Ecological Modelling 306 (2015) 278–286

Contents lists available at ScienceDirect

Ecological Modelling

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

Approximate Bayesian computation to recalibrate individual-based models with population data: Illustration with a forest simulation model

Guillaume Lagarrigues [a,*], Franck Jabot [b], Valentine Lafond [a], Benoit Courbaud [a]

[a] Irstea, Ecosystèmes Montagnards, 2 rue de la Papeterie, BP 76, 38402 Saint-Martin-d'Hères Cedex, France
[b] Irstea, Laboratoire d'Ingénierie pour les Systèmes Complexes, 9 avenue Blaise Pascal, CS 20085, 63178 Aubière, France

ARTICLE INFO

Article history:
Available online 13 October 2014

Keywords:
Approximate Bayesian computation
Inverse modeling
Forest dynamics
Sensitivity analysis
Uneven-aged forests
Model calibration

ABSTRACT

Ecology makes an increasing use of complex simulation models. As more processes and model parameters are added, a comprehensive model calibration with process-level data becomes costly and predictions of such complex models are therefore often restricted to local applications. In this context, inverse modelling techniques enable to calibrate models with data of the same type than model outputs (thereafter called population data for the sake of clarity, although other data types can be used according to model outputs), which are usually simpler to collect and more readily available. This study aims at demonstrating how such data can be used to improve ecological models, by recalibrating the most influential parameters of a complex model in a Bayesian framework, and at providing general guidelines for potential users of this approach. We used the individual-based and spatially explicit forest dynamics simulation model Samsara2 as a case study. Considering the results of an initial calibration and of a sensitivity analysis as prerequisites, we assessed whether we could use approximate Bayesian computation (ABC) to recalibrate a subset of parameters on historical management data collected in forests with various ecological conditions. We propose guidelines to answer three questions that potential users of the approach will encounter: (1) How many and which parameters are we able to recalibrate accurately with such low-informative data? (2) How many ABC simulations are required to obtain a reasonable convergence of the parameter posterior estimates? (3) What is the variability of model predictions following the recalibration? In our case study, we found that two parameters by species could be recalibrated with forest management data and that a relatively low number of simulations (20,000) was sufficient. We finally pointed out that the variability of model predictions was largely due to model stochasticity, and much less to ABC recalibration and initial calibration uncertainties. Combining direct process-level calibration to ABC recalibration of the most influential parameters opens the door to interesting modelling improvements, such as the calibration of forest dynamics along environmental gradients. This general approach should thus help improve both accuracy and generality of model-based ecological predictions.

1. Introduction

Ecology makes an increasing use of complex simulation models, like individual-based models (Grimm, 1999), physiological models (Deckmyn et al., 2008) or dynamic global vegetation models (Scheiter et al., 2013). Detailed simulation models integrate available knowledge in a consistent framework, and aim at producing quantitative predictions on specific systems (Evans et al., 2013). For that purpose, mechanisms are added in models at different organization levels, such as physiological processes (Deckmyn et al., 2008), variability among individuals (Vieilledent et al., 2010), spatio-temporal processes (Balzter et al., 1998) or feedbacks between plant growth and abiotic factors like soil resources (Raynaud and Leadley, 2004) or light conditions (Courbaud et al., 2003). All these features are thought to improve the prediction quality of ecological models (Evans et al., 2013).

The inclusion of ecological details, however, comes at a price: as more processes and model parameters are added, models become themselves complex systems whose behaviour is difficult to understand and which are challenging to calibrate with data. This trade-off between model complexity and calibration accuracy is well-known by statisticians (Burnham and Anderson, 2002). The

* Corresponding author. Tel.: +33 4 76 76 27 07.
E-mail address: guillaume.lagarrigues@irstea.fr (G. Lagarrigues).

http://dx.doi.org/10.1016/j.ecolmodel.2014.09.023

typical data scarcity in ecological systems makes this calibration problem particularly acute. Two general types of data commonly used for model calibration can be distinguished. First, process-level data, like individual resource acquisition, growth, fecundity, mortality or local environmental monitoring provide direct information on the modelled processes. However, because such detailed data are typically scarce, model calibration is often performed using data from unconnected and local studies, making the joint calibration of the different processes hardly feasible. In contrast, data of the same type than model outputs, like canopy cover for forest dynamics model or population data for individual-based models, are more often available but do not provide direct information on process parameters. This type of data (there after simply called population data for the sake of clarity, although other data types can be used according to model outputs) requires inverse modelling techniques to infer the ecological processes (Grimm and Railsback, 2005; Hartig et al., 2011).

