

This discussion paper is/has been under review for the journal Biogeosciences (BG).
Please refer to the corresponding final paper in BG if available.

Technical Note: Approximate Bayesian parameterization of a complex tropical forest model

F. Hartig^{1,2}, C. Dislich^{1,3}, T. Wiegand¹, and A. Huth¹

¹UFZ – Helmholtz Centre for Environmental Research, Department of Ecological Modelling, Permoserstr. 15, 04318 Leipzig, Germany

²University of Freiburg, Department of Biometry and Environmental System Analysis, Tennenbacher Str. 4, 79085 Freiburg, Germany

³University of Göttingen, Department of Ecosystem Modelling, Büsgenweg 4, 37077 Göttingen, Germany

Received: 28 May 2013 – Accepted: 9 July 2013 – Published: 7 August 2013

Correspondence to: F. Hartig (florian.hartig@biom.uni-freiburg.de)

Published by Copernicus Publications on behalf of the European Geosciences Union.

BGD

10, 13097–13128, 2013

**Approximate
bayesian calibration
of a forest model**

F. Hartig et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Abstract

Inverse parameter estimation of process-based models is a long-standing problem in ecology and evolution. A key problem of inverse parameter estimation is to define a metric that quantifies how well model predictions fit to the data. Such a metric can be expressed by general cost or objective functions, but statistical inversion approaches are based on a particular metric, the probability of observing the data given the model, known as the likelihood.

Deriving likelihoods for dynamic models requires making assumptions about the probability for observations to deviate from mean model predictions. For technical reasons, these assumptions are usually derived without explicit consideration of the processes in the simulation. Only in recent years have new methods become available that allow generating likelihoods directly from stochastic simulations. Previous applications of these approximate Bayesian methods have concentrated on relatively simple models. Here, we report on the application of a simulation-based likelihood approximation for FORMIND, a parameter-rich individual-based model of tropical forest dynamics.

We show that approximate Bayesian inference, based on a parametric likelihood approximation placed in a conventional MCMC, performs well in retrieving known parameter values from virtual field data generated by the forest model. We analyze the results of the parameter estimation, examine the sensitivity towards the choice and aggregation of model outputs and observed data (summary statistics), and show results from using this method to fit the FORMIND model to field data from an Ecuadorian tropical forest. Finally, we discuss differences of this approach to Approximate Bayesian Computing (ABC), another commonly used method to generate simulation-based likelihood approximations.

Our results demonstrate that simulation-based inference, which offers considerable conceptual advantages over more traditional methods for inverse parameter estimation, can successfully be applied to process-based models of high complexity. The methodology is particularly suited to heterogeneous and complex data structures and

BGD

10, 13097–13128, 2013

Approximate bayesian calibration of a forest model

F. Hartig et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



can easily be adjusted to other model types, including most stochastic population and individual-based models. Our study therefore provides a blueprint for a fairly general approach to parameter estimation of stochastic process-based models in ecology and evolution.

1 Introduction

Parameter estimation of process-based models is a long-standing problem in ecology and evolution. Early proponents of process-based modeling have stressed the importance of deriving predictions from basic physical processes, with physical parameters that can be experimentally determined (Bossel, 1992). In practice, however, various reasons, ranging from time limitations to fundamental observability restrictions, result in the situation that most process-based models include parameters for which direct measurements are not available (Hartig et al., 2012). These parameters need to be estimated inversely, meaning that they are adjusted by comparing model outputs to observed data.

To make this comparison, Bayesian methods have become increasingly popular in ecological research during the last decades (e.g. O'Hara et al., 2002; Clark, 2005; Purves et al., 2007; Higgins et al., 2012). Additional to their flexibility and their explicit treatment of parameter uncertainty, a particularly appealing property of Bayesian statistics is the possibility to combine existing information on likely parameters values with the information that is generated inversely (Hartig et al., 2012). As other inverse parameterization approaches, Bayesian methods require the definition of a metric that measures how well model predictions fit to the observed data. In non-statistical inversion approaches, such metrics are often called goal functions, objective functions or cost functions (e.g. Schröder and Seppelt, 2006). Bayesian approaches use a particular statistical metric, the probability of obtaining the observed data given the current model and parameter values, usually referred to as the likelihood. Most previous applications of Bayesian statistics to complex stochastic models derive this probability by

Approximate bayesian calibration of a forest model

F. Hartig et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



BGD

10, 13097–13128, 2013

Approximate bayesian calibration of a forest model

F. Hartig et al.

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)

making distributional assumptions about how observations vary around mean model predictions that are independent of the processes in the model, either ad-hoc or based on the observed variance in the data (e.g. Martínez et al., 2011; van Oijen et al., 2013). This is usually justified with the idea in mind that there is observation uncertainty or variability in environmental conditions that is not accounted for in the model. The approach of constructing likelihoods from such assumptions is the current state-of-the-art, but it has a major limitation: many process-based ecological models are already stochastic and predict variability of certain model outputs. Additional process-based observer models that describe how field data was collected can easily be added (Zurell et al., 2009). In principle, one would therefore much rather use this mechanistically derived variability for deriving the likelihood, because this accounts for our ecological understanding in creating expectations for the probabilities of observing particular deviations from mean model predictions. However, while theoretically always possible, this route was blocked in practice for most models due to the difficulties of making the necessary probability calculations technically tractable (Hartig et al., 2011).

This technical limitation has weakened in recent years. Simulation-based approximation techniques have been developed that allow treating any stochastic model in a formal statistical inference framework. Of those, Approximate Bayesian Computing (Beaumont, 2010) has arguably attracted most attention, but there are other approaches as well. Their common principle is very simple: what is needed for including a stochastic simulation model in a formal inferential framework is the likelihood $p(D|M(\phi))$ for an outcome D to occur under a model M with parameters ϕ (Diggle and Gratton, 1984). Simulation-based likelihood approximations estimate this probability by generating draws from the stochastic model. Subsequently, different methods are used to approximate the likelihood or posterior (Hartig et al., 2011). Often, this involves comparing model output and observed data by means of data aggregations, also called patterns (Wiegand et al., 2004; Grimm and Railsback, 2012) or summary statistics (Beaumont, 2010; Wood, 2010). For brevity, we will refer to these methods in gen-



BGD

10, 13097–13128, 2013

**Approximate
bayesian calibration
of a forest model**

F. Hartig et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

ent growth properties and light demands. For highly diverse systems such as tropical rainforests exhibit, species are usually grouped into plant functional types (PFTs) that represent a group of species with similar functional properties. Parameters and model predictions per plant functional types then represent a mean over the species that are represented by this type. Gap formation, that is, the death of large trees, maintains the modeled forest in a dynamic equilibrium. As a result, forest gap models do not merely predict a mean value for outputs such as biomass, species composition, or tree size distributions. Rather, they deliver samples of different possible values for these outputs and therefore allow assigning probabilities to different community or biomass states. These predictions of spatio-temporal variation in community composition is what we will use later to derive a probabilistic distance measure between model output and observed data.

