Fisheries Research
June 2013, Volume 143, Pages 57–66
http://dx.doi.org/10.1016/j.fishres.2013.01.008

Selection and validation of a complex fishery model using an uncertainty hierarchy

Sigrid Lehuta[a,] *, Pierre Petitgas[a], Stéphanie Mahévas[a], Martin Huret[b], Youen Vermard[c], Andrés Uriarte[d], Nicholas R. Record[e]

[a] IFREMER, rue de l'île d'Yeu BP 21105, 44311 Nantes Cedex 03, France
[b] IFREMER, Centre de Brest, B.P. 70, 29280 Plouzané, France
[c] IFREMER, 150 quai Gambetta, B.P. 699, 62321 Boulogne/Mer Cedex, France
[d] AZTI tecnalia, Herrera Kaia-Portualdea z/g E20110, Guipuzcoa, Spain
[e] Earth & Oceanographic Science Department, Bowdoin College, Brunswick, ME 04011, USA

*: Corresponding author : Sigrid Lehuta, Tel.: +33 3 21 99 50 61 ; email address : sigridlehuta@gmail.com

Abstract:

Assessing the validity of a model is essential for its credibility especially when the model is used as decision making tool. Complex dynamic fishery models are recommended to investigate the functioning of fisheries and to assess the impact of management strategies, particularly spatial fishing regulations. However, their use is limited due to the difficulty and computational cost of parameterizing and gaining confidence, particularly for parameter rich models. These difficulties are compounded by uncertainty regarding parameter values, many of which are often taken from literature or estimated indirectly. Here we propose a methodology to improve confidence and understanding in the model, easily transferable to any complex model. The approach combines sensitivity analysis, scalability of parameters, optimization procedures, and model skill assessment in order to parameterize, validate and achieve the most plausible formulation of a model given the available knowledge while reducing the computational load. The methodology relies on five steps: (1) sensitivity analysis, (2) classification of parameters into a hierarchy according to their sensitivity and the nature of their uncertainty, (3) building of alternative formulations, (4) calibration and (5) skill evaluation. The approach is illustrated here by reviewing the parameterization of the ISIS-Fish model of the anchovy fishery in the Bay of Biscay. By using this approach, it is possible to make a thorough assessment of lacking information (e.g. accessibility to fishing and adult mortality) and to identify the strengths and weaknesses of the model in the context of different hypotheses. When applied to the ISIS-Fish model, the results suggest higher egg and adult mortality than formerly estimated, as well as new estimates for the migration towards spawning areas. They show the reliability of the model in terms of correlations with observations and the need for further efforts to model purse seiner catches. The methodology proved to be a cost-efficient tool for objectively assessing applied model validity in cases where parameter values are a mix of literature, expert opinion and calibration.

Highlights

► A five-step methodology is proposed to parameterize and validate a complex fishery model. ► It combines scalability of parameters, optimization procedure and model skill assessment. ► It is used to review the ISIS-Fish model of the anchovy fishery in the Bay of Biscay. ► It modifies our understanding of anchovy dynamics and evidences model strengths and weaknesses.

Keywords: Bay of Biscay anchovy ; Dynamic fishery model ; Model selection ; Uncertainty ; Validation

1. Introduction

Assessing the validity of a model is essential for its credibility especially when the model is used as decision making tool (Stow et al., 2009). Although crucial, validation is a non-trivial exercise, which, when attempted, is often qualitative and largely subjective (Allen et al., 2007). In fishery science, models have emerged as the most widely used tools to support management decisions. However, in fishery models, the modelling process is complicated by (1) substantial observational uncertainty and (2) limited data. Observations are often indirectly obtained, and there is confusion between measurement error and intrinsic variability, also called process error, which arises from unpredictable natural variability (Charles, 1998; Punt & Donovan, 2007). Short time series and limited data hardly allow for full understanding of all the mechanisms involved. These limitations can significantly confound validation efforts.

In most cases, parameterization is achieved using optimization procedures, which tune parameters and ensure the best fit of the model outputs to the available historical data (Drouineau et al., 2010; Pech et al., 2001). However, such procedures have a number of pitfalls. In many cases, the optimization procedure requires the use of every available time series, and an independent validation cannot be carried out, often leading to confusion between validation and calibration. With the evolution of management strategies, which now include spatial regulations and/or combination of management rules, models have become more complex, for example in terms of time resolution, explicit representation of space, or number of processes described (Pelletier and Mahévas, 2005). The number of parameters, computation time, difficulty to build a likelihood function, or likely confounding effects make simultaneous optimization of all parameters often unfeasible, and parameters must be assessed independently, using available knowledge and integrating this information in the model (Fulton et al., 2004; Kraus et al., 2009; Lehuta et al., 2010; Travers et al., 2006). In these cases, consistency between the parameters of the model has to be verified through rigorous validation.

Given the large uncertainty in parameter values (either due to measurement error when parameters are directly measured, estimation error due to incorrect estimation method and imperfect data, intrinsic variability, or multiple sources of information), but also in model structure (for instance from various interpretations of a mechanism), it is likely that the modeling process will result in a multitude of equally plausible formulations (here used to refer to specification and parameter estimates) of the model, as opposed to a single, unambiguously optimal formulation. Moreover, the optimal formulation and the validity of the model depend on the criteria used to assess them, and those criteria are thus crucial (Allen et al., 2007). To accommodate the fact that there cannot be an absolute, nor objective or single test of validity of a model (Sterman, 1984), several authors have promoted the use of specific metrics that reflect various aspects of model skill (correlation, efficiency, accuracy) and have provided more quantitative elements for model evaluation (Jolliff et al., 2009). Additionally, multivariate validation using several output variables at different scales is recommended to demonstrate the appropriateness of model structure (Allen and Somerfield, 2009; Cury et al., 2008; Wiegand et al., 2004).

Given the broad uncertainty and the likely ambiguity of model optimality, it is important to consider alternative model formulations. Consequently, selecting and gaining confidence in the model parameterization are inseparable processes, and they require an iterative process of formulation, parameterization and validation (Wiegand et al., 2003). We propose a methodology that relies on an uncertainty hierarchy to build alternative parameterizations of a model. The methodology then uses several validation criteria and several historical variables to compare their validity. This approach allows us to efficiently and thoroughly explore and assess the space of alternative model formulations and select the most appropriate for a given model purpose.

We applied this approach to the ISIS-Fish model of the anchovy fishery in the Bay of Biscay (Pelletier et al, 2009; Lehuta et al. 2010). It is an example of a deterministic complex fishery model that was developed to evaluate spatial management strategies for anchovy (Lehuta et al. 2010). It was not possible to parameterize the model through an integrated optimization procedure because of the high number of parameters (about 700), the discrete nature of some uncertainty domains and the long duration of simulations (10 minutes for 10 years). We propose to review the modelling of this fishery using our methodology to improve model adequacy and credibility.

The Methods section describes the generic approach proposed—that is, to organize parameters in a hierarchy and to rationalize the building of alternative formulations. The Materials section presents the model under study. The Results section presents the application of this methodology to the ISIS-Fish model of the Anchovy fishery in the Bay of Biscay. We detail the classification of parameters in categories; investigate the quality of the calibration, compare the validity of the formulations and assess the influence of time varying parameters. We then discuss the approach and its limits and provide recommendations for applying it more generally.

2. Methods

The methodology is designed to guide the modeler in the determination of parameter values, in building multiple formulations of the model, and in evaluating and comparing their respective skill.

2.1. Parameter assessment and alternative formulations

Building a parameterization begins with a review of knowledge on modeled processes and confidence in the model‘s parameters. Parameters are classified within a hierarchy according to the nature of their uncertainty and their effect on model sensitivity. The first step is a sensitivity analysis, to identify the parameters that significantly influence model results. It is typically a small proportion of the parameters, as evidenced by Lehuta et al. (2010) and stated by Saltelli et al. (2000), which considerably reduces the number of parameters to investigate. Sensitivity analyses are carried out by changing values of input parameters according to a simulation design and analysing the impact on model output using statistical methods. Many sensitivity analysis methods exist and the appropriate method must be selected based on the objectives of the analysis, the costs of model runs, the number of parameters and the nature of their uncertainty domains (see Saltelli et al., 2000 for a review of sensitivity analysis methods). The second step is to characterize the uncertainty range for each sensitive parameter. Some parameters have well-known values, unambiguously described in literature. Others parameters are accompanied by various forms of uncertainty, which we divided into three categories. The first encompasses parameters linked to processes that vary in time, and the mechanisms underlying the variability are unknown and thus not modelled. Since our model is deterministic, inter-annual variability cannot be taken into account in prediction through random parameter values and it must be decided whether a fixed value is appropriate or if scenarios must be tested. The second category describes parameters inconsistently estimated, for which several values are available in literature. This happens when various procedures or measuring tools have led to inconsistent estimations or when values obtained in other regions are used to estimate an unknown parameter. The third includes parameters inaccurately estimated or unknown because the assessment methods or data are lacking (Figure 1).

These steps taken together produce five categories of parameters, and each category requires a different treatment. Category (i) parameters are those to which the model is not sensitive, and any available value is sufficient. Category (ii) parameters are those with highly certain values, and here again the values are not questioned. Category (iii) parameters have values that are variable in time. As a first step, we propose to consider these as forcing variables in hind-cast simulations. This allows us to evaluate whether an average value is appropriate in prediction or whether time-varying values are necessary. Category (iv) parameters are those for which discrete alternative values are identified in literature.

Calibration could be a solution to decide on the most appropriate value. However here, the interval on which the parameters are defined is discrete rather than continuous, a situation not handled well by optimization algorithms (Nocedal and Wright, 2006). As an alternative to calibration, we propose to build alternative formulations of the model using all the possible values for category (iv) parameters and combining them systematically. The formulations could later be compared to identify the best combination. This constitutes the third step. Finally, sensitive parameters whose values are unknown (category v) have to be estimated using an optimization procedure; this is step four. Information available through literature, scientific sampling, and fisheries should as much as possible be used to derive parameter estimates and limit the number of parameters to calibrate. There is an abundant literature on optimization procedures for non-linear models (Kleijnen, 1998; Nocedal and Wright, 2006; Walters et al., 1991). This also requires the building of an objective function and the selection of the most appropriate data to fit the model to, keeping in mind this series should preferably not have been used in setting the values for other parameters (see Duboz et al., 2010; Pasandideh and Niaki, 2006). The optimization procedure must be repeated for each formulation of the model created in step 3 since the value of optimized parameters is a function of the other parameter values. If changes are made to the model, the optimization problem is solved anew (Nocedal and Wright, 2006). Once a formulation is calibrated, it presents the best achievable fit to the time series used in the objective function, which may not be perfect, and comparing the fit of various formulations could be a first step toward model selection.

2.2. Model skill assessment

Once all the parameter values are determined, the validation step is carried out by comparison of model outputs with corresponding observations, usually time series.

Information available on the case study should be reviewed including time series of observed data (either from scientific surveys or fishery dependent) and literature knowledge. For validation, hindcast simulations are run with each calibrated formulation during the period for which time series of observations are available, and simulated variables are compared to observations. Unlike calibration, the fit has not been constrained by parameter tuning and the time series have not been used in parameterization. Thus the fit reflects the real predictive power of the model given the formulation assumptions. This power can be quantified against all the available time series using summary statistics. Three commonly used summary statistics have been selected to reflect three complementary aspects of model fit (Stow et al., 2009):

- correlation of time series: $r = \left(\sum_{i=1}^{n} (O_i - \overline{O})(P_i - \overline{P}) \right) / \left(\sqrt{\sum_{i=1}^{n} (O_i - \overline{O})^2 \sum_{i=1}^{n} (P_i - \overline{P})^2} \right)$
- model prediction accuracy (mean square error): $MSE = \frac{1}{n} \times \sum_{i=1}^{n} (P_i - O_i)^2$
- modelling efficiency (Stow et al., 2009):

$$MEF = \left(\sum_{i=1}^{n} (O_i - \overline{O})^2 - \sum_{i=1}^{n} (P_i - O_i)^2 \right) / \left(\sum_{i=1}^{n} (O_i - \overline{O})^2 \right)$$

where P_i is the i_{th} of n model predictions, O_i the i_{th} of n observations and $\overline{O}$ and $\overline{P}$ are the observation and prediction averages respectively. Correlation ensures that trends are captured, MSE identifies potential biases and reflects accuracy of absolute predictions and MEF measures how well the model predicts relative to the average of the observations. Thus, using the three metrics allows us to characterize trends (r), accuracy (MSE) and efficiency (MEF).

