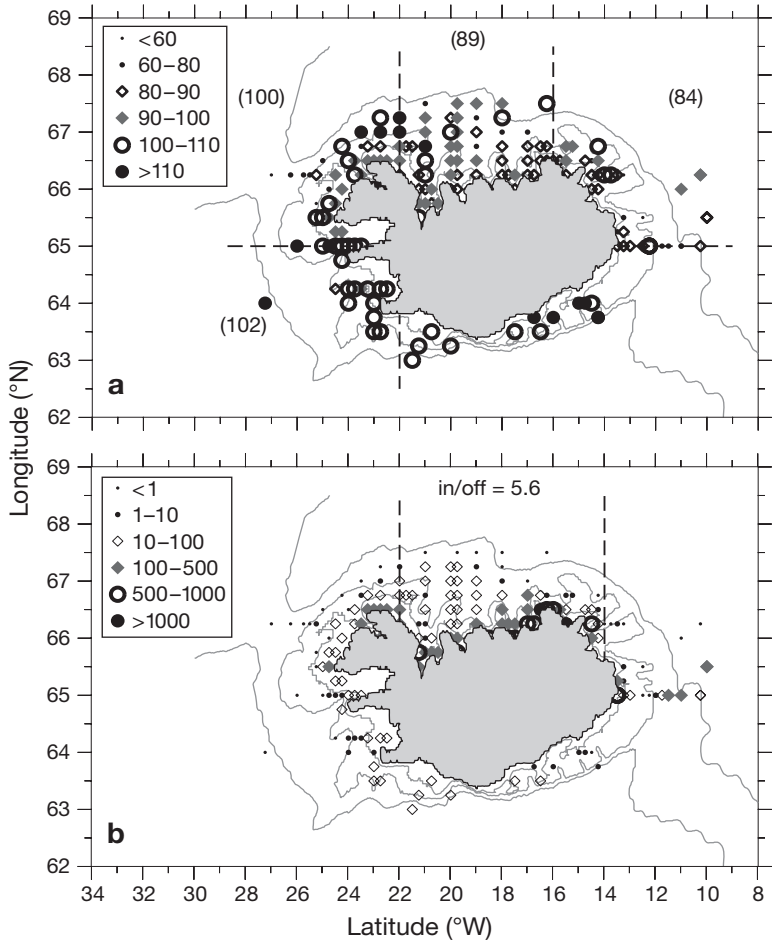

Fig. 6. (a) Spatial age (parentheses: spatial age gradient) and (b) abundance distributions, for Model 2 (cost_*f*n = cost_ab + cost_age2), no-settlement module. Bulk measures of spatial age and inshore(in)/offshore(off) abundance gradients are shown in the 2 panels, respectively

and an excellent job in reproducing the climatological age distribution (Model 2 = [102, 100, 89, 84]; data = [102, 102, 87, 84]). The basic model (Model 1: cost_*f*n = cost_ab + cost_age1, no-settlement model) did a reasonable job with the abundance distribution and the north coast inshore/offshore abundance gradient (in/off = 4.7 versus 5.1 for climatology), but was unable to reproduce the spatial age gradient (Model 1 = [107, 91, 94, 88]). (The Model 1 result was similar enough to Model 2 to not warrant a figure.)

To assess the difference between these 2 results, all components of the cost function (cost_ab, cost_age1, cost_age2) were computed for both cases. We found that Model 2 results in an almost 20-fold reduction in cost_age2, while only incurring about 7% increases in cost_ab and cost_age1. In other words, focusing specifically on the age gradient metric—of special interest to the 0-group distribution problem—leads to a large reduction in this error, with only a small effect on the other distributional measures.

To illustrate where errors occur spatially, a cost function was computed for Model 2 as the square root of cost_ab + cost_age1 on a grid-by-grid basis (Fig. 7). Major areas of error occur in waters deeper than ~150 m along the southwest and west coasts, and in deeper waters along the north and northeast coasts. From visual inspection of model results versus data (Fig. 6 versus Fig. 3), these errors occur in regions where the model predicts zero abundance, indicating a possible underdispersion of particles relative to the climatological data. We return to this point in the 'Discussion' section. Analysis of model output reveals 70 'presence/absence' errors out of 270 data grid points, of which 65 were cases in which larvae are present in the data, but the model predicts zero abundance (and thus zero age). In these 65 instances, the average abundance is low (~10 larvae km^{-1} towed), but the average age is high (96 d). Thus, in real terms, most of the model error is due to age discrepancies in regions of presence/absence errors.