FORMIND, the forest model used for this study, is a stochastic, individual-based forest model designed in the tradition of classical forest gap models (Köhler, 2000). It has been applied for estimating forest succession, variability and disturbances impacts in various tropical locations around the world (e.g. Rüger et al., 2007; Köhler and Huth, 2010; Dislich and Huth, 2012; Gutiérrez and Huth, 2012). The simulation area (plot) in FORMIND, which can be of variable size (we use 1 ha throughout the paper) is subdivided into 20 m × 20 m grid cells. Tree individuals are assigned to one of these cells and interact with each other on this cell, but do not have an explicit spatial position within the cells. The model state is entirely described by species respectively functional type, size (measured in diameter at breast height dbh), and location (cell) of all trees. Other variables, such as tree height and crown dimensions, are derived through fixed allometric relationships.

At each time step (we use 5 yr time steps), the light climate for each tree is calculated from the position of all trees and their respective crowns in the same cell. Subsequently, establishment (light-dependent, stochastic), mortality (stochastic) and tree growth (light-dependent) act on the plot, respectively the tree individuals. Important parameters in the model (Table 1) are recruitment and mortality rates, parameters

that describe the size-specific maximum growth rates, and the allometric relationships that determine height and crown dimensions. The model is Markov in the sense that the probabilities for reaching any possible state in the next time step are entirely determined by the values of the state variables in the previous time step. Details of these processes, together with a more exact description of the model scheduling, are provided in the online Supplement (see also Köhler, 2000; Dislich et al., 2009).

2.2 Bayesian parameter estimation with simulation-based likelihood approximations

We use a Bayesian approach for parameter estimation. Hartig et al. (2011) recommend using Bayesian methods (or at least MCMCs) with simulation-based likelihood approximations, because MCMCs (unlike optimization approaches) are more robust towards variance in likelihood estimates generated by the approximation. They are also less sensitive to interactions between parameters that are to be expected in process-based models. In principle, however, one could use the likelihood-approximation used in this study with an optimization algorithm in a maximum-likelihood framework as well.

There are a number of introductions to Bayesian statistics. A detailed reference is Gelman et al. (2003), for shorter introductions see Ellison (2004). We therefore give only a brief summary here. The outcome of a Bayesian inference is a probability distribution $P(\boldsymbol{\phi}|\mathbf{D}_{\text{obs}})$ for the parameters $\boldsymbol{\phi}$ given the observed data \mathbf{D}_{obs} . This distribution, called the posterior, is calculated as

$$p(\boldsymbol{\phi}|\mathbf{D}_{\text{obs}}) = c \cdot p(\mathbf{D}_{\text{obs}}|\boldsymbol{\phi}) \cdot p(\boldsymbol{\phi}), \quad (1)$$

where c is a normalization constant, the prior probability density $p(\boldsymbol{\phi})$ quantifies parameter uncertainties before comparing the model to the observed data, and the likelihood function $p(\mathbf{D}|\boldsymbol{\phi})$ describes the probability of obtaining the observed data conditional on a parameter set $\boldsymbol{\phi}$, which can, broadly speaking, be interpreted as a measure of fit.

Because our main concern in this paper is the approximation of the likelihood, we chose wide uniform (flat) priors for all parameters and fit types, which means that the

BGD

10, 13097–13128, 2013

Approximate bayesian calibration of a forest model

F. Hartig et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



posterior and likelihood are strictly proportional to each other across the possible prior range. Tables with the widths of these uniform priors are provided in the online Supplement. Given that we knew that the model reacts nonlinearly to many parameters, other uninformative priors choices would have been possible (e.g. Kass and Wasserman, 1996), but we felt that for this study it is more useful to ensure proportionality of likelihood and posterior to make the results better interpretable.

2.2.1 Generating approximate likelihoods

The technical key novelty in this study is the definition of the likelihood $p(D|\phi)$. In “conventional” Bayesian or maximum likelihood studies, this conditional probability is obtained by formulating an error model that quantifies probabilities of deviations between model predictions and observations to occur (e.g. Van Oijen et al., 2005). This model may be mechanistically motivated, for example by knowledge about measurement uncertainties during obtaining the data. In practical situations, however, there are usually a number of error sources that interact, and error models are therefore typically either fixed ad hoc (van Oijen et al., 2013) or derived from the observed variability in the data (Martínez et al., 2011). Hence, “conventional” likelihoods are usually independent of the mechanisms in the process-model that is fit.

The method that we apply here goes beyond such an independent error model towards an approach where both the mean model prediction, but also the probability of observing deviations from the mean, are derived from the same stochastic ecological processes. This is particularly promising in systems where process-stochasticity dominates observation errors. For many tropical forest inventories, this is the case – typical observation errors are known (Chave et al., 2004), and based on these results we can assume that, for small plots (1 ha), observation uncertainty is small compared to the local biomass variation due to successional dynamics (Chave et al., 2003). FORMIND simulations of the aforementioned successional dynamics triggered by gap formation explain the extent of this variability well (e.g. Köhler and Huth, 2010) and can therefore

BGD

10, 13097–13128, 2013

Approximate bayesian calibration of a forest model

F. Hartig et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



be used to generate statistical expectations for field observations such as, for example, plot biomass (Fig. 1).

