Progress in Oceanography 151 (2017) 227–244

Contents lists available at ScienceDirect

Progress in Oceanography

journal homepage: www.elsevier.com/locate/pocean

A sequential approach to calibrate ecosystem models with multiple time series data

Ricardo Oliveros-Ramos [a,b,c,*], Philippe Verley [b,c], Vincent Echevin [d], Yunne-Jai Shin [b,c]

[a] Instituto del Mar del Perú (IMARPE), Gamarra y General Valle s/n Chucuito, Callao, Peru
[b] Institut de Recherche pour le Développement (IRD), UMR MARBEC 248, LMI ICEMASA, Place Eugène Bataillon, Bâtiment 24 RDC, CC 093, 34095 Montpellier Cedex 5, France
[c] University of Cape Town, Marine Research (MA-RE) Institute, Department of Biological Sciences, Private Bag X3, Rondebosch 7701, South Africa
[d] Laboratoire d'Océanographie et de Climatologie: Expérimentation et Analyse Numérique (LOCEAN), Institut Pierre-Simon Laplace (IPSL), UPMC/CNRS/IRD/MNHN, 4 Place Jussieu, Case 100, 75252 Paris cedex 05, France

ARTICLE INFO

Article history:
Received 20 September 2016
Received in revised form 5 January 2017
Accepted 7 January 2017
Available online 9 January 2017

Keywords:
Stochastic models
Ecosystem model
Model calibration
Model fitting
Inverse problems
Parameter estimation
Data time series
Humboldt Current Ecosystem
Peru

ABSTRACT

When models are aimed to support decision-making, their credibility is essential to consider. Model fitting to observed data is one major criterion to assess such credibility. However, due to the complexity of ecosystem models making their calibration more challenging, the scientific community has given more attention to the exploration of model behavior than to a rigorous comparison to observations. This work highlights some issues related to the comparison of complex ecosystem models to data and proposes a methodology for a sequential multi-phases calibration (or parameter estimation) of ecosystem models. We first propose two criteria to classify the parameters of a model: the model dependency and the time variability of the parameters. Then, these criteria and the availability of approximate initial estimates are used as decision rules to determine which parameters need to be estimated, and their precedence order in the sequential calibration process. The end-to-end (E2E) ecosystem model ROMS-PISCES-OSMOSE applied to the Northern Humboldt Current Ecosystem is used as an illustrative case study. The model is calibrated using an evolutionary algorithm and a likelihood approach to fit time series data of landings, abundance indices and catch at length distributions from 1992 to 2008. Testing different calibration schemes regarding the number of phases, the precedence of the parameters' estimation, and the consideration of time varying parameters, the results show that the multiple-phase calibration conducted under our criteria allowed to improve the model fit.

1. Introduction

The implementation of an ecosystem approach to fisheries not only requires a thorough understanding of the impact of fishing on ecosystem functioning and of the ecological processes involved, but also quantitative tools such as ecosystem models to provide useful information and predictions in support of management decision. Yet, the use of ecosystems models as decision making tools would only be possible if they are rigorously compared to data by means of accurate and robust parameter estimation methods and algorithms (Bartell, 2003). In many respects, the calibration of ecosystem models is a complex task. So far, minimum realism models (MRM) and models of intermediate complexity (MICE) are among the most complex multispecies models fitted to data through the optimization of an objective function and taking data uncertainty into account (Plagányi, 2007; Plagányi et al., 2014). In particular, the dynamics represented in complex ecosystem models allow species-specific parameters to have an impact on one another through ecological interactions, which may lead to correlated parameters. In addition, critical information and observations on non-commercial species can be missing or poor. The large number of parameters and the long duration of the simulations can also be an obstacle to calibrate a model. These diverse reasons hampered the development of flexible and generic calibration algorithms and methodology for ecosystem models, and only sparse documentation has been produced on fitting complex models (Bolker et al., 2013).

Given that the calibration of complex ecosystem models require large datasets and potentially involves a large number of parameters to be estimated, common practice in the field has been to (i) reduce the number of parameters to be estimated by directly using estimates from other models (Marzloff et al., 2009; Lehuta et al., 2010) or available parameters for similar species or ecosystems

* Corresponding author at: Instituto del Mar del Perú (IMARPE), Gamarra y General Valle s/n Chucuito, Callao, Peru.

E-mail address: ricardo.oliveros@gmail.com (R. Oliveros-Ramos).

http://dx.doi.org/10.1016/j.pocean.2017.01.002
0079-6611/

(Bundy, 2005; Ruiz and Wolff, 2011), (ii) use outputs from other models as data to calibrate the model (Mackinson and Daskalov, 2007), or (iii) use both strategies (Shannon et al., 2003; Guénette et al., 2008; Friska et al., 2011; Travers-Trolet et al., 2014). These different strategies expedite the calibration of complex models while attempting to synthesize the maximum of available information. However, since the borrowed parameters and outputs rely on different model assumptions, they may lead to inaccuracy and inconsistency in parameter estimation by trying to reproduce other models' dynamics.