2.3. Assessment of inter-annual variability impact

As a last diagnostic, we propose to assess the impact of inter-annual variation in category (iii) parameters to determine if their value can be averaged in prediction. As suggested by Mackinson et al. (2009), exploratory runs are performed with inter-annual forcing factors relaxed one by one and replaced by their average value. Results are explored using the summary statistics presented above with a focus on correlation.

3. Materials: ISIS-Fish model of the anchovy fishery in the Bay of Biscay

The ISIS-Fish model (Mahévas and Pelletier, 2004; Pelletier et al., 2009) has been specifically developed to assess management measures with special emphasis on spatial measures. It uses a common framework for the description of marine populations and exploitation dynamics and describes processes on a discretized 2D space (based on a regular grid defined by the user). ISIS-Fish is a matrix model that accounts for mortality, growth, reproduction and migration to update population numbers monthly, spatially and per population stage. It relies on the basic equation for survival (exponential decay model) and computes fishing mortality based on the dynamically allocated effort of the fishing fleets in the areas where populations and fishing overlap. A management dynamics sub-model enforces regulation constraints on exploitation monthly and spatially as well as fishermen reactions to those constraints. An application of ISIS-Fish has been developed to assess the impact of spatial management measures on the anchovy fishery in the Bay of Biscay. A summary of the equations and parameters is provided in appendix A. The software and the database of this application can be freely downloaded (http://www.isis-fish.org/download.html). The model consists of around 700 parameters for 17 population stages, 12 months, 15 areas (possibly overlapping, Figure A1), 5 fleets, 14 métiers and 3 gear types (see tables A.1, A.2, A.3 and A.4 for parameter lists and values). The time step is one month. The model describes anchovy life cycle including growth, reproduction, recruitment and migrations. The activity of five fleets fishing on anchovy is described through allocation of effort on métiers. Parameters in the initial formulation (Lehuta et al., 2010) were set up by integrating knowledge from literature and experts and by statistical analyses of log-book data for fishing parameters. The model was calibrated on catches at age and visually validated (Lehuta et al., 2010). Parameter values are assessed monthly or seasonally but assumed to be fixed over years. A sensitivity analysis has been performed on the model using group screening, a fractional factorial to explore the parameter space, and a PLS regression (Tenenhaus, 1995) to apportion the variation in multiple model outputs to each group of parameters (Lehuta et al., 2010). The analysis demonstrated that the most sensitive parameters were the biological parameters related to mortality, growth, reproduction and migration. The model was less sensitive to parameters related to fishing. The two exceptions were accessibility and the gear standardisation factor, which standardize effort between gear. Pelletier et al. (2009) and Lehuta et al. (2010) respectively provide full details of model structure and on the formulation initially proposed for the anchovy fishery in the Bay of Biscay. The purpose of this paper is to objectively review this initial formulation, to gain confidence in parameter values, increase transparency in modeling choices and select the most credible parameterization to be used in management strategy evaluation.

4. Results

Here we present a review of the initial parameterization of the ISIS-Fish model of the Bay of Biscay anchovy and its validation through the methods described above.

4.1. Classification of parameters in categories and alternative formulations

Based on the results of the sensitivity analysis (Lehuta et al., 2010), the parameters of the model were reviewed and classified in the five categories (Table A.1, A.2, A.3, A.4).

Parameters that were not sensitive were classified in category (i) and kept at their reference value. Growth parameters (Table A.1.V1, 3 parameters) and standardization factors used to standardize effort between gear (Table A.4, 3 parameters), although sensitive and uncertain, were considered properly assessed with the best knowledge at hand, and no alternative methods were proposed to review their values. They were thus assigned to category (ii). Fecundity (Table A.1.R1 and R2, 10 parameters) was also sensitive, but the value is accurately known (Motos, 1996) and belongs to category (ii) as well.

Anchovy spawning spatial distributions (Table A.1.S3, 10 parameters) were recorded during scientific surveys. Time spent fishing (Table A.3, 60 parameters) and fishing strategies (Table A.2, 168 parameters) were both assessed from log-books for the period 2000-2008. These parameters all belong to category (iii), that is to say parameters varying in time. Also, larval survival is known to show intrinsic variability, depending on environmental conditions. A time series of estimates of larval survival by year, month and area for the period 2000-2007 was reconstructed based on an individual-based larval drift and survival model (Huret et al., 2010). The bio-physical model predicts the potential survival of larvae depending on spawning area and date of birth (Table A.1.M4, 5 parameters). The effects of annual environmental conditions on larval survival were estimated by fitting an additive linear model to survival rates. In hind-cast simulations over the period 2000-2008, annual values of larval survival, observed stock distribution across spawning areas, and observed fishing effort and strategies of each fleet were used as forcing variables.

Three processes of the model are sensitive and inconsistently or vaguely described in literature. The parameters related to these processes are attributed to Category (iv). The first concerns natural mortality of the first stages, representing the most sensitive parameter of the model. The mortality rate at each stage (15 stages, from eggs to young adults) is determined by adjusting a mortality curve following a Pareto decay (Lo et al., 1995; Table A.1.M3) from birth, using the value of egg mortality, to the end of the first reproduction (455 days) assuming that mortality rate is the same as the adults‘ rate from this stage on (Table A.1.M2, 3 parameters). Two possible estimates of egg mortality were found for *Engraulis encrasicolus* making the choice uncertain: a survey-based estimate of 0.266 day^{-1} in the Bay of Biscay (2000-2005) (ICES, 2007) and an egg mortality value of 0.565 day^{-1} assessed by Pertierra et al. (1997) in the Catalan Sea (Table A.1.M1). Although the environmental conditions can explain such a difference in estimates for the same species, differences due to measurement error and assessment methods cannot be excluded. Secondly, although quite accurate, fecundity was sensitive which indicates the importance of assumptions related to reproduction parameters in the model. Among them, individual spawning duration has not been characterized precisely (Table A.1.R7, 8 parameters). Pertierra et al. (1997) estimated spawning to last for three months, based on the number of batches and spawning frequency, while Motos (1996) gave a duration of 2.5 months and differentiated between fish of length above and below 14 cm. Thirdly, as seen previously, migration is a sensitive process. The time of departure to the spawning grounds is inaccurately known and described by Uriarte et al. (1996) as occurring in winter, which can correspond to four months (Table A.1.S2).

In order to limit the number of formulations that need to be built to reflect the uncertainty in category (iv) parameters, we decided to limit the number of alternatives per uncertain parameter to two. We kept the two alternative values found for egg mortality (ICES estimate **Megg1** and the Pertierra et al. (1997) estimate **Megg2)** as possible starting points for mortality curves (Table A.1.M1). For the spawning duration of small fish (five stages born the previous year in different months), we defined a first hypothesis **R1** in which the two smallest stages (born in July and August of the previous year) spawn during two months only while the other three stages (born earlier in the year) spawn during the full period of three months as adults do. In the alternative hypothesis **R2**, the three smallest stages (born from June to

August) spawn for two months only, while the larger fish spawn during three months (Table a.1.R7). Finally two contrasted months of migration to spawning grounds, corresponding to the extremes of the uncertainty range, were considered: January representing the beginning of the winter season (hypothesis **MigJ**), and April representing the end (hypothesis **MigA**). The possible formulations are built by combining the discrete values selected for each parameter (2 discrete values for 3 parameters result in 8 = 2x2x2 combinations).

4.2. Calibration of category (v) parameters

Adult mortality (Table A.1.M2, 3 parameters for 3 adult stages) and accessibility (Table A.1.A, 204 parameters for 17 stages and 12 months) were also sensitive parameters that were both uncertain and hardly measured. Natural mortality can be considered almost constant after the first reproduction (Chen and Watanabe, 1989), that is for anchovy for age 1+ fish. Natural mortality for mature fish is estimated at 1.2 year^{-1} by ICES (ICES, 2009). To assess natural mortality rates for every stage from larvae to recruits, a Pareto model of mortality was fitted using an adult mortality of 1.2 and each of the two egg mortalities previously cited (0.266 day^{-1} (ICES, 2007) or 0.56 day^{-1} (Pertierra et al. 1997)). The resulting survival rates for the first year were of about 10^{-4}. This is high and inconsistent with the value of 10^{-5} estimated by Petitgas and Massé (2003) based on biomass and recruitment time series and average fecundity, and confirmed using other methods (IBM model, Allain et al., 2007;) and on another anchovy (*E. mordax*, Petterman et al., 1988). Higher egg mortality rates or adult mortality rates would be necessary to reach the estimated survival. Estimates of egg mortality being available from field studies, we considered the value 1.2 of adult mortality as questionable and requiring calibration.

Accessibility in ISIS-Fish is the biological component of catchability and quantifies changes in the probability of a fish to be caught due to the biology or the ecology of the stock (ex. behaviour: burying, schooling)(accessibility and vulnerability sensu Mahevas et al. 2011). It is a key parameter of the model as it is one of the parameters which convert effort into fishing mortality. No value was available in the literature for the accessibility parameter. Consequently we relied on expert judgement for the description of the parameter. According to experts, the accessibility of anchovy is mainly age-dependent since area and time variations due to changes in fish density are explicitly modelled, and the schooling behaviour mostly changes after recruitment (P. Petitgas, pers.comm.). Three values of accessibility were thus considered: q0, corresponding to age-0 fish (until the first reproduction); q1, for age-1 and q2, for age-2+. Accessibility parameters and the natural mortality of adults are considered category (v) and needed to be assessed by calibration for each of the eight parameterizations of the anchovy model.

For category (v) parameters, the methodology recommends calibration. As natural mortality and accessibility may have confounding effects, a sequential calibration is used to assess parameters related to the two processes independently. We first assess natural mortality of adults. Since the fishery was closed between 2005 and 2008, accessibility values are not necessary to simulate the population over the period. Because the values of natural mortality for larvae and juveniles depend on adult mortality through the Pareto model, the coefficients of the Pareto model are estimated using a range of values for adult mortality ([0.5;3] with 0.01 increments) and then used to simulate population dynamics over the period 2005-2008. Population growth rate is chosen to be the optimization criterion assuming that mortality is an important driver of biomass trends. The slope of the SSB time series over the period 2005-2008 (ICES, 2009) is thus compared to the slope simulated with the set of Pareto models, and the best Pareto model is retained for each of the 8 formulations. Corresponding adult mortality rates are much higher than the value of 1.2 used by ICES (from 1.63 to 3.03 depending on the parameterization) (Table 1).

Secondly, we assess accessibility during a period when the fishery was opened (2000-2004). Parameter values for accessibility coefficients q0, q1, q2 are optimised using the variable step simplex algorithm (Walters et al., 1991). The objective function is the sum of squared differences (MSE) (least square minimisation) between the catches at age by quarter reported by ICES (annual reports of the working group for 2001 to 2005, (ICES, 2006; ICES, 2005; ICES, 2004; ICES, 2003; ICES, 2002)) and the simulated ones. Catches at age are chosen because they are directly related to catchability at age through the Baranov equation. Calibrated accessibility values of juvenile fish (before the first reproduction q0) are 100 times lower than those of adults regardless of the model formulation (Table 1). The largest differences between accessibility values (particularly for q1) occur between formulations in which egg mortality differed (Table 1). According to the least squares minimisation between simulated and observed catches, the best fit (smallest MSE) is obtained between simulated and observed catches for the formulation with hypotheses **Megg2**, **R2** and **MigJ** (Table 1) and more generally for formulations assuming **Megg2**. This indicates that egg mortality could be higher than the value estimated in the Bay of Biscay.

At the end of the parameterisation stage, the use of the uncertainty hierarchy enabled us to reduce the number of parameters to assess or refine from about 700 to 254 (catchability, natural mortality of eggs, larvae per area and of adults, migrations rates to spawning grounds, migration date, fishing time, strategies) among which only 4 are calibrated (category (v): adult natural mortality, q0, q1, q2). The calibration process (in two stages natural mortality first then catchability using the simplex) required about 1500 simulations for all 8 formulations.