Several techniques have been suggested for achieving approximations of the likelihood $p(\mathbf{D}_{\text{obs}}|\boldsymbol{\phi})$ from the variability of the stochastic simulation outputs. Most prominent is arguably the method of Approximate Bayesian Computing (ABC) (Csilléry et al., 2010; Beaumont, 2010), which has attracted much attention in recent years. However, as discussed in Hartig et al. (2011), there are a number of closely related methods that are currently not counted as examples of ABC, but that apply nearly the same principles. In this study, we apply a simulation-based likelihood approximation used by (Wood, 2010), called a “parametric likelihood approximation” in Hartig et al. (2011).

The principle of this method is to estimate $p(\mathbf{D}_{\text{obs}}|\boldsymbol{\phi})$, for any $\boldsymbol{\phi}$ desired, by fitting a parametric distribution to the output of the stochastic simulation, and estimating the probability of obtaining \mathbf{D}_{obs} from this distribution (Fig. 2). We used a multivariate normal distribution because it fitted well to the simulation outputs, and it allows a convenient estimation of the covariance structure, but normality is by no means a fundamental requirement. For this multivariate normal approximation, the likelihood of obtaining the observed data \mathbf{D}_{obs} is

$$p(\mathbf{D}_{\text{obs}}|\boldsymbol{\phi}) \approx c \cdot |\boldsymbol{\Sigma}_{\text{sim}}(\boldsymbol{\phi})|^{1/2} \exp \left[-(\mathbf{D}_{\text{obs}} - \bar{\mathbf{d}}_{\text{sim}}(\boldsymbol{\phi}))^T \boldsymbol{\Sigma}_{\text{sim}}^{-1}(\boldsymbol{\phi}) (\mathbf{D}_{\text{obs}} - \bar{\mathbf{d}}_{\text{sim}}(\boldsymbol{\phi})) \right] \quad (2)$$

where c is $(2\pi)^{k/2}$, k the dimension of \mathbf{D}_{obs} , $\bar{\mathbf{d}}_{\text{sim}}(\boldsymbol{\phi})$ is the corresponding vector of mean simulation outputs, and $\boldsymbol{\Sigma}_{\text{sim}}^{-1}(\boldsymbol{\phi})$ the covariance matrix of the simulation outputs that summarizes the variability within each dimension of the model output, but also correlations between those dimensions. Pseudocode for the entire parameter algorithm is provided in the Supplement.

2.2.2 Representation of the data

Similar as in ABC, it is desirable to represent the data used in Eq. (2) in a low-dimensional form so that the estimation particularly of $\boldsymbol{\Sigma}_{\text{sim}}^{-1}(\boldsymbol{\phi})$ can be achieved in

BGD

10, 13097–13128, 2013

Approximate bayesian calibration of a forest model

F. Hartig et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Approximate bayesian calibration of a forest model

F. Hartig et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



a computationally efficient way. Therefore, we do not use the raw data (which would be in our case the size and identity of each tree on the plot), but some aggregation (summary statistics) of the data. Ideally, one would choose the lowest-dimensional aggregation of the original data for which there is no loss of information with respect to the inference (minimal sufficient statistics), but there is no generally accepted rule about how to find this representation (but see Fearnhead and Prangle, 2012; Blum et al., 2013). We therefore decided to use mainly two aggregations of forest data that have been frequently used for summarizing local inventory data in forest modeling. The first aggregation is stem size distributions per PFT, which count the number of tree individuals in (here) 10 cm size classes. The second is the size specific mean growth, which quantifies the mean stem diameter growth for different tree size classes. We also experimented with other forest attributes or aggregations of the data (see Table 2).

2.2.3 Posterior estimation

Subsequent posterior estimation based on the approximate likelihood was done with an adaptive Metropolis-Hastings MCMC (Haario et al., 2001). We always ran several chains and checked convergence visually and with Gelman–Rubin diagnostics (Gelman and Rubin, 1992, see Supplement for further details). As we required typically several seconds to evaluate a single parameter combination with FORMIND, posterior estimations cost substantial computing time. The exact number, length and burn-in of chains are provided in the figure captions. Figure 2 visualizes a summary of the analysis method.

2.3 Field data, model setup and analysis

We used two data sets to fit the parameters of the model, a “virtual” 1 ha inventory with three plant functional types (PFT1 = pioneer, PFT2 = mid-successional, PFT3 = late successional) that was created from the FORMIND model itself (which has the advantage that the “true” parameter values are known), and a 5 ha forest inventory from

**Approximate
bayesian calibration
of a forest model**

F. Hartig et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

a montane tropical rainforest in Ecuador that is described in (Dislich et al., 2009). The purpose of the virtual dataset is to test the parameter estimation method for different data types in a situation where true parameters are known, while the data from Ecuador provides a realistic case study by testing the method in a situation that had previously been dealt with by informal model calibration.

To create the “virtual” inventory, we used a base parameterization that was adjusted for exhibiting biomass values and successional patterns typical to a wet tropical lowland rainforest. With this setting, we simulated 1000 model runs, and created virtual datasets from the mean equilibrium values of these replicates for different types of output variables (summary statistics) such as biomass, stem diameter growth rates and stem size distributions. We also experimented with a different number of parameters to be estimated. A summary of these options, labeled with V1–V5, is provided in Table 2. While the question of sufficiency of different data types for the parameterization is a fundamental concern in this type of analysis (e.g. Jabot and Chave, 2009), the number of estimated parameters is more a practical issue: we knew already before that FORMIND exhibits interactions between parameters with respect to these outputs, which slows down posterior estimation and makes the analysis of the posterior distribution difficult.

For the fit of the model to field data in Ecuador, a tree-species grouping into seven PFTs was used that is described in detail in Dislich et al. (2009). Due to data availability, we used only the stem size distributions for the parameter estimation which was labeled E1.

3 Results

3.1 Fit to virtual field data (tropical lowland rainforest, V1–V5)

For fitting the model to the virtual field data, we considered a number of options that differ in the aggregation of model outputs, and in the number of estimated parameters.

We concentrate here on the case V1 in Table 2 (detailed data, not all parameters under calibration). Results for the other cases are discussed shortly below. Detailed results are provided in the Supplement.