In this study, we used a calibration strategy in which both process-level and population data were used. This strategy relies on three steps: (1) carry out an initial calibration with process-level data, (2) perform a sensitivity analysis to identify the influential model parameters and (3) recalibrate the most influential parameters with population data. We used an individual-based and spatially explicit model of forest dynamics, called Samsara2, in which the processes of light interception, growth, reproduction and mortality are simulated at the individual tree level (Supplementary data, Appendix A). This model relies on a detailed representation of light interception by tree crowns, taking into account both stand slope and exposition (Courbaud et al., 2003). The steps (1) and (2) have been carried out prior to this study: Samsara2 was initially calibrated in a restricted area, using detailed process-level data mainly collected in the Northern Alps (see Section 2.2); the sensitivity analysis was also detailed in (Supplementary data, Appendix B). After a quick reminder of the results of these two steps, we will detail the third one and we will propose general guidelines to tackle three issues that we encountered while performing this recalibration task. In what follows, we present these issues.

The first difficulty arose along with the large number of input parameters which commonly characterize complex models such as Samsara2. In inverse modelling, modellers usually calibrate all model parameters jointly, so that model predictions fit as well as possible to observed data (Jabot and Bascompte, 2012; Van Oijen et al., 2005; Vrugt and Sadegh, 2013; Wiegand et al., 2003). However, when the model has a large number of interacting input parameters, this method leads inevitably to a high uncertainty on parameter estimates (Wiegand et al., 2003), or to marginal posterior distributions close to prior distributions when using a Bayesian framework (Hartig et al., 2014; Jabot and Bascompte, 2012), at least for some parameters. If the focus is on fitting well the model to data (with the perspective of a local adaptation of the model, for example), that is not really problematic. But the finality of the recalibration can also be model improvement (Dong et al., 2008; Fenicia et al., 2008). For instance, one may be interested in how model parameters vary when fitted in different study sites or at different dates (Foll and Gaggiotti, 2006; Jabot et al., 2008). In such cases, reducing uncertainty in parameter estimates might become a central objective. Our study falls within this scope since we are interested in calibrating tree demographic processes along environmental gradients using population data from several managed mountain forests. The first issue was therefore to determine how many and which parameters could be accurately recalibrated.

The second point is linked to the computing time, which is commonly substantial in complex models. As inverse modelling methods are based on the repetition of a large number of simulations, modellers are often reluctant to use them with time-consuming models and most of applications have been applied to fast-running models (Beaumont, 2010; Jabot and Bascompte, 2012; Vrugt and Sadegh, 2013), although the increasing of computing power now allows calibrating more complex models with such methods (Hartig et al., 2014; Lenormand et al., 2013; Van Oijen et al., 2005). Still, the computing time can be a limiting factor when applying such a calibration process on a large number of data sets. In the recalibration context presented previously, the number of simulations may be decreased as we intended to calibrate only some model parameters, especially as we had prior knowledge on parameters coming from initial calibration. The issue was then to determine a minimum number of simulations needed to accurately recalibrate a subset of parameters, while preserving the usability of the method with a large amount of data.

The third issue was to assess the efficiency of the recalibration. For that purpose, we aimed at comparing the effects of the different sources of uncertainty on the variability of model predictions, including the uncertainties on the recalibrated parameters resulting from the recalibration task. Here, we had to deal with model stochasticity, which is also a common feature in complex ecological models that makes statistical inference particularly challenging (Hartig et al., 2011, 2014). We propose a method to quantify the contributions of each source of uncertainty to variability of predictions, so as to disentangle the effects of model stochasticity from those of parameter uncertainties in this specific context where only some parameters have been recalibrated.

2. Material and methods

2.1. Data

In this study, we used historical management data collected in the Prenovel forest (46.5°N–5.8°E, altitude ranging from 920 m to 1030 m, mean annual temperature of 7.1 °C, annual precipitation of 1700 mm per year), located in the Jura massif. This forest is a public forest which has been managed for more than one century as a selection forest system based on continuous natural regeneration and simultaneous presence of uneven-aged trees. The Prenovel forest is divided into smaller management units, thereafter called stands, with a surface area ranging from 5 to 20 ha. In this study, we applied our recalibration method on the stand *n*° 24.

For this forest, management plans are renewed each 20 years, at so-called inventory years. In these inventories, all trees with a diameter at breast height (dbh) larger than the commercial size (17.5 cm) are recorded, identifying their species and ranking them in 5 cm-wide diameter classes. Harvests are recorded in yearly reports that indicate the volume and number of trees harvested, grouping species in only two categories: resinous or broadleaves. Inventories and harvest reports have been recorded at the stand level and archived for several decades by the French National Forest Office (ONF). Historical management data are available from 1953 to 2011 for this forest. Silver fir (*Abies alba* Mill.) and Norway spruce (*Picea abies* (L.) Karst.) are dominant species in this forest, representing together 90% of the total basal area. Climate conditions are comparable to those of the Queige forest (Northern Alps, 45.7°N–6.5°E, altitude ranging from 1100 m to 1550 m, mean annual temperature of 5 °C, annual precipitation of 1700 mm per year), for which the initial calibration of growth and allometric equations has been performed. However, mean temperatures are slightly warmer and soil fertility is overall larger in Prenovel than in Queige, resulting in larger growth rates and tree size, particularly for silver fir.