4.3. Model skill assessment

Available observations for validation purposes are: the absolute index of biomass and recruitment biomass annually obtained by acoustics (PELGAS surveys, IFREMER, ICES, 2009), the average distribution of egg production per month over the spawning season (Allain et al., 2007), the monthly biomass of catch by fleet reported in logbooks for the period 2000-2004 (French database Harmonie, FIS, IFREMER; Spanish data, Leire Ibaibarriaga, pers. comm.), and the annual total biomass of catch over the period 2000-2004 (ICES, 2006). Although used for calibration, the values of summary statistics for the seasonal catch in numbers at age per country are considered in the assessment of model skill (ICES 2006; ICES 2005; ICES, 2004; ICES, 2003, ICES, 2002).

The three summary statistics are examined successively. Temporal correlations range from 0.18 for the catches at age 2 to 0.99 for spawning distributions (Figure 2, r). Although the models are calibrated for minimum squared differences on catches at age, the temporal correlations seldom exceed 0.8 for these output variables. Correlation values do not differ much between formulations, except for catches at age 2, for which hypotheses **Megg2** and **MigA** appears more suitable (Figure 2, r).

MSE values differ considerably between output variables because of the variety of metrics and temporal scales. To be able to compare alternative formulations across output variables, the value displayed on the radar plot is the rank of each formulation according to its MSE value (8 corresponding to the lowest MSE value: best, 1 to the highest: worst) (Figure 2, MSE). The rank of a formulation highly depends on the output variable considered. Annual catches (Catch/y), catches at age (Cage), biomass and spawning distribution (Sp. Distr.) appear more accurately reproduced using models with hypotheses **Megg2** and **MigA**, while **Megg1** gives better fit to catches of trawler fleets (CPel) (Figure 2, MSE). The representation using ranks enables formulations to be compared according to spawning distribution when they perform identically according to MEF and correlation (Sp. Distr., Figure 2, MSE).

Spawning distribution is probably the most direct consequence of the hypotheses on reproduction patterns. Confrontation of simulated and observed time series (not shown), evidences that the spatial and temporal distribution of egg production is close to the average pattern reported by Allain et al. (2007), although inter-annual variations are important particularly in July. Differences mostly concern the month in which egg production peaks. Combining **Megg2** and **R1**is necessary to simulate a peak in egg abundance in June as observed. Therefore spawning duration for most of the age-1 fish could be as long as that of age-2 (R1) and spawning timing is probably influenced by age structure (Megg 2).

Finally, model efficiency is relatively low with frequent negative values indicating that the observation average would be a better predictor than the model results (Figure 2, MEF). In particular, modelled catches of purse seiners profile 1 (CBol1), and trawlers profile 2 (CPel 2) are far from observations. This is mainly due to the opportunism of these fleets. Anchovy is not their main target, and they catch it occasionally when they encounter a school, which is hardly described by the model because the time step is the month and the movements of fish schools are not simulated. Here again, the best model depends on the output variable considered (Figure 2, MEF). Formulations with **Megg2** give poor results when considering catches of trawler fleets (particularly CPel2), mostly because catches in the first year are overestimated, but these formulations are appropriate for the majority of the other outputs. Given the good performances across almost all output variables, we use the formulation with **Megg2**, **R1** and **MigA**, in the following to investigate the impact of inter-annual variability on the modelled dynamics.

4.4. Impact of inter-annual variability

Exploratory simulations are run with the selected formulation (egg mortality hypothesis **Megg2**, first reproduction function (**R1**) and migration of adults in April (**MigA**)). Each output variable reacts differently and reveals the process that mostly impacts its dynamics (Figure 3). The effect of relaxing a forcing (using average values) is seen by comparing the forced model (black line) to the model with average values for the process considered. Biomass, catches at age 0, recruitment and catches of purse seiners fleets, are highly influenced by the forcing by larval survival (Figure 3, ―Average Survival”). Releasing the forcing generally improves the correlation except for catches at age 0, showing that the processes occurring between egg and recruit stages are not fully resolved using the individual-based larval drift and survival model. However, visual validation showed that the use of the forcing by the larval survival time series improves the reproduction of the biomass trend with the exception of year 2002 (Figure 4). Average migration rates have a low impact and generally decrease correlation while average effort substantially decreases the correlation for purse seiners profile 1 (Figure 3). Consequently, scenarios are needed for larval survival and migration, as they largely influence the dynamics. On the other hand, the average effort and strategies should be appropriate for most of the fleets except for purse seiner profile1 (CBol1).

5. Discussion

Ideally integrated estimation using multiple time series and objective functions is performed to assess parameter values in a complex model. Many methods exist to address parameter estimation (both discrete and continuous) and model selection, in both frequentist and Bayesian frameworks, as long as a likelihood function can be built and simulation costs are not limiting. Autodifferentiation as proposed by the software ADMB is often used for fishery models (http://www.otter-rsch.com/ ; Drouineau et al., 2010). Markov processes and Bayesian frameworks when usable offer transparent ways to deal with various sources and the nature of uncertainty (Hillary et al., 2012; Williams, 2011; Bertorelle et al., 2010).

Two major difficulties linked to the nature of our model prevented the use of an integrated optimization procedure and motivated the development of the proposed methodology: (1) the duration of simulations, which constrained the model to be deterministic and the optimization procedure to be limited to few parameters; and (2) its complexity, which prevents a likelihood function to be defined. We also needed a procedure which could deal with discrete parameters, like integers (months) or formulas (reproduction duration) or even more complex (spatial limits of areas) as well as continuous parameters, a situation not well handled by common optimization algorithms (based on computation of derivative for instance). In addition, an integrated estimation (e.g. using likelihood) would require many data sources and their compliance with model assumptions, a situation seldom encountered (Pech et al., 2001). An alternative would have been the building of a metamodel, for instance using response surface methodology, to decrease simulation time and allow the methods cited above to be used (Myers et Montgomery, 2002).

However, the building of a metamodel could itself be time consuming and delicate. In such cases, our procedure offers a pragmatic and transparent way to use all available information to provide parameter estimates and evaluate the credibility of the model. To limit computation time and prevent identifiability problems, the optimization was restrained to as few parameters as possible based on a sensitivity analysis and a classification of sensitive uncertain parameters into 3 categories according to the nature of their uncertainty. Calibration was used for only one of these categories and alternative formulations were considered. Although an effort was made towards rational attribution of parameters to categories, it cannot be denied that the classification is partly subjective. Similarly, the solutions provided to treat each category do not pretend to be unique and other treatments could be proposed. The assessed values for category (ii) for instance, although less uncertain, are unlikely to be exact, and confidence intervals or distribution laws could have been used to create predictive distributions for model outputs. Category (iii) processes were treated as forcing variables. In a less computer-intensive model, they could have been considered latent processes and a stochastic procedure run to assess the process error associated, as is done in state-space modelling in Bayesian frameworks (Harwood & Stokes, 2003). In a deterministic context, one can also propose alternative underlying mechanisms possibly responsible for these variations and compare their fit; they would then belong to category (iv). The alternative values found for category (iv) were interpreted as inconsistent or conflicting. It is actually rational that different cases study or methodologies (empirical, model estimates) or time periods of observations give different estimates, and the uncertainty could be described as resulting from process or measurement error. Optimization for these parameters could have resulted in a new estimate, possibly different from those found in the literature. Keeping alternative model formulations can be compulsory when, for instance, the inconsistency concerns incompatible model structures. Since the number of alternative formulations can rapidly increase, it could be necessary to limit their definition by adopting other sampling strategies of the parameter space or by eliminating unlikely combinations.

Keeping alternative formulations presents the advantage of guaranteeing that parameters are meaningful outside the model context. Indeed, the value of parameters estimated through an optimization procedure is conditional to assumptions made elsewhere in the model, and may thus compensate for the value of other parameters to achieve a good fit. The estimate of adult mortality can be discussed as an example. First, the short duration of the calibration period (only four years) gives low robustness to the mortality estimate. Second, the values appeared very sensitive to other assumptions of the model. The value of 3.03 for instance was out of range according to experts. It should be interpreted in the context of our model as it compensates for higher early survival and reproduction success estimates.

Both for calibration and validation, the observations should be distinguished from the truth (Lynch et al., 2009). Observation data may be subject to errors, for instance landings reported in log-books may be incomplete or erroneous. Ideally, validation data must be directly observed, and not the result of an estimation process, because the estimation methods might use assumptions inconsistent with the assumptions of the model. Here the model was validated using indices of biomass estimated by acoustics, which do not involve any biological knowledge on anchovy. Anchovy biomass estimated by ICES was also available, however it is the result of an assessment model (ICES, 2009) for which assumptions regarding recruitment (random) and adult mortality (1.2) differed from ours. Even the indices of biomass estimated using the Daily Egg Production Method required parameters also used in the model (fecundity and egg mortality). Although the use of acoustic data was recommended for validation, the comparison with the other time series (Figure 4) informed of the uncertainty around biomass and increased our confidence in model realism.

Anchovy being well documented in the Bay of Biscay, we were able to corroborate model outputs against multiple time series not used in parameterization. This could be seen as a concrete application of pattern oriented modelling through the validation of ‗secondary predictions‘ (Wiegand et al., 2003). The use of several time series of outputs, which result from the interaction of multiple parameters, guaranteed the model ‗structural realism'. In effect, the model cannot reproduce simultaneously multiple patterns observed at different scales and hierarchical levels, if key processes are not captured realistically (Cury et al., 2008). Secondly, because of the multiple, confounding factors that can have synergistic or antagonistic effects on fishery dynamics, one needs more than a single series of observations to distinguish between several hypotheses of description for a phenomenon (Mackinson et al., 2009). It enables model developers to filter unlikely combinations of hypotheses even if their effects are generally confounded. For instance, model fit on biomass was equivalent across parameterizations, as biomass is mainly driven by variation in recruitment. However, the underestimation of egg mortality in parameterizations with **Megg1** induced incompatible low proportions of age-2+ in the population and very low catches for these classes. Allen and Somerfield (2009) also propose to validate the relationships between model variables to guaranty the realism of model emergent properties.

Here we tried to make the validation step more objective by quantifying model skill through three statistical criteria. This approach is largely applied for validation of coupled physical-biological models (Allen and Somerfield, 2009; Jolliff et al., 2009; Stow et al., 2009) that often make use of geo-referenced validation data. No agreement exists on the best and number of criteria to use, and they must be selected according to the kind of confidence the user wants to acquire on its model (see Sterman, 1984, for a classification of possible confidence tests). The rational for the three we chose was that they summarised 3 important aspects of model fit we were interested in and are probably of importance for any fishery model. Indeed, (i) small MSE ensures that absolute values can be used in prediction (TAC, reference points, etc); (ii) good correlations prove that mechanisms responsible for past trends are understood and (iii) efficiency informs on variables that could be replaced by the average of the observations and should not be trusted in prediction. They offer complementary information, for instance showing that although catch realised by the trawler fleet profile 2 is well correlated with observations, it does not perform better than the average time series. Other criteria proposed in the literature were reliability index (Stow et al., 2003), the normalised standard deviations (Jolliff et al., 2009), and rank correlations (Allen and Somerfield, 2009). Further developments of summary statistics for time series and spatial data would also be interesting to consider in further studies. It was implicitly done here as output variables had various time and spatial aggregation levels (but see Bellassen et al., 2011 for explicit account of three spatial scales in validation). We presented multivariate results using radar plots for each criterion to allow visual comparison of the variables. Alternatively, Jolliff et al. (2009), proposed summary diagrams such as Taylor diagrams together with reference points to help summarise one model's skill, while Allen and Somerfield (2009) relied on multivariate analyses.

Finally, validation will ever be in some ways arbitrary as long as no reference values for summary statistics are unanimously accepted. An advantage of our method is that it transparently evidenced the trade-offs, strengths and weaknesses of modeling choices. Even if reproducing time series were to give confidence in the model, as advocated by Mackinson et al. (2009), failures in fitting the model were also informative and offered the opportunity to reveal gaps in the current understanding of the system and provide indications of where further knowledge could be usefully gained. We identified weaknesses in the simulated dynamics of catches at age 0 and 2 and catches of purse seiners from Brittany.