3.1.1 Marginal distributions

Figure 3a shows the estimated marginal posterior densities (Eq. 1) for the parameterization V1 in Table 2, which represents the probability assigned to the different possible values for the respective parameter. We find that most parameter values are retrieved correctly and show moderate uncertainties of the order of 20–50 % of the mean. In interpreting these plots, note that “marginal” means that we display the values of one particular parameter in the posterior sample without consideration of the corresponding values of the other parameters. The latter is important because marginally, the uncertainty of a parameter appears often substantially larger than it effectively is due to correlations between parameters. We examine this in Sect. 3.1.2.

3.1.2 Correlations

Marginal distributions represent a cross-section of the posterior sample along a particular parameter axis, which neglects potential trade-offs between parameters with respect to the data that is used for the fit. Statistical models are usually designed in a way that such correlations are avoided wherever possible. For process-based models, the correspondence to concrete biological mechanisms is usually the main design criterion. It is therefore likely that such correlations will appear when estimating parameters of process-based models, as evidenced by Fig. 3b. Moreover, it is to be expected that the correlation structure depends on the data used to fit the model – less informative data will typically lead to more parameter combinations that can reproduce this data, affecting the correlation structure in the posterior sample.

These expectations are largely confirmed by our results. We find strong negative correlations particularly for recruitment and mortality of early successional types, as

Approximate bayesian calibration of a forest model

F. Hartig et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



one would expect, because, for those PFTs, increased mortality can be compensated for to some extent by increased recruitment. Also, we find that the correlation structure changes with the data types used. A detailed analysis of the correlation structure for the different data types (summary statistics) tested by us is provided in the Supplement.

5 3.1.3 Choice of data type and number of fitted parameters

For the parameter estimations V3–V5 in Table 2 that used less information (more aggregated model outputs or summary statistics), posterior parameter estimates were clearly wider than for our baseline V1 (Table 2, for details, see Supplement). It can therefore be concluded that all further aggregations of the data used in V1 lose information for the purpose of estimating the considered parameters (see also Wiegand et al., 2004). Similarly, increasing the number of fitted parameters (V2) increases the width of the posterior distribution. For all cases, the results indicate that more coarse aggregations or more parameters under calibration lead to additional correlations between parameters with respect to the objective of reproducing the respective data type, with the consequence of wider marginal distributions and slower convergence of the MCMC algorithms (Table 2, see Supplement for details).

3.1.4 Reduction of predictive uncertainty

Based on the posterior distribution of parameter values, we may also estimate the posterior predictive uncertainty for output values of the model, that is, the uncertainty estimates that are generated for the output of the model given the posterior parameter uncertainty for the parameters of the model. We compare three cases, the inherent stochastic uncertainty of the model with the “true” parameters, the uncertainty resulting from parameters drawn from the prior distribution (i.e. before parameter estimation), and the uncertainty for parameters drawn from the posterior distribution (i.e. after parameter estimation). We see that the posterior predictive mean is similar to that of the “true” parameters, with predictive uncertainty only slightly larger than for the “true”,

Approximate bayesian calibration of a forest model

F. Hartig et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



any stochastic ecological model. This technique provides a universal and unambiguous way to connected stochastic ecological models to field data.

The present study is one of the first to apply this method to a large ecological model. We use a parametric likelihood approximation, proposed by Wood (2010), to fit FOR-MIND, a relatively complex individual-based forest gap model, to a range of different virtual field data sets created from the model as well as to real field data from an Ecuadorian tropical forest.

4.1 Validation of the method with virtual field data

Fitting the model to different virtual field data sets allowed us to assess uncertainty and bias of the fit for situations where true parameters were known. For the most detailed data (abundance and growth distributions, case V1 in Table 2), parameter values were largely unbiased, with correlations between a few of the parameters (Fig. 4, as well as Figs. 2 and 3 in the Supplement). With increasing level of aggregation, (V3–V5), parameter values showed increasing correlations, bias and uncertainty. Note that it is important to take correlations into account when interpreting marginal parameter uncertainties such as Fig. 3a: if there are correlations between parameters, marginal uncertainties appear wider than in the multivariate correlation plots. This remains true for higher-order correlations which are likely present for more aggregated data types such as V4 and V5, but which are not covered by our analysis. Comparing the extent to which model parameters are constraint by the data by the width of their marginal posterior distribution only can therefore be misleading in the presence of strong correlations.

Correlations in the posterior indicate a trade-off between parameters with respect to the data that is used for the fit. For the data V1, for example, correlations occurred mostly within the parameters of the same PFT, with higher mortality requiring higher recruitment, with the strongest correlation for the early-successional type and weakest correlation for the late-successional type. The explanation is that if the mortality of a species is increased, we have to increase recruitment or growth to maintain similar

BGD

10, 13097–13128, 2013

Approximate bayesian calibration of a forest model

F. Hartig et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Approximate bayesian calibration of a forest model

F. Hartig et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



species numbers in the community. However, as V1 contained data on growth data, growth rates are tightly constrained, so the only option to maintain a similar species number is to increase recruitment, those effects are most strong for the pioneer species that has the highest turnover. The fits to data V3 without growth data, on the other hand, show correlation also with mortality and growth, because growth is unconstrained for this data type.

It is an advantage of the Bayesian analysis (or rather the use of an MCMC) that these interactions can be made explicit and interpreted. Thinking about the reasons for correlations may be helpful to understand and improve the model structure, but a correlation in the posterior does not necessarily mean that two parameters share a particular ecological connection. It just means that changes in one of the parameters may be counterbalanced by the other to maintain the same value of the model output under consideration. Thus, correlations are connected to a data type, and they inform us which parameters cannot be fully constraint by this data type. For example, correlations and bias increase from V1 to V3, indicating that even for fitting recruitment, mortality and growth parameters only, static data such as stem size distributions does not provide sufficient information to constrain all parameters at once.

Bias or correlations observed in the virtual test case seemed to originate predominantly from data limitations and not from problems with the simulation-based likelihood approximation. We have no indications that would suggest that the parametric model (multivariate normal) used in the likelihood approximation created any problems or bias by not adequately summarizing model outputs, which would be theoretically possible. Due to the computational complexity of this study, however, it was not possible for us to make a more systematic analysis of this question, for example by using virtual replicates of the field data sets or less aggregated data types.