The weaknesses of these common practices can be overcome by implementing a multiple-phase calibration approach (Nash and Walker-Smith, 1987; Fournier et al., 2012). In this multiple-phase approach, some parameters can be fixed at initial values obtained from independent data, other models or expertise (Nash and Walker-Smith, 1987). In particular, assigning initial guesses for completely unknown parameters before proceeding to a full calibration of all parameters can ease the estimation of model parameters (Nash and Walker-Smith, 1987; Fournier et al., 2012). This multiple-phase calibration approach is supported by some optimization softwares, like specialized R packages or the AD Model Builder software (Bolker et al., 2013). However, it is difficult to find in the literature a clear roadmap or strategy to guide the users and help them to determine what parameters should be estimated in the successive phases. It appears that the final organization of the calibration phases is most often an empirical process and is the result of trials and errors in the calibration procedure (Fournier, 2013).

The objective of this paper is to highlight some issues related to the comparison of complex ecosystem models to data and propose a methodology to a sequential calibration of ecosystem models, illustrating it with the calibration of the ecosystem model OSMOSE (Shin and Cury, 2004; Travers et al., 2009) applied to the Northern Humboldt Current Ecosystem. The first important step in a calibration is to be able to categorize the parameters of a model. To do so, we propose two criteria: the model dependency and the time variability of the parameters. Then, we use these criteria and the availability of initial guesses of the parameters to determine which parameters need to be estimated, and their precedence in the sequential calibration process. We finally compare our sequential approach with the results of a single step calibration of all parameters.

2. Material and methods

2.1. Parameterization and calibration

Several classifications of model parameters can be found in the literature (e.g. Jorgensen and Bendoricchio, 2001) according to different criteria and for different purposes. In this work, we classified the parameters according to two criteria: (1) the dependence of the parameter on the model structural assumptions, and (2) the time variability of the parameter in relation to its use in the model. The categorization of the parameters is defined as follows.

Model dependency: Parameters are considered to be model-dependent when their values can vary between models due to different model structures or assumptions. For example, fishing mortality can be categorized as being model-dependent, because it depends on the value of natural mortality, structural equations of the fishing process and assumptions on the selectivity or seasonal distribution of fishing effort. On the contrary, model-independent parameters can be estimated directly from data and observations by simple models or theoretical relationships. For example, parameters for the length-weight relationships or for the von Bertalanffy growth function can be considered independent of the overarching ecosystem model structure and assumptions.

Time variability: Some parameters of the model are expected to have temporal variability at the time scale of the model and the data. For example, fish larval mortality rates that determine the fish annual recruitment success and which are related to environmental conditions are expected to vary annually. Other parameters of an ecosystem model are not expected to have significant temporal variability at the time scale of the model and the data time series, for example the parameters of predators' functional response. It is also important to notice that time-varying parameters for one model may be outputs from a process submodel in another model.

The classification of the parameters in terms of model dependency is necessary in order to avoid the misleading use of parameters' values which have been estimated in other models and not directly from observations. If some parameters are fixed at values inconsistent with the model structure currently used to fit the data, the estimates of other parameters obtained from the calibration can be highly uncertain and only artifacts to fit the data. This can also impede the convergence of the objective function and lead to a calibration failure (Gaume et al., 1998; Whitley et al., 2013). Additionally, this practice leads to an underestimation of the model uncertainty by assuming some parameters to be known when they are not.

The classification in terms of temporal variability can be more arbitrary since many parameters (especially the ones characterizing the populations) are expected to vary with time. The cutoff we propose for a parameter to be considered as time-varying results from the following considerations: (i) the identification of a process leading to such time variability, (ii) the existence of theoretical assumptions about the importance of such process in the dynamics of the modeled ecosystem, (iii) the non-explicit representation of the process in the model, and (iv) the significance of the time variability compared to the time scale of the model and the length of the data time series. Some parameters can be assumed to be constant at shorter time scales (e.g. a few years) but can exhibit variability at longer time scales (e.g. several decades). For example, the length at maturity for a given species can decrease in response to heavy fishing (Shin et al., 2005), but can be considered as constant in the model for periods short enough, or if the variability is not considered to cause significant changes in the functioning of the multispecies assemblage.

Despite the apparent dichotomous classification presented, the degree of model-dependency and temporal variability in the parameters can vary, and a qualitative classification of the parameters should be attempted. In the OSMOSE ecosystem model, such classification could be proposed for the parameters characterizing modeled multispecies fish assemblages (Fig. 1; see Appendix A for details about the parameterization of OSMOSE).

2.2. Approach for the sequential calibration

2.2.1. Progressive time resolution of the parameters

The number of parameters to be estimated in a model can be high, particularly when time-varying parameters are included, so that fitting the model to data can be challenging (e.g. see Schnute, 1994). Additionally, the way a model is parameterized will define the objective function to be optimized to estimate the parameters; just by rescaling or transforming the parameters this objective function can be changed and the parameter estimation process can be improved (Bolker et al., 2013).

There are several ways to model the time variability in the parameters, taking into account the assumed shape of the variability and the degree of time resolution one wants to introduce (see Megrey, 1989; Methot and Wetzel, 2013 for examples in fishery models). However, in practical terms, there is a limit in the number of parameters that can be estimated, which depends on the quality