In the present case, none of the formulations was unanimously supported, neither across summary statistics, nor across output variables. It is nevertheless possible to conclude that a high value of egg mortality, surprisingly closer from the value observed in the Mediterranean than in the Atlantic, was necessary to reproduce the dynamics of validation data (except catches of pair trawler fleets). The choice of a single formulation among the eight evaluated requires trade-offs. In many cases, experts help in sorting hypotheses or weight knowledge sources according to their reliability (Stefansson, 1998). However, reliability could be hard to determine. When the sources are so conflicting that the model cannot explain all the data sources simultaneously, Pech et al. (2001) proposed an iterative procedure of partial fit, and Stefansson (1998) suggested questioning model structure. If a valid likelihood function can be constructed, the Akaike criteria or Bayesian approaches are well adapted to assign weights to alternative assumptions (see for instance Patterson, 1999). Recent emergence of Approximate

Bayesian Computation offers a more flexible framework to choose between models (Beaumont, 2010). Model averaging is also an option and is used for instance if the structures of the available models are very different (climate models for instance, Hill et al., 2007). Otherwise alternative hypotheses should normally be stored for inclusion in uncertainty analyses (Hill et al., 2007). Here we gave preference to a formulation that was performing satisfyingly on every variable rather than the best on one variable; this choice is of course debatable. Despite the weaknesses evidenced, we deem that the main processes of the anchovy fishery dynamics were understood and successfully modelled. The model was developed to assess the impact of marine protected areas on the anchovy fishery and given the validation results, we are confident that any of the eight formulations (ideally all eight) could be used to support management decisions. It would have to be run within an uncertainty framework that ensures that all sources of uncertainty are covered. Additionally, scenarios or random trajectories of inter-annual variations in larval survival and migration rates should be tested.

Acknowledgements

This study was carried out with financial support from the Commission of the European Communities as a contribution to FP6 Specific Targeted Research Project 022717 (UNCOVER). Authors are grateful to Jacques Masse (IFREMER) for his expertise on the fishery. We thank Leire Ibaibarriaga (AZTI) for supplying the data on the Spanish Basque fishery 2000-2003 and for expert advice on its parameterization. We thank Kathy Mills for her precious editing comments. Technical support on ISIS-Fish was warranted by the private company Code Lutin. The official French catch and effort data were extracted from Harmonie, a database containing the Fisheries Information System managed by IFREMER.

References

Allain, G., Petitgas, P., Lazure, P., and Grellier, P., 2007. Biophysical modelling of larval drift, growth and survival for the prediction of anchovy (Engraulis encrasicolus) recruitment in the Bay of Biscay (NE Atlantic). Fisheries Oceanography. 16(6), 489-505.

Allen, J.I., Somerfield, P.J., and Gilbert, F.J., 2007. Quantifying uncertainty in high-resolution coupled hydrodynamic-ecosystem models. Journal of Marine Systems. 64, 3-14.

Allen, J.I., and Somerfield, P.J., 2009. A multivariate approach to model skill assessment. Journal of Marine Systems. 76, 83-94.

Beaumont, M.A.. Approximate Bayesian Computation in Evolution and Ecology. In *Annual Review of Ecology Evolution and Systematics*. Futuyma, DJ and Shafer, HB and Simberloff, D (Ed.). Vol. 41. 2010.

Bellassen, V., le Maire, G., Guin, O., Dhôte, J., Ciais, P., and Viovy, N., 2011. Modelling forest management within a global vegetation model—Part 2: Model validation from a tree to a continental scale. Ecological Modelling. 222, 57–75.

Bertorelle, G., Benazzo, A., and Mona, S., 2010. ABC as a flexible framework to estimate demography over space and time: some cons, many pros. Molecular Ecology. 19, 2609-2625.

Charles, A., 1998. Living with uncertainty in fisheries: analytical methods, management priorities and the Canadian groundfishery experience. Fisheries Research. 37, 37-50.

Chen, S., and Watanabe, S., 1989. Age dependence of natural mortality coefficient in fish population dynamics. Nippon Suisan Gakkaishi. 55(2), 205-208.

Cury, P.M., Shin, Y., Planque, B., Durant, J.M., Fromentin, J., Kramer-Schadt, S.et al., 2008. Ecosystem oceanography for global change in fisheries. Trends in Ecology and Evolution. 23, 338-346.

Drouineau, H., Mahevas, S., Bertignac, M., and Duplisea, D., 2010. A length-structured spatially explicit model for estimating hake growth and migration rates. ICES Journal of Marine Science. 67, 1697-1709.

Duboz, R., Versmisse, D., Travers, M., Ramat, E., and Shin, Y., 2010. Application of an evolutionary algorithm to the inverse parameter estimation of an individual-based model. Ecological Modelling. 221, 840-849.

Fulton, E.A., Smith, A.D.M., and Johnson, C.R., 2004. Biogeochemical marine ecosystem models I: IGBEM—a model of marine bay ecosystems. Ecological Modelling. 174, 267-307.

Harwood, J., Stokes, K., 2003. Coping with uncertainty in ecological advice: lessons from fisheries. 18, 617-622.

Hill, S.L., Watters, G.M., Punt, A.E., McAllister, M.K., Le Quere, C., and Turner, J., 2007. Model uncertainty in the ecosystem approach to fisheries. Fish and Fisheries. 8, 315-336.

Hillary, R.M., Levontin, P., Kuikka, S., Manteniemi, S., Mosqueira, I., and Kell, L., 2012. Multi-level stock-recruit analysis: Beyond steepness and into model uncertainty. Ecological Modelling. 242, 69-80.

Huret, M., Petitgas, P., and Woillez, M., 2010. Dispersal kernels and their drivers captured with a hydrodynamic model and spatial indices: A case study on anchovy (Engraulis encrasicolus) early life stages in the Bay of Biscay. Progress in Oceanography. 87, 6-17.

ICES. 2002. Report of the Working Group on the Assessment of Mackerel, Horse Mackerel, Sardine and Anchovy. ICES CM 2002/ACFM:06: pp.

ICES. 2003. Report of the Working Group on the Assessment of Mackerel, Horse Mackerel, Sardine and Anchovy. ICES CM 2003/ACFM:07: pp.

ICES, 2004. Report of the Working Group on the Assessment of Mackerel, Horse Mackerel, Sardine and Anchovy. ICES CM 2004/ACFM:08. , .

ICES. 2005. Report of the Working Group on the Assessment of Mackerel, Horse Mackerel, Sardine and Anchovy. ICES CM 2005/ACFM:08: pp.

ICES. 2006. Report of the Working Group on the Assessment of Mackerel, Horse Mackerel, Sardine and Anchovy. ICES CM 2006/ACFM:36ICES CM 2006/ACFM:36.

ICES. 2007. Report of the working group on the assessment of mackerel, horse mackerel, sardine and anchovy. ICES CM 2007/ACFM:31.

ICES. 2009. Report of the Working Group on Anchovy and Sardine. ICES CM 2009/ACOM:13, 354pp.

Jolliff, J.K., Kindle, J.C., Shulman, I., Penta, B., Friedrichs, M.A.M., Helber, R.et al., 2009. Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. Journal of Marine Systems. 76, 64-82.

Kleijnen, J.P.C. 1998. Experimental Design for Sensitivity Analysis, Optimization, and Validation of Simulations Models. *In* Handbook of simulation. Principles, Methodology, Advances, Applications and Practise. Banks, J. Wiley, New York. Engineering and Management Press (Ed.). 173-224.

Kraus, G., Pelletier, D., Dubreuil, J., Moellmann, C., Hinrichsen, H., Bastardie, F.et al., 2009. A model-based evaluation of Marine Protected Areas: the example of eastern Baltic cod (Gadus morhua callarias L.). ICES Journal of Marine Science. 66, 109-121.

Lehuta, S., Mahévas, S., Petitgas, P., and Pelletier, D., 2010. Combining sensitivity and uncertainty analysis to evaluate the impact of management measures with ISIS–Fish: marine protected areas for the Bay of Biscay anchovy (Engraulis encrasicolus) fishery. ICES journal of marine science. 67, 1063–1075.

Lo, N., Smith, P., and Butler, J., 1995. Population growth of the northern anchovy and pacific sardine unsing stage-specific matrix models. Marine Ecology Progress Series. 127, 15-26.

Lynch, D.R., Jr, D.J.M., and Werner, F.E., 2009. Skill assessment for coupled biological/physical models of marine systems. Journal of Marine Systems. 76, 1 - 3.

Mackinson, S., Daskalov, G., Heymans, J.J., Neira, S., Arancibia, H., Zetina-Rejón, M.et al., 2009. Which forcing factors fit? Using ecosystem models to investigate the relative influence of fishing and changes in primary productivity on the dynamics of marine ecosystems. Ecological Modelling. 220, 2972–2987.

Mahevas, S., Trenkel, V.M., Doray, M., Peyronnet, A., 2011. Hake catchability by the French trawler fleet in the Bay of Biscay: estimating technical and biological components. ICES Journal of Marine Science. 68, 107-118.

Mahévas, S., and Pelletier, D., 2004. ISIS-FISH, a generic and spatially explicit simulation tool for evaluating the impact of management measures on fisheries dynamics. Ecological Modelling. 171, 65-84.

Motos, L., 1996. Reproductive biology and fecundity of the Bay of Biscay anchovy population (Engraulis encrasicolus L). Scientia Marina. 60, 195-207.

Myers, R H, and Montgomery, D.C. 2002. Response Surface Methodology, second ed. Wiley Series in Probability and Statistics.

Nocedal, J., and Wright, S.J. 2006. Numerical Optimization, 2nd Ed. Mikosch, T. V., Resnick, S. I. , and Robinson, S. M. (Eds.). Springer Series in Operations Research, New York.

Pasandideh, S.H.R., and Niaki, S.T.A., 2006. Multi-response simulation optimization using genetic algorithm within desirability function framework. Applied Mathematics and Computation. 175, 366-382.

Patterson, K., 1999. Evaluating uncertainty in harvest control law catches using Bayesian Markov chain Monte Carlo virtual population analysis with adaptative rejection sampling and including structural uncertainty. Canadian Journal of Fisheries and Aquatic Sciences. 56, 208-221.

Pech, N., Samba, A., Drapeau, L., Sabatier, R., and Laloë, F., 2001. Fitting a model of flexible multifleet-multispecies fisheries to Snegalese artisanal fishery data. Aquatic Living Resources. 14, 81-98.

Pelletier, D., Mahevas, S., Drouineau, H., Vermard, Y., Thebaud, O., Guyader, O.et al., 2009. Evaluation of the bioeconomic sustainability of multi-species multi-fleet fisheries under a wide range of policy options using ISIS-Fish. Ecological Modelling. 220, 1013 - 1033.

Pelletier, D., and Mahévas, S., 2005. Fisheries simulation models for evaluating the impact of management policies, with emphasis on marine protected areas. Fish and Fisheries. 6, 307-349.

Pertierra, J., Lleonart, J., and Lo, N., 1997. Application of a stage-specific matrix model and length-cohort based analysis to assess the anchovy fishery in Catalan coastal waters (NW Mediterranean Sea). Fisheries Research. 30, 127-137.

Petitgas, P., and Massé, J.. 2003. Orders of magnitude for some biological processes in Biscay anchovy population. Working document of the ICES Working Group on the Assessment of Mackerel, Horse Mackerel, Sardine and Anchovy. Working document of the ICES Working Group on the Assessment of Mackerel, Horse Mackerel, Sardine and Anchovy. Copenhagen, 9-19 September 2003.

Peterman, R.M., Bradford, M.J., Lo, N.C., Methot, R.D., 1988. Contribution of early life stages to interannual variability in recruitment of Northern Anchovy. Canadian Journal of Fisheries and Aquatic Sciences. 45, 8-16.

Punt, A.E., Donovan, G.P., 2007. Developing management procedures that are robust to uncertainty: lessons from the International Whaling Commission. ICES Journal of Marine Science. 64, 603-612.

Saltelli, A., Chan, K., and Scott, E. 2000. Sensitivity Analysis. Saltelli A, Chan, K. , and Scott, E. (Eds.). Wiley.