4.2 Fit to Ecuadorian field data

Only static data was available to us for fitting the FORMIND model to field data from a montane forest in Ecuador. Our previous analysis suggested that these data would

not be sufficient to sensibly constrain all demographic parameters at once. To get ecologically interpretable results, we therefore fixed the recruitment parameters to the values used in Dislich et al. (2009), and calibrated mortality and growth parameters only. Prior uncertainty was considerably reduced by these data (Fig. 5), suggesting that the data was relatively informative to constrain the parameters under calibration. Marginal posterior parameter estimates are similar to those that were derived by Dislich et al. (2009) with a combination of literature data, expert knowledge and calibration (see Supplement, Table 3 for exact values).

From the fits to the virtual field data V3 (Fig. 7, Supplement), we expected correlations in the posterior mostly to occur between parameters of the same PFT. We find those correlations, but we find additionally correlations particularly between the mortality parameters of some PFTs (Fig. 13, Supplement). To understand this, one has to know that species grouping designed by Dislich et al. (2009) is hierarchical, consisting of 7 PFTs that were further divided into 4 growth groups with equal maximum diameter growth for the PFTs in each group, with the following relation between (PFT) and growth group: (1)-2,(2)-1,(3,4)-3,(5,6,7)-4. Diameter growth parameter 3, which is estimated lower, thus applies to the mid-successional PFTs 3 and 4, and diameter growth parameter 4, estimated higher, applies to the late successional PFTs 5, 6 and 7. This hierarchical species grouping is mirrored in the correlation structure, with particularly strong correlations in the mortality parameters of PFTs that belong to the same growth group. Our interpretation of this pattern is that PFTs in the same growth group are competing more strongly with each other than those that are in different growth groups.

Differences to the parametrization of Dislich et al. (2009) are particularly evident in the mortality parameters. Lower values were estimated for the mortality of the mid-successional PFTs 3 and 4, while mortality of the late successional PFTs 5, 6 and 7 was estimated higher. This pattern is mirrored in the maximum diameter growth rates of mid-successional species. Thus, our study points to less pronounced differences between mid and late-successional types than Dislich et al. (2009). We can only speculate about the reason for these differences. In general, one would think that the system-

BGD

10, 13097–13128, 2013

Approximate bayesian calibration of a forest model

F. Hartig et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



atic parameter estimation is more reliable than the manual calibration by Dislich et al. (2009). However, although Dislich et al. (2009) calibrated to the same data, they also considered the fit of other model outputs such as total biomass and expert opinions for fixing the parameters. Expert opinion in particular would favor more pronounced differences in mortality rates between mid and late successional species due to ecological expectations, although specific empirical data on tree mortality or on maximum growth rates under full light were not available. Secondly, there are significant correlations between the parameters, which allow gaining a similar fit with a range of different parameter values. And finally, we were using the model in this study at a lower temporal resolution (5 yr time steps) than Dislich et al. (2009) to reduce computing time, which can affect model dynamics and equilibrium distributions, meaning that slightly different parameter values would be estimated for the same model with different temporal resolution.

4.3 Advantages compared to conventional calibration methods

Our results demonstrate that inverse parameter estimation with a likelihood function derived from the stochasticity in the model outputs is feasible and provides good results, even for a relatively complex and runtime-intensive ecological model. This is encouraging in itself, as it is neither trivial to calibrate a parameter-rich model with heterogeneous data in general, nor to address all the technical challenges for performing the simulation-based likelihood approximation. A valid question, however, is whether the gain is worth the effort – after all, our approach is connected with considerable computational and conceptual costs, and all we gain are parameter estimates that could probably also have been derived with conventional inversion methods such as parameter optimization.

However, there are practical advantages of simulation-based likelihood approximations for ecological research that extend far beyond what we could demonstrate in this study. First of all, there is considerable interest in connecting models to large and heterogeneous data sources that become increasingly available (Luo et al., 2011; Hartig

BGD

10, 13097–13128, 2013

Approximate bayesian calibration of a forest model

F. Hartig et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



et al., 2012; Dietze et al., 2013). A practical problem in this context is that conventional methods provide no good answer for how to weight different data sources to construct a joint likelihood or objective function. Moreover, ecological processes nearly inevitably lead to correlations between those different data types, meaning that we would not expect errors to be independent, posing a challenge for conventional methods. Simulation-based likelihood approximations provide a natural answer to these problems. Assuming that the simulation model includes all major sources of stochasticity, likelihoods approximations automatically weight the importance of different model outputs and account for correlations between them. In our study, we can see this in the combined fit of growth rates and stem size distributions, which required no weighting of these two patterns and automatically accounted for second order correlations between them.

Moreover, under conventional inverse parametrization procedures, one might see that a certain pattern is not well represented, but it is often difficult to decide whether this is a random or a systematic problem. Simulation-based likelihood approximations allow making a definite statement about the probability of observed patterns given the current model (parameters). Thus, we can use the full arsenal of statistical procedures, including Bayesian and frequentist model selection, to compare alternative ecological hypotheses. The possibility of such rigorous statistical tests for alternative process-based models will likely increase the acceptance of process-based models as a tool for not only representing and predicting, but for statistically testing ecological knowledge.

4.4 Differences to ABC

The comments in the previous subsection apply to parametric and non-parametric likelihood approximation alike. However, it also seems interesting to discuss differences between the parametric likelihood approximation used in this study and the more widely used non-parametric approximation used in Approximate Bayesian Computing (ABC, Beaumont, 2010). As discussed in Hartig et al. (2011), unlike ABC, parametric likelihood approximations will nearly inevitably exhibit a certain amount of approximation

BGD

10, 13097–13128, 2013

Approximate bayesian calibration of a forest model

F. Hartig et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



BGD

10, 13097–13128, 2013

**Approximate
bayesian calibration
of a forest model**

F. Hartig et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

error because it is unlikely that a simple distributional model can emulate model output distributions in all respects (particularly in the tails of the output distribution). Yet, the parametric approximation also has practical advantages. Many ecological models have to be run into equilibrium before predictions can be made. Once such a model is in equilibrium, more draws for the parametric approximation can be generated relatively cheap, while a new run has to be started for each ABC step. In our example, the time required for the parametric approximation in one MCMC step was not much longer than for an ABC step, but the parametric approximation ensures a good acceptance probability. To reach the same acceptance probability with ABC, we would have to accept a relatively large ABC approximation error. This error may be corrected later, but the fact remains that for situations where the number of possible MCMC evaluation is fixed (complex models), both ABC and parametric approximations will have a non-negligible error. We conjecture that the balance could well be in favor of parametric approximations in situations such as the one encountered in this study.