For both Model 1 and Model 2, the addition of the age–settlement module made little visual difference to the results (not shown), and actually resulted in slight increases (~0.5%) in the cost functions. This is not impossible as the module essentially forces some settlement (equivalent to a reduction in predicted 0-group) when no settlement may be a better answer. This shows that the addition of an unmodeled process may actually be detrimental to overall model performance.

DISCUSSION AND CONCLUSIONS

In this paper, a new formulation for a BPM was presented. The computationally efficient method is based on converting the results from particle tracking into drift probability density functions and other attributes calculated in a grid-based coordinate system. In this way many model calculations are done offline, becoming sparse grid-based input data for the main program. It was shown how poorly known model parameters can be determined as the solution of a bound constrained optimization problem that minimizes model-data discrepancies.

The method, herein applied to data from 1 time frame, is readily adapted to handle multiple data frames (e.g. egg/larval surveys) with little loss in speed. As well, while this method is presented as an alternate formulation of a BPM, the gridded attributes

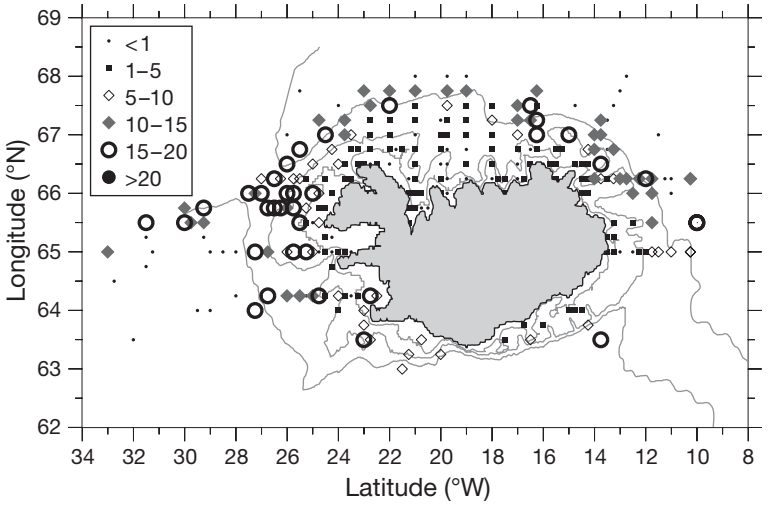

Fig. 7. Distribution of error for Model 2 expressed as the square root of the cost function, computed on a grid-by-grid basis. The values are dimensionless

computed in this case give identical results to an individual-based model (IBM; Heath & Gallego 1998, Hinrichsen et al. 2002) using the same parameters. This can be seen by computing, for an IBM, the contribution to ab(i, j, $t1$) from eggs released at $t0$ from SPG-k:

$$\mathrm{ab}(i,j,t1) = \sum_{\mathrm{p}=1}^{n} E_{\mathrm{p}} \mathrm{M}_{\mathrm{p}} \quad (8)$$

where $E_{\mathrm{p}} = E(t0)/N_k$ is the number of eggs assigned to each particle released from SPG-k at $t0$, n is the number of particles found in grid cell (i, j), and M_{p} is the mortality along the drift path as defined in Eq. (4). Rearranging the above, using the fact that the mortality used in the PDF method $M_{\mathrm{T}}(i, j, t0, t1) = (\sum M_{\mathrm{p}})/n$ and recalling the definition of P, we get:

$$\mathrm{ab}(i,j,t1) = M_{\mathrm{T}} \times E(t0) \times \left\{ \frac{n}{N_k} \right\} = M_{\mathrm{T}} \times E(t0) \times P \quad (9)$$

which is the same as Eq. (5).

Note that this method can be adapted to all formulations that can be run offline. This includes the possibility of optimizing the mortality parameters, as well as the situation where the larvae are embedded in a nutrient/phytoplankton/zooplankton model and their interaction with the prey (and/or predator) field affects their behavior. The technique remains efficient provided that parameters affecting propagule behavior are not part of the set to be optimally determined. An optimization technique for this latter case remains a key problem in biophysical modeling.