Stefansson, G. 1998. Comparing different information sources in a multispecies context. *In* Fishery Stock Assessment Models. Alaska Sea Grant College Program (Ed.). 741-758.

Sterman, J. D. 1984. Appropriate summary statistics for evaluating the historical fit of system dynamics models. Dynamica. 10 (Winter), 51-66.

Stow, C.A., Jolliff, J., McGillicuddy, Jr., Dennis J., Doney, S.C., Allen, J.I., Friedrichs, M.A.M.et al., 2009. Skill assessment for coupled biological/physical models of marine systems. Journal of Marine Systems. 76, 4-15.

Stow, C.A., Roessler, C., Borsuk, M., Bowen, J., and Reckhow, K., 2003. A comparison of estuarine water quality models for TMDL development in the Neuse River Estuary. Journal of Water Resources Planning and Management. 129, 307-314.

Tenenhaus, M., Gauchy, J.P., Ménardo, C., 1995. Régression PLS et applications. Tome 43, n°1, 7-63.

Travers, M., Shin, Y., Shannon, L., Cury, P., 2006. Simulating and testing the sensitivity of ecosystem-based indicators to fishing in the southern Benguela ecosystem. Canadian Journal of Fisheries and Aquatic Sciences. 67, 943-956.

Uriarte, A., Prouzet, P., Villamor, B., 1996. Bay of Biscay and Ibero Atlantic anchovy populations and their fisheries. Scientia Marina. 60 (Suppl. 2), 237-255.

Walters, F.H., Parker, L.R., Morgan, S.L., and Deming, B.R. 1991. Sequential Simplex optimization: a technique for improving quality and productivity in research, development, and manufacturing (Chemometrics series). Walters, F. H., Parker, L. R., Morgan, S. L., and Deming, B. R. (Eds.). CRC Press LLC, Wiegand, T., Jeltsch, F., Hanski, I., and Grimm, V., 2003. Using pattern-oriented modeling for revealing hidden information: a key for reconciling ecological theory and application. Oikos. 100, 209-222.

Wiegand, T., Revilla, E., and Knauer, F., 2004. Dealing with uncertainty in spatially explicit population models. Biodiversity and Conservation. 13, 53-78.

Williams, B.K., 2011. Resolving structural uncertainty in natural resources management using POMDP approaches. Ecological Modelling. 222, 1092-1102.

Appendix A: ISIS-Fish model, equations and parameters of the Bay of Biscay anchovy application

ISIS-Fish is a matrix model based on three sub-models for populations, fleets and management. The population sub-model describes the processes that occur seasonally in the different areas defined for the population(s) (growth, natural mortality, reproduction, migrations). At each time step, the model updates population numbers per age- or size class and area according to those biological processes and to the natural and fishing mortality suffered (eq. 1).

$$N(t+1) = SR(t)R(t) + D^{mig}_{season}CC_{season}N(t) + N^{immig}_{season} \quad \text{(eq.1)}$$

where N(t) and N(t+1) are the area- and class-specific matrix of population numbers at time t and t+1. R(t) is the recruitment vector, D^{mig}_{season} is the migration matrix, N^{immig}_{season} is the immigration vector, CC_{season} is the matrix depicting change of class due to aging in the case of an age-structured model, and to individual growth in the case of a stage-structured model and SR(t) is the diagonal matrix of survival rates of the population between time t and t+1.

The anchovy population was modeled with 17 stages, 15 from spawning to first reproduction to account for mortality at each stage and variation in size when first reproducing due to early/late hatching. The population migrate seasonally from the north of the Bay of Bicay to spawning areas and back (Figure A.1, table A.1).

Survival rate is computed according to the classic exponential decay model (eq 2):

$$sr(c, zpop, t) = \exp(-(F(c, zpop, t) + M(c, zpop)/12)) \quad \text{(eq.2)}$$

where c is age- or size-class, zpop is population zone, F is the fishing mortality rate (month^{-1}) and M natural mortality rate (year^{-1}) that can vary with seasons and areas.

The fleet dynamics sub-model describes the spatio-seasonal distribution of fishing effort according to vessels characteristics, métiers, and annual fishing strategies. It computes fishing mortality for each population area based on fishing effort in the overlapping fishing areas. Effort is deduced from fishing time depending on several standardization factors for métiers (target factors) and gear (gear standardization factor and selectivity coefficient) (table A.2). The anchovy fishery is described by 5 fleets according to gear used (purse seine or pair trawl), home harbor (Brittany, Basque Country, Spain, La Tuballe). Number of vessels in a fleet is the 2000-2003 average. Each fleet displays different fishing strategy, characterized by the distribution of monthly effort on each possible métier (table A.2). Métiers were identified at the fishing operation scale by gear used, area of practice and catch profile. They are characterized by target factors (mean percentage of anchovy in the landings per métier trip per month) (table A.2). Gears standardization factors were computed by statistical analyses of catch per unit of effort data and no selectivity function is assumed; all fish above 9cm are caught (table A.4). Total fishing time per vessel and month for each fleet is the average over boats and over the period 2000-2004 (Table A.3). It is distributed among métiers according to monthly proportions (Table A.2) computed as the 2000-2004 averages.

Management dynamics sub-model enforces regulation constraints on exploitation monthly and spatially as well as fishermen reactions to those constraints. We assumed no constraints on the fishery for the period 2000-2004 preceding the closure.

Except for the last part of the study (impact of inter-annual variations) parameter values are fixed for the entire simulation period.

Table

Hindcast models forced by past time series	Parameterizations			Calibrated parameters				Comparison within formulations		Comparison across formulations	MSE
	Egg mortality	Repro-duction duration	Migration date	Adult morta-lity	q0	q1	q2	q0/q2	q1/q2	q1(for-mulation i) / q1(for-mulation 1)	
	Megg1 = 0.266 day^{-1}	**R1**	April	3.03 yr^{-1}	3.56 e-5	4.68 e-3	2.20 e-3	0.02	2.12	1.00	2.66e18
			January		3.23 e-5	4.46 e-3	2.41 e-3	0.01	1.85	0.95	2.70e18
		R2	April	2.97 yr^{-1}	3.59 e-5	4.48 e-3	2.18 e-3	0.02	2.05	0.96	2.62e18
			January		3.30 e-5	4.29 e-3	2.36 e-3	0.01	1.82	0.92	2.66e18
	Megg2 = 0.565 day^{-1}	**R1**	April	1.67 yr^{-1}	8.98 e-5	2.66 e-3	2.70 e-3	0.03	0.99	0.57	2.36e18
			January		8.35 e-5	2.27 e-3	3.31 e-3	0.03	0.69	0.49	2.35e18
		R2	April	1.63 yr^{-1}	8.44 e-5	2.59 e-3	2.70 e-3	0.03	0.96	0.55	2.29e18
			January		7.67 e-5	2.18 e-3	3.23 e-3	0.02	0.68	0.47	2.28e18

Table 1: Calibration results of the adult natural mortality and accessibility coefficients (q0, q1 and q2) and corresponding minimum squared errors (MSE) between simulated and observed catches at age per quarter, for different scenarios of egg mortality, reproduction duration and migration date. Ratios of values at age (q0/q2; q1/q2) ease comparison within a formulation. Ratios of q1values obtained for each formulation help comparison between formulations.

Table

<table>
<tr><th></th><th>Process</th><th>Modelling assumptions</th><th>Parameter values</th><th># of parameters</th><th>Category</th><th>Reference</th></tr>
<tr><td>S</td><td colspan="6">Spatial organization</td></tr>
<tr><td>S1</td><td>Population areas</td><td>Six population areas corresponding to habitats occupied seasonally by the different age groups</td><td>Figure A.1a</td><td></td><td></td><td>Vaz et al., 2002; Uriarte et al., 1996; Motos et al., 1996</td></tr>
<tr><td>S2</td><td>Migration to spawning area</td><td>Adults (age-2+): winter</td><td>Juveniles (age-1): assume January
Adults (age-2+): either January (MigJ) or April (MigA)</td><td></td><td>iv</td><td>Uriarte et al., 1996, spatial distribution of fishing effort.</td></tr>
<tr><td>S3</td><td>Spatial distribution in spawning areas</td><td>Determined by migration coefficients</td><td>Percent of observed number-at-age in May per area each year. Average values (%) are reported below :<table><tr><th>Area</th><th>Age-1</th><th>Age-2+</th></tr><tr><td>Gironde</td><td>47</td><td>29</td></tr><tr><td>Landes coastal</td><td>11</td><td>15</td></tr><tr><td>Landes Large</td><td>8</td><td>18</td></tr><tr><td>Rochebonne</td><td>27</td><td>33</td></tr><tr><td>North</td><td>7</td><td>5</td></tr></table></td><td>10</td><td>iii</td><td>Motos et al., 1996 ;
Pelgas surveys (2000-2008)
Vaz et al., 2002; Motos et al., 1996</td></tr>
<tr><td>S4</td><td>Migration along Cantabrian coast</td><td>Supposed to occur only in years of great Anchovy abundance</td><td>Ignored</td><td></td><td></td><td>Uriarte et al., 1996</td></tr>
<tr><td>S5</td><td>Feeding area</td><td>Area “North” defined based on fishing spatial distribution assuming that fishermen follow fish migration and cover the entire anchovy distribution area</td><td>Figure A1</td><td></td><td></td><td></td></tr>
<tr><td>S6</td><td>Migration to feeding area in autumn</td><td>Adults migrate in August, recruits (1 year old) in September.</td><td></td><td></td><td></td><td>Uriarte et al., 1996; Catch analysis.</td></tr>
</table>

S7	Population spatial distribution in autumn and winter	Determined by migration coefficients	65% of age-1+ biomass is in area "North" (Figure 1)	10	ii	Evohe surveys
S8	Larval drift and juvenile concentration in coastal waters	Definition of a coastal "Recruit" area where they migrate at the age of 3 months.	Figure A1a			Allain et al., 2007
V	***Vital rates***					
V1	Growth (cm)	Function of age (monthly scale) with update each month for juveniles, each year for adults.	Von Bertalanffy growth function Linf = 18.77cm $K = 1.25\ y^{-1}$ t0 = -0.17 y	3	ii	Pelgas, 2000-2005
V2	Weight (kg)	Function of length for adult in spring and summer and juveniles all year. Function of age in autumn and winter.	$4.18 \times length^{3.2} 10^{-6}$ age-1=0.018; age-2=0.031; age-3=0.04	5	i	Pelgas, 2000-2005 Market sampling (ICES, 2000-2004)
M	***Mortality***					
M1	Egg mortality rate (day^{-1})		0.266 (**Megg1**) 0.565 (**Megg2**)	1	iv	Eggs survey, Somarakis et al., 2004; ICES 2006, 2007, 2008, 2009; Pertierra, 1997
M2	Adults mortality ($year^{-1}$)	U-shaped curve: age-2 mortality lower than age-1 and age-3	1.2 [0.5 ;3] calibrated on 2005-2008 period	3	v	ICES, 2002-2009; Chen and Watanabee, 1989
M3	Juvenile mortality ($month^{-1}$)	Function of age, monthly updated for juveniles (until the end of the first reproduction)	Exponential decay (Pareto regression) between egg mortality and adult mortality.	2	ii	Lo et al., 1995
M4	Weighting factor per area of the mortality of larvae	Average values derived from results of the IBM of larval drift and survival	Gironde=0.9; Landes coastal=1.33; Landes offshore=0.96; Rochebonne=0.97; North=0.95	5	iii	Allain et al., 2007; Hinrichsen et al., 2011

		through linear modelling.				
R	***Reproduction***					
R1	Fecundity	Function of month and dry weight (90% of fresh weight)	Apr.=200, May=500, Jun-Aug = 650	5	ii	Motos, 1996
R2	Spawning fraction		Apr=0.18, May-Aug=0.25	5	ii	Motos, 1996
R3	Maturity	All individuals mature after their first winter		17	ii	Motos, 1996
R4	Sex ratio		0.5	1	ii	Motos, 1996
R5	Reproduction function	Linear. Biomass in the last years considered low and consequently far from saturation threshold.	Number of eggs (t,area) = fecundity x spawning biomass(t, area) x spawning fraction(t, area)		ii	
R6	Spawning timing and location	Depend on area and date.	Start mid-April in Gironde, Landes coast and Landes offshore; in May in Rochebonne; in June in North.		ii	Uriarte et al., 1996 ; Allain et al., 2007
R7	Spawning duration	Depends on individual length at the beginning of the spawning season	Size at the time of spawning depends on month of birth. According to the growth function, fish of length 14cm are born in June. R1: 2 months if born after June 3 months if born before June R2: 2 months if born after May 3 months if born before May	8	ii/iv	Pertierra et al. (1997) estimated spawning to last for three months, while Motos (1996) gave a duration of 2.5 months and differentiated between fish of length smaller and larger than 14 cm.
A	***Accessibility***	Probability for a fish in a given area at a given season to be fished by a standard fishing unit with non-selective gear	Calibrated on 2000-2004 period.	3	v	

Table A.1: Population processes integrated in the model, with modelling assumptions, parameter values, attribution to categories and references.