5 Conclusions

Our results suggest that likelihood approximations, in particular parametric likelihood approximations, are a promising route for the parametrization of stochastic ecological models. Using them is technically more challenging than the “traditional” Bayesian approach where likelihoods are based on phenomenological error models. The advantage, however, is that error models are based on the same ecological mechanisms as all other model predictions. Thus, they allow a more rigorous test of the mechanistic model assumptions, because the mechanisms have to explain both the mean and the variance in the data. Moreover, likelihood approximations account for the relative importance and correlations between different data types predicted by the model, which makes them interesting when models have to be coupled to heterogeneous data. In this study, additional computational costs of the approach were moderate (factor 2–5) compared to a standard Bayesian approach due to the fact that the model had to be

run into equilibrium in any case. Such runtime differences appear secondary compared to the methodological advantage of rigorously testing our mechanistic understanding of ecosystems against field data, including the sampling and measurement process. Parametric likelihood approximations therefore seem particularly promising for models that have to be run into equilibrium, contain the dominant stochastic processes, use heterogeneous data, and can still be well summarized by simple distributions.

Supplementary material related to this article is available online at <http://www.biogeosciences-discuss.net/10/13097/2013/bgd-10-13097-2013-supplement.pdf>.

Acknowledgements. F. Hartig gratefully acknowledges financial support from ERC Advanced Grant 233066. C. Dislich gratefully acknowledges financial support from DFG Research Unit 816.

The service charges for this open access publication have been covered by a Research Centre of the Helmholtz Association.

References

- Beaumont, M. A.: Approximate Bayesian computation in evolution and ecology, *Annu. Rev. Ecol. Evol. Syst.*, 41, 379–406, doi:10.1146/annurev-ecolsys-102209-144621, 2010. 13100, 13105, 13115
- Blum, M. G. B., Nunes, M. A., Prangle, D., and Sisson, S. A.: A comparative review of dimension reduction methods in approximate Bayesian computation, *Stat. Sci.*, 28, 189–208, 2013. 13106
- Bossel, H.: Real-structure process description as the basis of understanding ecosystems and their development, *Ecol. Model.*, 63, 261–276, doi:10.1016/0304-3800(92)90072-M, 1992. 13099

BGD

10, 13097–13128, 2013

Approximate bayesian calibration of a forest model

F. Hartig et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Approximate bayesian calibration of a forest model

F. Hartig et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- Bugmann, H. K. M.: A simplified forest model to study species composition along climate gradients, *Ecology*, 77, 2055–2074, doi:10.2307/2265700, 1996. 13101
- Chave, J., Condit, R., Lao, S., Caspersen, J. P., Foster, R. B., and Hubbell, S. P.: Spatial and temporal variation of biomass in a tropical forest: results from a large census plot in Panama, *J. Ecol.*, 91, 240–252, 2003. 13104
- Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S., and Perez, R.: Error propagation and scaling for tropical forest biomass estimates, *Philos. T. R. Soc. B*, 359, 409–420, doi:10.1098/rstb.2003.1425, 2004. 13104
- Clark, J. S.: Why environmental scientists are becoming Bayesians, *Ecol. Lett.*, 8, 2–14, doi:10.1111/j.1461-0248.2004.00702.x, 2005. 13099
- Csilléry, K., Blum, M. G. B., Gaggiotti, O. E., and François, O.: Approximate Bayesian Computation (ABC) in practice, *Trends Ecol. Evol.*, 25, 410–418, doi:10.1016/j.tree.2010.04.001, 2010. 13105
- Dietze, M. C., Lebauer, D. S., and Kooper, R.: On improving the communication between models and data, *Plant Cell Environ.*, 36, 1575–1585, doi:10.1111/pce.12043, 2013. 13115
- Diggle, P. J. and Gratton, R. J.: Monte Carlo methods of inference for implicit statistical models, *J. Roy. Stat. Soc. B Met.*, 46, 193–227, 1984. 13100
- Dislich, C. and Huth, A.: Modelling the impact of shallow landslides on forest structure in tropical montane forests, *Ecol. Model.*, 239, 40–53, doi:10.1016/j.ecolmodel.2012.04.016, 2012. 13102
- Dislich, C., Günter, S., Homeier, J., Schröder, B., and Huth, A.: Simulating forest dynamics of a tropical montane forest in South Ecuador, *Erdkunde*, 63, 347–364, doi:10.1098/rstb.2003.1425, 2009. 13103, 13107, 13113, 13114, 13122, 13128
- Ellison, A. M.: Bayesian inference in ecology, *Ecol. Lett.*, 7, 509–520, doi:10.1111/j.1461-0248.2004.00603.x, 2004. 13103
- Fearnhead, P. and Prangle, D.: Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation, *J. Roy. Stat. Soc. B Met.*, 74, 419–474, 2012. 13106
- Gelman, A. and Rubin, D.: Inference from iterative simulation using multiple sequences, *Stat. Sci.*, 7, 457–472, 1992. 13106
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B.: *Bayesian Data Analysis*, 2nd edn., Chapman & Hall, London, 2003. 13103

Approximate bayesian calibration of a forest model

F. Hartig et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Grimm, V. and Railsback, S. F.: Pattern-oriented modelling: a “multi-scope” for predictive systems ecology, *Phil. Trans. R. Soc. B*, 367, 298–310, doi:10.1098/rstb.2011.0180, 2012. 13100

Gutiérrez, A. G. and Huth, A.: Successional stages of primary temperate rainforests of Chiloé Island, Chile, *Perspect. Plant Ecol. Evol. Syst.*, 14, 243–256, doi:10.1016/j.ppees.2012.01.004, 2012. 13102

Haario, H., Saksman, E., and Tamminen, J.: An adaptive Metropolis algorithm, *Bernoulli*, 7, 223–242, 2001. 13106