The mortality function was taken to be inversely proportional to larval length. For Icelandic cod, no specific temperature-dependent growth model exists, so larval length was taken to be a linear function of age (Begg & Marteinsdottir 2000). If such a relationship did exist, it would be simple to incorporate it into the mortality attribute. However, it is questionable what difference this would make, because, as shown by Brickman et al. (2001) for haddock on the Scotian Shelf of eastern Canada, the incorporation of an age–temperature growth model into the mortality function has a low-order effect. This is due to the form of the mortality function, the temperature field experienced by the larvae, and the fact that a length–age relation captures most of the growth variability.

The model was applied to the problem of simulating the Icelandic cod climatological 0-group survey data. Spawning was taken to occur in 15 spawning grounds, resulting in 45 (egg production model) parameters that were known within bounded estimates. Two different cost functions were used, measuring model–data discrepancies in abundance and age distributions. One penalized the model age prediction on a grid cell by grid cell basis (Model 1). The other (Model 2) directly penalized differences in the spatial age gradient, a feature of the data that is of particular interest for Icelandic researchers.

We found that both models did a reasonable job in simulating the observed inshore/offshore abundance gradient along the north coast. However, the Model 1 solution was unable to reproduce the observed juvenile age gradient, while Model 2 (Fig. 6, Table 1) did much better in this regard. Evaluation of the components of the 2 cost functions showed that the improved fit to the age gradient using Model 2 was only slightly offset by poorer fits to the gridded abundance and age distributions. This observation was corroborated by running a version of the model that penalized only the age gradient (not shown), which resulted in virtually zero misfit in this bulk metric, with little further degradation in the gridded abundance and age errors. The relative insensitivity of the model to these latter 2 metrics can be attributed to the fact that large errors occurred in the offshelf regions (Fig. 7), so that subtle, but important, changes onshelf do not greatly affect total gridded error measures. In general, these large errors were found to be due to systematic presence/absence discrepancies, where the model predicted zero abundance (and age), while the observations indicated older larvae (albeit in low abundance).

As pointed out by Brickman et al. (2007a,b), the particle-tracking algorithm seems to retain particles closer to shore than the climatological data indicate. Whether or not this is due to the use of climatological flow fields, which would be expected to underestimate variability, or due to a deficiency in the particle-tracking algorithm is difficult to determine. Inspection of the annual 0-group distributions (Fig. 3 of Begg & Marteinsdottir

2000) indicates that, while there is interannual variability (especially with respect to the drift toward Greenland), generally the model underpredicts abundance in areas where the survey data are not characterized by episodic events. As well, this underdispersion is also evident in the Brickman et al. (2007a) results, although not to as great an extent. This favors the possibility that this apparent underdispersion may be due to some sort of dispersive larval behavior, such as horizontal swimming motion, which increases as they age. It is possible to add this in a number of ways to the particle-tracking algorithm, and then use the optimized BPM to determine what algorithm produces the best result. This is similar to what was done in Brickman et al. (2007a) to determine the best ontogenetic vertical migration algorithm. In any case, the reason for this systematic error is currently unresolved.

A simple age-based settlement module was added to the BPM to see if it improved the model fit. We found that despite adding 2 more parameters to the BPM, the overall solution actually slightly deteriorated (<1%). This was attributed to the fact that this module effectively forces extra pelagic juvenile mortality, which can lead to a worse fit. The incorporation of more processes (e.g. biology) into a model is usually considered to be an improvement. The modeling technique presented in this paper allows a quantitative evaluation of this procedure. The result points out that additions of more biophysics to a model may not be beneficial, and thus must be approached with caution. It does not, however, preclude that a different settlement module could improve model performance.

The model finds the optimal set of egg production model parameters (Table 1) within the confines of the imperfections of its inputs and its construct, and the results should always be assessed with these limitations in mind. For example, the fraction of eggs contributed from SPG-3 (about 29%) is higher than expected, as is the total from the northern spawning grounds. Whether this is real or due to imperfections in the particle-tracking routine, the simple EPM or the mortality function, for example, is difficult to determine without more biological and physical data. The problems of model limitations and poorly represented or unmodeled processes are inherent to all biophysical models. The technique presented in this paper allows quantitative evaluation of various model processes in a computationally efficient framework.

Acknowledgements. This work was partly funded by the EU (METACOD—Project No. Q5RS-2001-00953), the Icelandic Research Council, and the Icelandic Ministry of Fisheries. The authors thank anonymous reviewers and, particularly, the editorial board for suggestions that significantly improved this manuscript.