Table

<table>
<tr><th>Fleet
(number of fishing units)</th><th>Metier
targeting anchovy</th><th></th><th>Jan</th><th>Feb</th><th>Mar</th><th>Apr</th><th>May</th><th>Jun</th><th>Jul</th><th>Aug</th><th>Sep</th><th>Oct</th><th>Nov</th><th>Dec</th></tr>
<tr><td rowspan="8">French pair trawlers Profile 1
(62)</td><td rowspan="2">Rochebonne</td><td>prop</td><td>0.22</td><td>0.08</td><td>0.03</td><td>0</td><td>0</td><td>0.33</td><td>0.25</td><td>0.15</td><td>0.18</td><td>0.09</td><td>0.19</td><td>0.03</td></tr>
<tr><td>targetF</td><td colspan="12">1</td></tr>
<tr><td rowspan="2">Gironde</td><td>prop</td><td>0.11</td><td>0.15</td><td>0.14</td><td>0</td><td>0</td><td>0.07</td><td>0.01</td><td>0.06</td><td>0.09</td><td>0.05</td><td>0.12</td><td>0</td></tr>
<tr><td>targetF</td><td colspan="12">0.98</td></tr>
<tr><td rowspan="2">Landes</td><td>prop</td><td>0</td><td>0.01</td><td>0</td><td>0</td><td>0</td><td>0.12</td><td>0.03</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>targetF</td><td colspan="12">1</td></tr>
<tr><td rowspan="2">North</td><td>prop</td><td>0.03</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0.03</td><td>0.25</td><td>0.47</td><td>0.48</td><td>0.56</td><td>0.32</td><td>0.02</td></tr>
<tr><td>targetF</td><td colspan="12">0.98</td></tr>
<tr><td rowspan="8">French pair trawlers Profile 2
(42)</td><td rowspan="2">Rochebonne</td><td>prop</td><td>0.02</td><td>0.01</td><td>0.01</td><td>0</td><td>0</td><td>0.07</td><td>0.03</td><td>0.05</td><td>0.02</td><td>0.01</td><td>0.01</td><td>0</td></tr>
<tr><td>targetF</td><td colspan="12">0.98</td></tr>
<tr><td rowspan="2">Gironde</td><td>prop</td><td>0.03</td><td>0.04</td><td>0.07</td><td>0</td><td>0</td><td>0.02</td><td>0.01</td><td>0.02</td><td>0.02</td><td>0.01</td><td>0</td><td>0</td></tr>
<tr><td>targetF</td><td colspan="12">0.96</td></tr>
<tr><td rowspan="2">Landes</td><td>prop</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0.08</td><td>0.01</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>targetF</td><td>0.71</td><td>0.42</td><td colspan="2">0.71</td><td colspan="3">0.99</td><td colspan="5">0.71</td></tr>
<tr><td rowspan="2">North</td><td>prop</td><td>0.01</td><td>0.03</td><td>0.01</td><td>0</td><td>0</td><td>0</td><td>0.07</td><td>0.07</td><td>0.15</td><td>0.08</td><td>0.04</td><td>0</td></tr>
<tr><td>targetF</td><td colspan="3">0.83</td><td colspan="3">0.86</td><td colspan="5">0.98</td><td>0.9</td></tr>
<tr><td rowspan="4">Purse seiners Basque Country
(9)</td><td rowspan="2">Landes</td><td>prop</td><td>0</td><td>0</td><td>0</td><td>0.7</td><td>0.84</td><td>0.59</td><td>0.22</td><td>0.12</td><td>0.06</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>targetF</td><td colspan="3">0</td><td>0.88</td><td>0.69</td><td>0.72</td><td>0.6</td><td colspan="2">1</td><td colspan="3">0</td></tr>
<tr><td rowspan="2">Gironde</td><td>prop</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0.02</td><td>0.03</td><td>0.01</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>targetF</td><td colspan="6">0</td><td>0.95</td><td>0.75</td><td>0.68</td><td colspan="3">0</td></tr>
<tr><td rowspan="4">Purse seiners Brittany
(17)</td><td rowspan="2">Brittany</td><td>prop</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0.03</td><td>0</td><td>0</td><td>0.16</td><td>0.5</td><td>0.41</td><td>0.2</td><td>0</td></tr>
<tr><td>targetF</td><td colspan="4">0.3</td><td>0.32</td><td colspan="2">0.33</td><td colspan="4">0.32</td><td>0.3</td></tr>
<tr><td rowspan="2">Landes</td><td>prop</td><td>0</td><td>0</td><td>0</td><td>0.19</td><td>0.12</td><td>0.02</td><td>0</td><td colspan="2">0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>targetF</td><td colspan="3">0.64</td><td colspan="2">0.79</td><td>0.46</td><td colspan="6">0.64</td></tr>
<tr><td rowspan="4">Spanish purse seiners
(222)</td><td rowspan="2">South corner</td><td>prop</td><td>0</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>targetF</td><td colspan="3">0</td><td>0.43</td><td>0.85</td><td>0.44</td><td colspan="6">0</td></tr>
<tr><td rowspan="2">Cantabria</td><td>prop</td><td>1</td><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>targetF</td><td colspan="6">0.04</td><td>0.24</td><td colspan="5">0.04</td></tr>
<tr><td rowspan="2">Spanish purse seiners - Païta
(107 vessels harvesting 2T each every two weeks as live baits)</td><td>Gironde</td><td></td><td colspan="5"></td><td colspan="2">Age 1</td><td>Age 1 /big Age 0</td><td colspan="4"></td></tr>
<tr><td>South 46°</td><td></td><td colspan="8"></td><td colspan="2">Age 0</td><td colspan="2"></td></tr>
</table>

Table A.2: The table contains all parameters relative to fleets, métiers and strategies. Number of vessels in each fleet is indicated in brackets. Métiers are designated in the table according to their area of practice and characterized by a target factor (targetF) depending on seasons which represent the intensity of search of this métier on anchovy. Target factors are classified in category (i). The proportion (prop) of fishing time spent on each métier each month belongs to category (iii) and average values are reported.

Table

Month	Br_PS	BC_PS	Prof1_PT	Prof2_PT	Sp_PS
1	58	25	68	86	15
2	42	20	95	85	11
3	26	19	121	83	129
4	20	17	109	103	206
5	37	23	63	84	381
6	37	34	91	83	245
7	68	41	81	73	84
8	69	49	84	64	54
9	55	42	71	59	62
10	19	23	52	60	71
11	26	19	48	73	84
12	17	11	14	49	40

Table A.3: Average total fishing time (hours) per boat for each fleet each month (category iii). Sp_PS: Spanish purse seiner fleet; BC_PS: Basque Country purse seiner fleet, Br_PS: Brittany purse seiner fleet, Prof1_PT: profile 1 pair trawler fleet; Prof2_PT: profile 2 pair trawler fleet

Table

Gear	Standardization factor (SF_{std})	Minimum length of catch
French purse seine	1	9cm
Spanish purse seine	0.0686	
Pair trawl	0.8	

Table A.4: Standardization factors of each gear (category ii) and minimum length of catch (category i).