Hartig, F., Calabrese, J. M., Reineking, B., Wiegand, T., and Huth, A.: Statistical inference for stochastic simulation models – theory and application, *Ecol. Lett.*, 14, 816–827, doi:10.1111/j.1461-0248.2011.01640.x, 2011. 13100, 13103, 13105, 13115

Hartig, F., Dyke, J., Hickler, T., Higgins, S. I., O’Hara, R. B., Scheiter, S., and Huth, A.: Connecting dynamic vegetation models to data – an inverse perspective, *J. Biogeogr.*, 39, 2240–2252, doi:10.1111/j.1365-2699.2012.02745.x, 2012. 13099, 13114

Higgins, S. I., O’Hara, R. B., Bykova, O., Cramer, M. D., Chuine, I., Gerstner, E.-M., Hickler, T., Morin, X., Kearney, M. R., Midgley, G. F., and Scheiter, S.: A physiological analogy of the niche for projecting the potential distribution of plants, *J. Biogeogr.*, 39, 2132–2145, doi:10.1111/j.1365-2699.2012.02752.x, 2012. 13099

Huth, A. and Ditzer, T.: Simulation of the growth of a lowland Dipterocarp rain forest with FORMIX3, *Ecol. Model.*, 134, 1–25, doi:10.1016/S0304-3800(00)00328-8, 2000. 13101

Jabot, F. and Chave, J.: Inferring the parameters of the neutral theory of biodiversity using phylogenetic information and implications for tropical forests, *Ecol. Lett.*, 12, 239–248, doi:10.1111/j.1461-0248.2008.01280.x, 2009. 13101, 13107

Jabot, F. and Chave, J.: Analyzing tropical forest tree species abundance distributions using a nonneutral model and through approximate Bayesian inference, *Am. Nat.*, 178, E37–E47, doi:10.1086/660829, 2011. 13101

Kass, R. E. and Wasserman, L.: The selection of prior distributions by formal rules., *J. Am. Stat. Assoc.*, 91, 1343–1370, 1996. 13104

Köhler, P.: Modelling anthropogenic impacts on the growth of tropical rain forests, Ph.D. thesis, University Kassel, 2000. 13102, 13103

Köhler, P. and Huth, A.: Towards ground-truthing of spaceborne estimates of above-ground life biomass and leaf area index in tropical rain forests, *Biogeosciences*, 7, 2531–2543, doi:10.5194/bg-7-2531-2010, 2010. 13102, 13104

Approximate bayesian calibration of a forest model

F. Hartig et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- Kohyama, T.: Size-structured tree populations in gap-dynamic forest – the forest architecture hypothesis for the stable coexistence of species, *J. Ecol.*, 81, 131–143, 1993. 13101
- Luo, Y., Ogle, K., Tucker, C., Fei, S., Gao, C., LaDeau, S., Clark, J. S., and Schimel, D. S.: Ecological forecasting and data assimilation in a data-rich era, *Ecol. Appl.*, 21, 1429–1442, doi:10.1890/09-1275.1, 2011. 13114
- Martínez, I., Wiegand, T., Camarero, J. J., Batllori, E., and Gutiérrez, E.: Disentangling the formation of contrasting tree line physiognomies combining model selection and Bayesian parameterization for simulation models, *Am. Nat.*, 5, E136–E152, doi:10.1086/659623, 2011. 13100, 13104
- May, F., Giladi, I., Ristow, M., Ziv, Y., and Jeltsch, F.: Metacommunity, mainland-island system or island communities?, assessing the regional dynamics of plant communities in a fragmented landscape, *Ecography*, 36, 842–853, doi:10.1111/j.1600-0587.2012.07793.x, 2013. 13101
- McCarthy, J.: Gap dynamics of forest trees: a review with particular attention to boreal forests, *Environ. Rev.*, 9, 1–59, doi:10.1139/a00-012, 2001. 13101
- O’Hara, R. B., Arjas, E., Toivonen, H., and Hanski, I.: Bayesian analysis of metapopulation data, *Ecology*, 83, 2408–2415, 2002. 13099
- Pacala, S. W., Canham, C. D., Saponara, J., Silander, J. A., Kobe, R. K., and Ribbens, E.: Forest models defined by field measurements: estimation, error analysis and dynamics, *Ecol. Monogr.*, 66, 1–43, doi:10.2307/2963479, 1996. 13101
- Purves, D. W., Zavala, M. A., Ogle, K., Prieto, F., and Benayas, J. M. R.: Environmental heterogeneity, bird-mediated directed dispersal, and oak woodland dynamics in mediterranean spain, *Ecol. Monogr.*, 77, 77–97, doi:10.1890/05-1923, 2007. 13099
- Rüger, N., Gutiérrez, I. G., Kissling, W. D., Armesto, J. J., and Huth, A.: Ecological impacts of different harvesting scenarios for temperate evergreen rain forest in southern Chile – a simulation experiment, *Forest Ecol. Manage.*, 252, 52–66, doi:10.1016/j.foreco.2007.06.020, 2007. 13102
- Schröder, B. and Seppelt, R.: Analysis of pattern-process interactions based on landscape models – overview, general concepts, and methodological issues, *Ecol. Model.*, 199, 505–516, doi:10.1016/j.ecolmodel.2006.05.036, 2006. 13099
- Shugart, H. H.: *A Theory of Forest Dynamics: the Ecological Implications of Forest Succession Models*, Springer, New York, USA, New York, 1984. 13101
- Shugart, H.: *Terrestrial Ecosystems in Changing Environments*, Cambridge University Press, Cambridge, UK, 1998. 13101

BGD

10, 13097–13128, 2013

Approximate bayesian calibration of a forest model

F. Hartig et al.

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)


Van Oijen, M., Rougier, J., and Smith, R.: Bayesian calibration of process-based forest models: bridging the gap between models and data, *Tree Physiol.*, 25, 915–927, doi:10.1093/treephys/25.7.915, 2005. 13104

5 Van Oijen, M., Reyer, C., Bohn, F., Cameron, D., Deckmyn, G., Flechsig, M., Härkönen, S., Hartig, F., Huth, A., Kiviste, A., Lasch, P., Mäkela, A., Mette, T., Minunno, F., and Rammer, W.: Bayesian calibration, comparison and averaging of six forest models, using data from Scots pine stands across Europe, *For. Ecol. Manage.*, 289, 255–268, doi:10.1016/j.foreco.2012.09.043, 2013. 13100, 13104