Editorial responsibility: Alejandro Gallego (Contributing Editor), Aberdeen, UK

LITERATURE CITED

Begg GA, Marteinsdottir G (2000) Spawning origins of pelagic juvenile cod *Gadus morhua* inferred from spatially explicit age distributions: potential influences on year-class strength and recruitment. Mar Ecol Prog Ser 202:193–217

Brickman D, Frank KT (2000) Modelling the dispersal and mortality of Browns Bank egg and larval haddock *Melanogrammus aeglefinus*. Can J Fish Aquat Sci 57:2519–2535

Brickman D, Smith PC (2002) Lagrangian stochastic modelling in coastal oceanography. J Atmos Ocean Technol 19: 83–99

Brickman D, Shackell NL, Frank KT (2001) Modelling the retention and survival of Browns Bank haddock larvae using an early life stage model. Fish Oceanogr 10: 284–296

Brickman D, Taylor L, Gudmundsdottir A, Marteinsdottir G (2007a) Optimized biophysical model for Icelandic cod larvae. Fish Oceanogr 16:448–458

Brickman D, Marteinsdottir G, Logemann K, Harms I (2007b) Drift probabilities for Icelandic cod larvae. ICES J Mar Sci 64:49–59

Heath M, Gallego A (1998) Bio-physical modelling of the early life stages of haddock, *Melanogrammus aeglefinus*, in the North Sea. Fish Oceanogr 7:110–125

Hinrichsen HH, Mollmann C, Voss R, Koster FW, Kornilovs G (2002) Biophysical modeling of larval Baltic cod (*Gadus morhua*) growth and survival. Can J Fish Aquat Sci 59: 1858–1873

Houde ED (1997) Patterns and trends in larval-stage growth and mortality of teleost fish. J Fish Biol 51(Suppl A):52–83

Jonsdottir ODB, Imsland AK, Danielsdottir AK, Marteinsdottir G (2002) Genetic heterogeneity and growth properties of different genotypes of Atlantic cod (*Gadus morhua* L.) at two spawning sites off south Iceland. Fish Res 55:37–47

Kolda TG, Lewis RM, Torczon V (2003) Optimization by direct search: new perspectives on some classical and modern methods. SIAM (Soc Ind Appl Math) Rev 45:385–482

Logemann K, Harms I (2006) High resolution modelling of the North Icelandic Irminger current (NIIC). Ocean Sci Discuss 3:1149–1189

Marteinsdottir G, Bjornsson H (1999) Time and duration of spawning of cod in Icelandic waters. ICES CM 1999/Y:34

Marteinsdottir G, Gunnarsson B, Suthers IM (2000a) Spatial variation in hatch date distributions and origins of pelagic juvenile cod in Icelandic waters. ICES J Mar Sci 57: 1182–1195

Marteinsdottir G, Gudmundsdottir A, Thorsteinsson V, Stefánson G (2000b) Spatial variation in abundance, size composition and viable egg production of spawning cod (*Gadus morhua* L.) in Icelandic waters. ICES J Mar Sci 56:824–830

Pepin P, Orr DC, Anderson JT (1997) Time to hatch and larval size in relation to temperature and egg size in Atlantic cod (*Gadus morhua*). Can J Fish Aquat Sci 54(Suppl 1):2–10

Petursdottir G, Begg GA, Marteinsdottir G (2005) Discrimination between Icelandic cod populations from adjacent spawning areas based on otolith growth and shape. Fish Res 80:182–189

Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1988) Numerical recipes in C: the art of scientific computing. Cambridge University Press, Cambridge

Rodean HC (1996) Stochastic Lagrangian models of turbulent diffusion. In: Meteorological monographs, No. 48. American Meteorological Society, Boston, MA, p 1–84

Röske F (2006) A global heat and freshwater forcing dataset for ocean modelling. Ocean Model 11:235–297

Werner FE, Page FE, Lynch DR, Loder JW, Lough RG, Perry RI, Greenberg DA, Sinclair MM (1993) Influences of mean advection and simple behavior on the distribution of cod and haddock early life stages on Georges Bank. Fish Oceanogr 2: 43–64

Submitted: July 4, 2006; Accepted: July 26, 2007
Proofs received from author(s): September 21, 2007