Figure Appendix A1a

Click here to download high resolution image

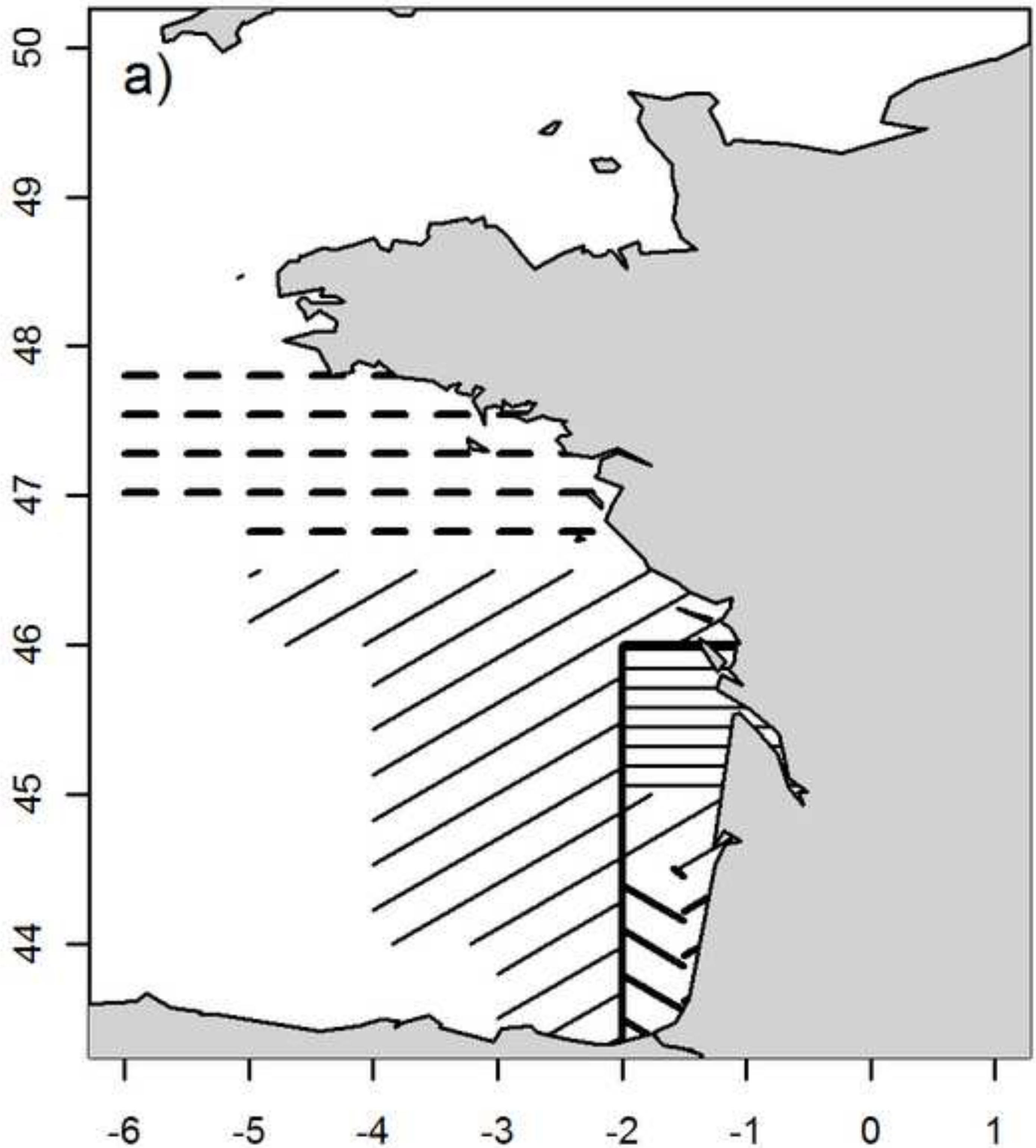


Figure Appendix A1b
Click here to download high resolution image

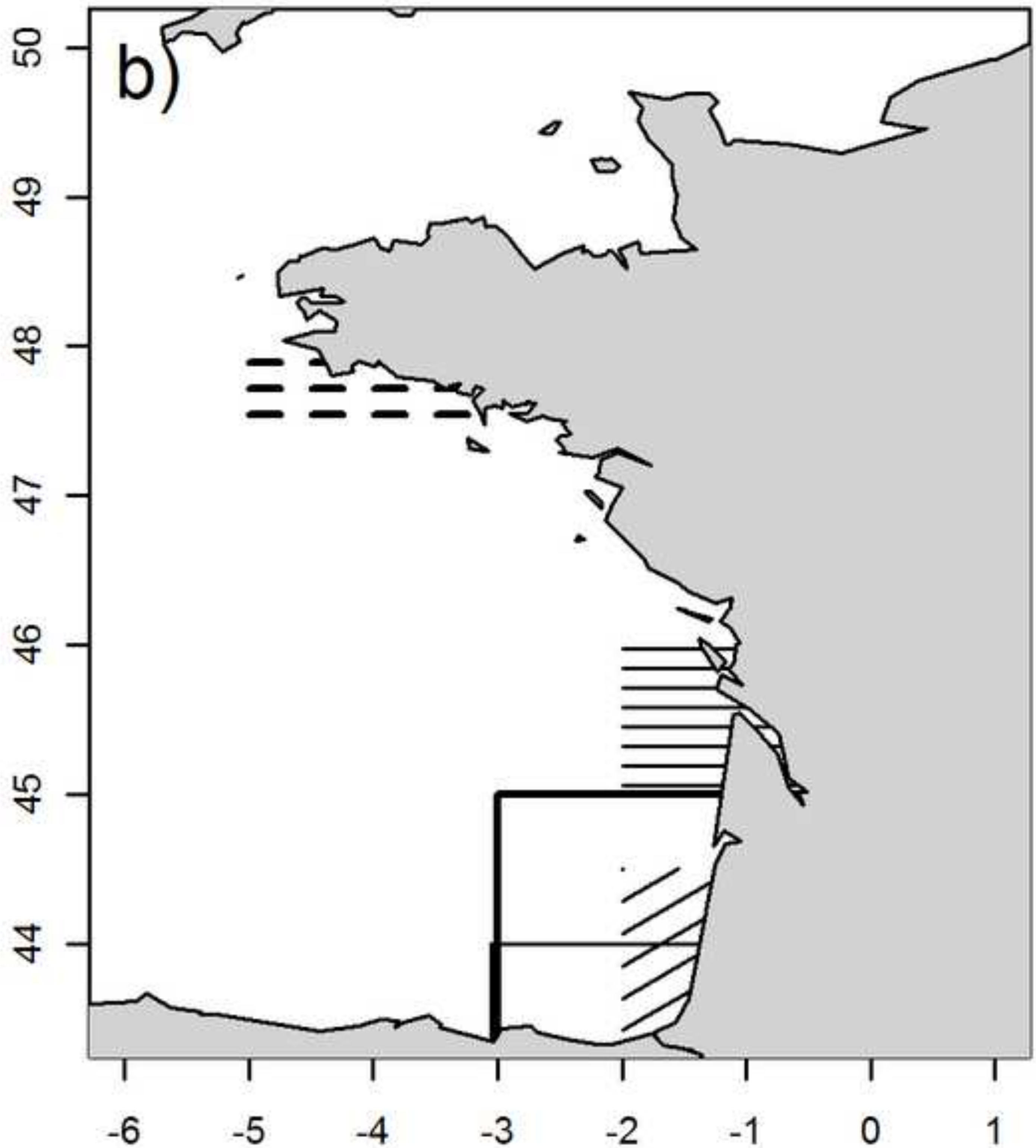
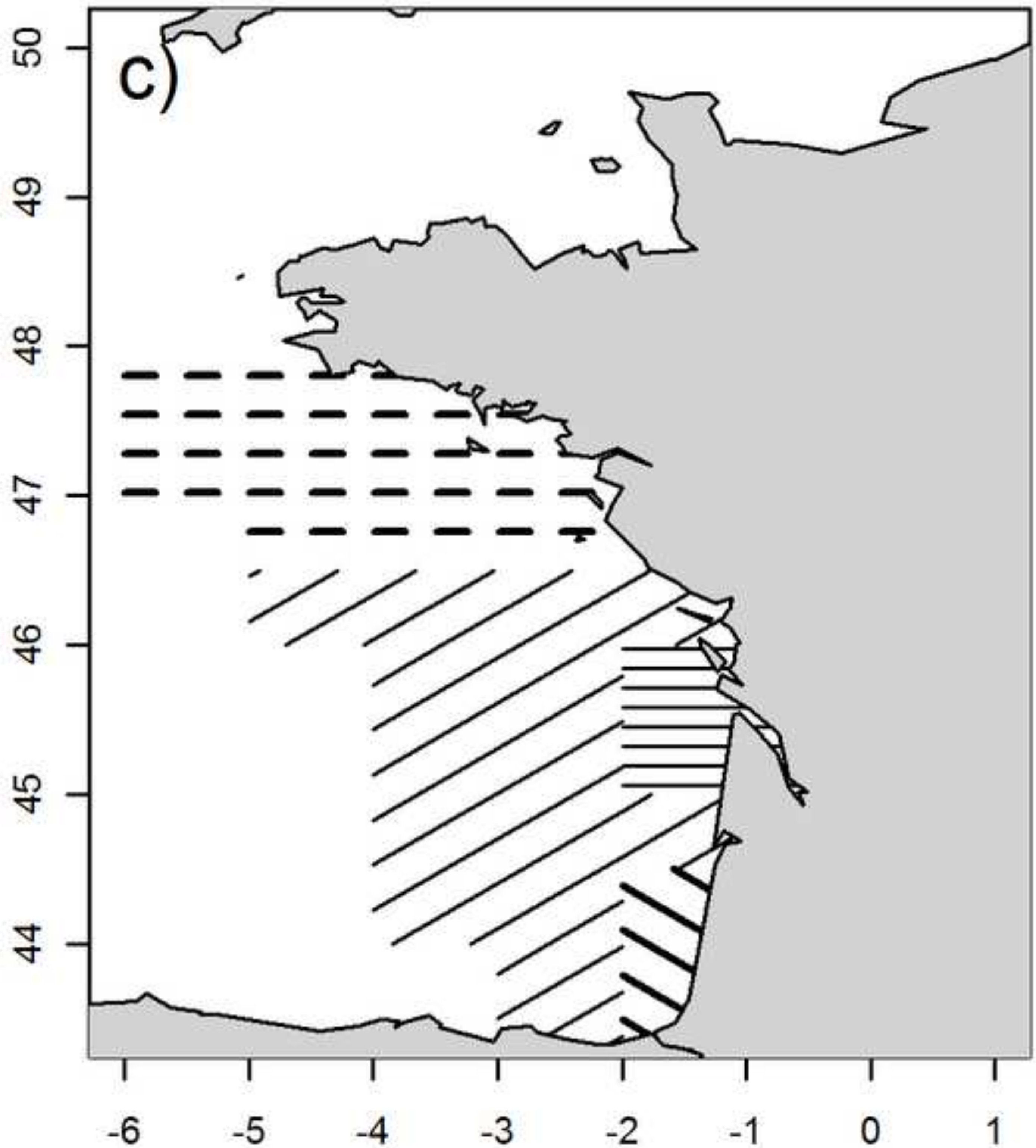


Figure Appendix A1c

Click here to download high resolution image



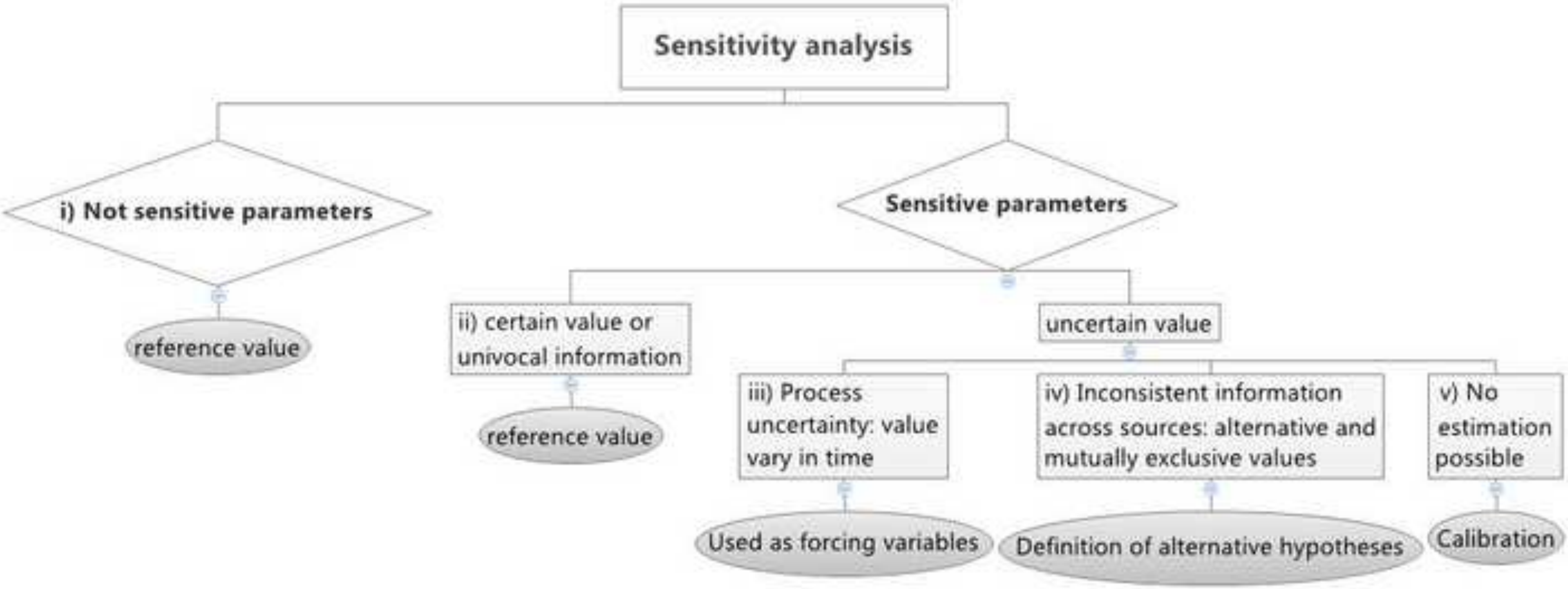

Sensitivity analysis
i) Not sensitive parameters
Sensitive parameters
reference value
ii) certain value or univocal information
uncertain value
reference value
iii) Process uncertainty: value vary in time
iv) Inconsistent information across sources: alternative and mutually exclusive values
v) No estimation possible
Used as forcing variables
Definition of alternative hypotheses
Calibration

Figure
Click here to download high resolution image

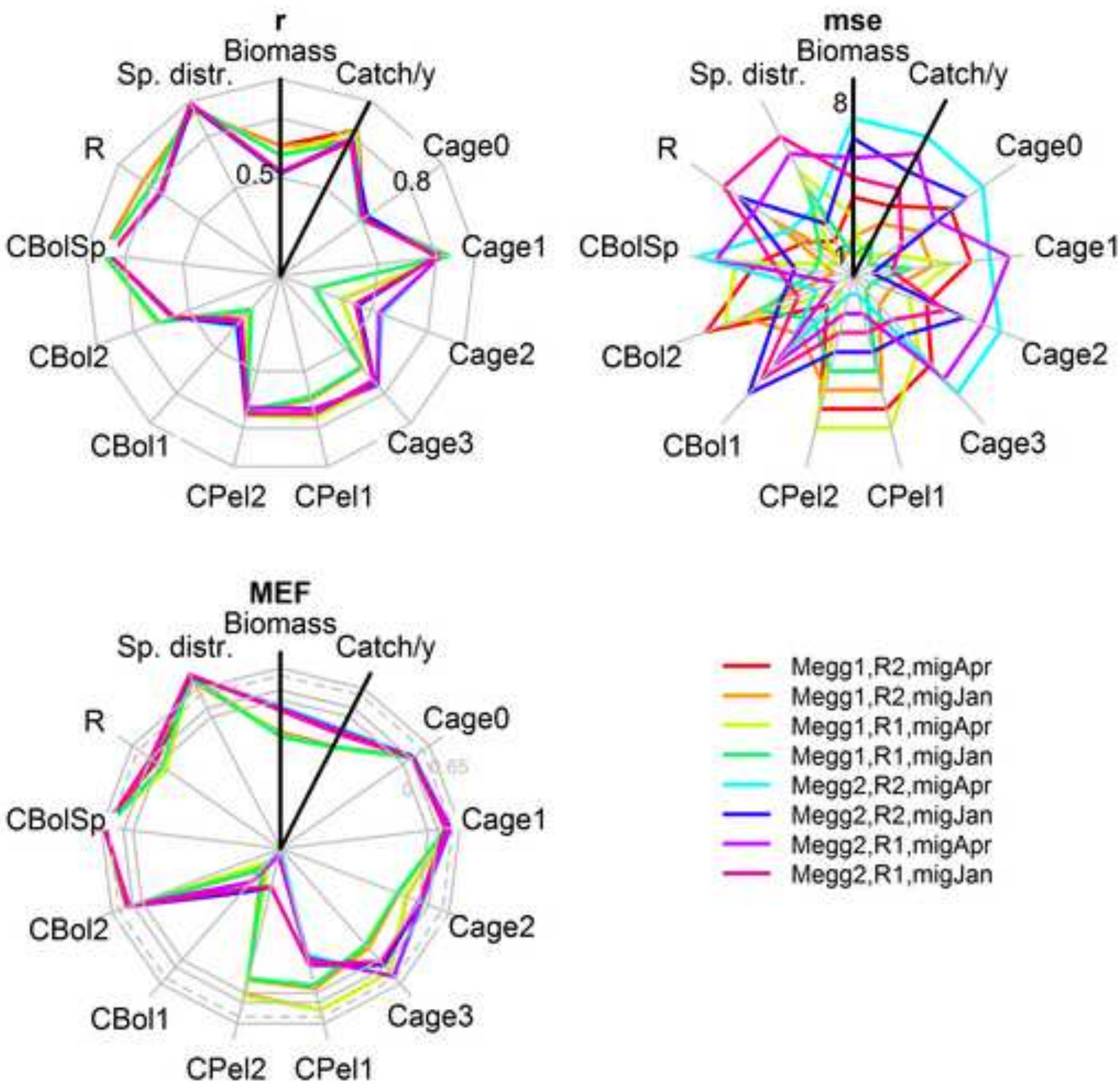


Figure
Click here to download high resolution image

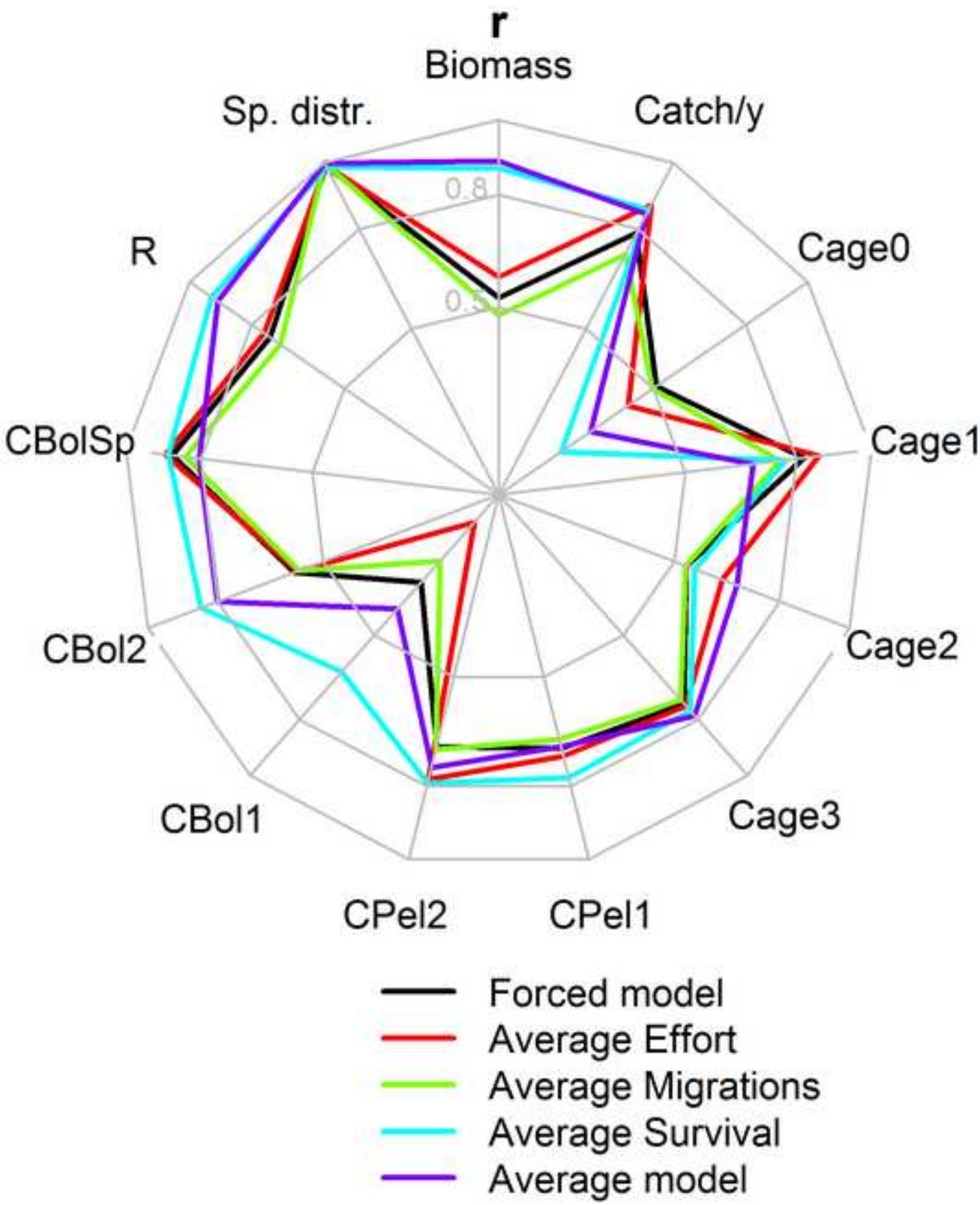


Figure
Click here to download high resolution image

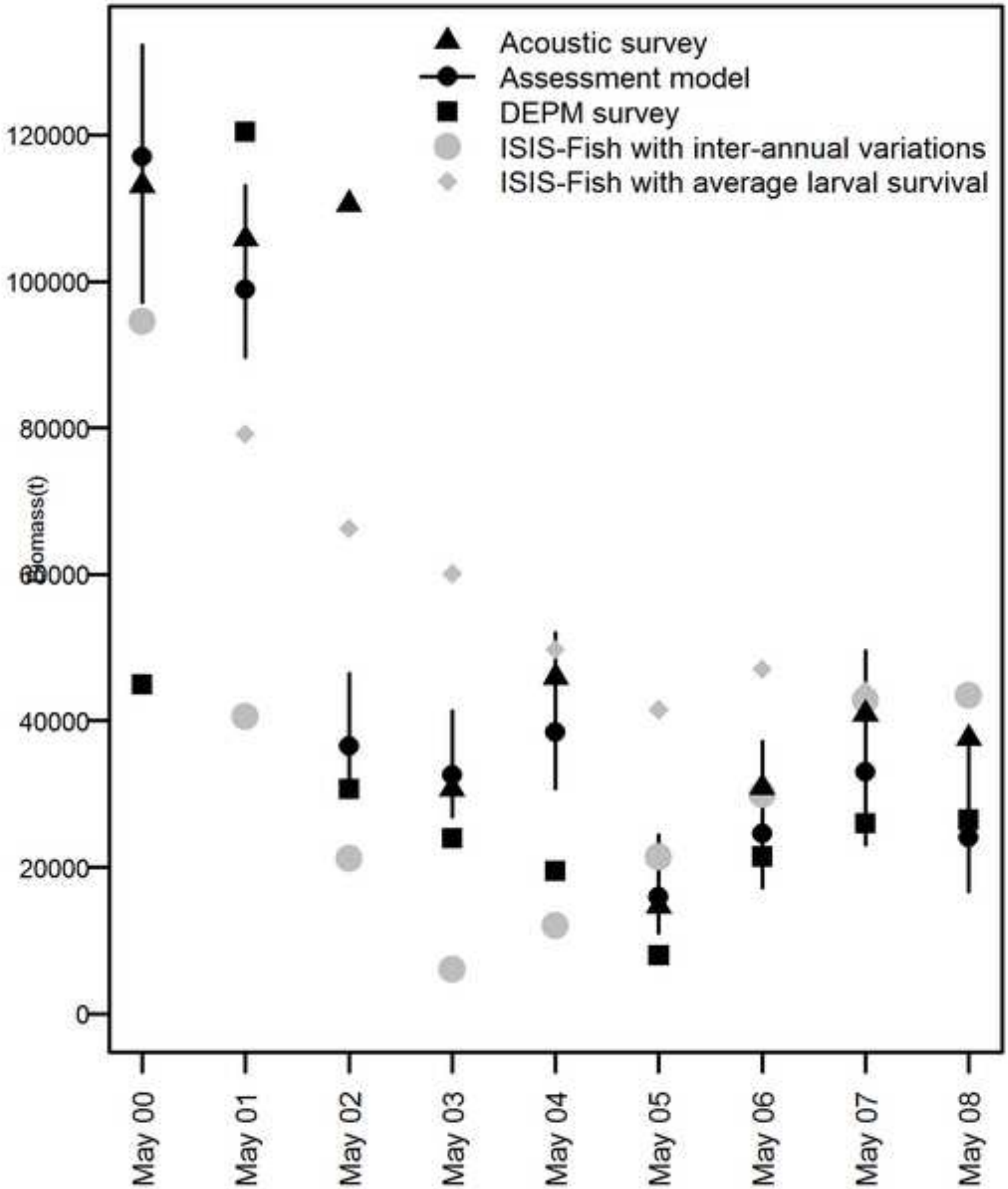


Figure 1: Parameter hierarchy: Discrimination tree to attribute parameters to categories according to their sensitivity and the level and nature of uncertainty on their value.

Figure 2: Model skill assessment: Radar plots of the scores of each alternative parameterisation (coloured lines) against each output variable (radial lines) and summary statistics with best parameterisation at the outer ends (top left: correlation value, top right: Mean squared error of the parameterisation, ranked (from 1 to 8), bottom left: model efficiency, with thin grey lines representing reference values of MEF <0.2: poor; <0.5: good; <0.65: very good; > 0.65 excellent, Allen et al., 2007). Biomass: time series of annual stock biomass 2000-2008, Catch/y: time series of annual catches 2000-2004, Cage0 (resp. Cage1, Cage2, Cage3): time series of the catches at age 0 per quarter (resp. age 1, age 2, age 3+), CPel1 (resp. CPel2, CBol1, CBol2, CBolSp): time series of monthly catches by French pair trawler profile 1 2000-2004 (resp. French pair trawlers profile 2, French purse seiners profile 1, French purse seiners profile 2, Spanish purse seiners), R: time series of annual recruitment biomass, Sp. distr.: average distribution of egg production per month over the spawning season. See text for details on statistics.

Figure 3: Forcing variables impact on correlation: The radar plot displays the correlation between observations and predictions (best at the outer end) for each simulation (coloured lines) and each output variable (radial lines). Simulations are run with all forcing variables (forced model), or with one forcing relaxed (average effort, average mortality, average migration) or with all forcing variables averaged (average model).

Figure 4: Time series of population biomass in May (2000-2008) as derived from the acoustic surveys, DEPM surveys, ICES assessment of anchovy (Assessment model) (in black) and hind-casted by ISIS-Fish with forced with larval mortality, spatial distribution and fishing effort (grey dots) or without the forcing by early survival (grey diamonds).

Appendix figures:

Figure A1: Maps of the fishery region. a) Populations areas for anchovy: wintering area (horizontal dashed lines); spawning areas: Gironde (horizontal lines), Rochebonne (upward hatched lines), Landes coast (bold upward hatched lines), Landes offshore (bold downward hatched lines); recruitment area (box). b) Purse seiners areas of practice: French fleets: Brittany (horizontal dashed lines), Gironde (horizontal lines), Landes (upward lines); Spanish fleets: South corner (thin line box), Cantabria (bold box). c) French pair trawlers areas of practice: North (horizontal dashed lines), Rochebonne (upward lines), Gironde (horizontal lines), Landes (downward bold lines).