10 Wiegand, T., Knauer, F., Kaczensky, P., and Naves, J.: Expansion of brown bears (*Ursus arctos*) into the Eastern alps: a spatially explicit population model, *Biodivers. Conserv.*, 13, 79–114, doi:10.1023/B:BIOC.0000004314.38828.db, 2004. 13100, 13109

Wood, S. N.: Statistical inference for noisy nonlinear ecological dynamic systems, *Nature*, 466, 1102–1104, doi:10.1038/nature09319, 2010. 13100, 13101, 13105, 13111

15 Zurell, D., Berger, U., Cabral, J. S., Jeltsch, F., Meynard, C. N., Münkemüller, T., Nehrbass, N., Pagel, J., Reineking, B., Schröder, B., and Grimm, V.: The virtual ecologist approach: simulating data and observers, *Oikos*, 119, 622–635, doi:10.1111/j.1600-0706.2009.18284.x, 2009. 13100

Approximate bayesian calibration of a forest model

F. Hartig et al.

Table 1. Important model parameters and their interpretation. The parameter dbh_m in the last line refers to the maximum diameter of a tree, which is a species or PFT specific parameter of the model (see Dislich et al., 2009).

Parameter	Notes	Units
light extinction coefficient	fraction of light intercepted per unit of LAI	$[\text{m}^2 \text{m}^{-2}]$
recruitment rate	new tree individuals at 1 cm dbh	$[\text{ha}^{-1} \text{yr}^{-1}]$
min light for establishment	for recruitment, relative to full irradiance	$[-]$
mortality rate	probability per year	$[\text{yr}^{-1}]$
falling probability	applies to large trees that die	$[-]$
leaf area index per tree	projected leaf area per unit area	$[\text{m}^2 \cdot \text{m}^{-2}]$
crown diameter	relative to diameter	$[\text{dbh}^{-1}]$
crown length	relative to tree height	$[-]$
max dbh growth (gro_m)	maximum stem diameter (sd) growth rate	$[\text{mm} \text{yr}^{-1}]$
start growth	sd growth rate for minimum stem diameter (relative to gro_m)	$[-]$
end growth	sd growth rate at maximum stem diameter (relative to gro_m)	$[-]$
max growth diameter	stem diameter of max growth (relative to dbh_m)	$[-]$

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Approximate bayesian calibration of a forest model

F. Hartig et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Table 2. Overview of parameter estimations with different models, parameters and summary statistics. Abbreviations for the data: SSD = stem size distribution (16 10 cm classes), GRO = mean stem diameter growth for each of the 16 10 cm stem diameter classes, BM = Biomass. If not stated otherwise, the data type was available for each PFT separately. If we use the mean over all PFTs, we label this by “total”. Full parameters means that all parameters listed in Tables 1 and 2 in the Supplement are estimated inversely. Reduced parameters means that from Table 1, only recruitment, mortality, maximum growth and maximum growth diameter are estimated. “Number of parameters” and “Data dimension” give the number of parameters and data points, respectively. “Posterior width” measures the posterior width of the marginal distributions by the ratio between marginal posterior standard deviation and uniform prior width averaged over all parameters. “Convergence ranking” provides a ranking of the speed of convergence of the MCMCs based on the convergence diagnostics discussed in the Supplement. Lower numbers indicate fastest convergence. As E1 uses different data and a different number of PFTs, the convergence ranking is not fully comparable and was set in parenthesis.

Label	Description	Number of Parameters	Data Dimension	Posterior Width	Convergence Ranking
Parameterization to virtual tropical forest (3 PFTs):					
V1	Field data: SDD, GRO, reduced parameters	12	96	0.019	1
V2	Field data: SDD, GRO, full parameters	26	96	0.096	5
V3	Field data: SDD, reduced parameters	12	48	0.073	2
V4	Field data: total SDD, reduced parameters	12	16	0.190	3
V5	Field data: BM, reduced parameters	12	3	0.192	4
Parameterization to Ecuadorian montane rain forest (7 PFTs):					
E1	Field data: SSD	18	112	0.036	(5)

Approximate bayesian calibration of a forest model

F. Hartig et al.

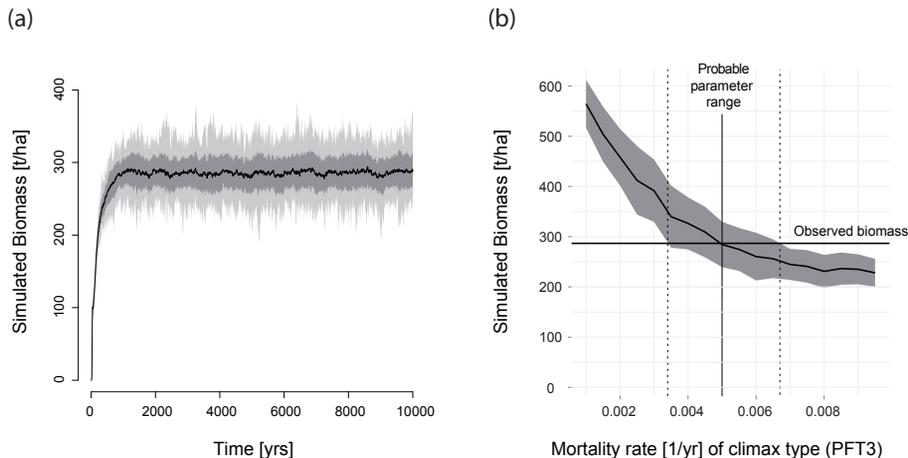


Fig. 1. Principle of statistical inference through stochastic simulation. **(a)** shows mean model predictions (black), standard deviation (gray) and extreme values (light gray) for the biomass of a 1 ha plot over 10 000 yr, starting from an empty plot. **(b)** shows the same mean equilibrium biomass (black) and two standard deviations (gray), but as a function of the mortality of the late-successional type PFT 3, all other parameters equal. We see that comparing the observed biomass from **(a)**, which was created with a mortality rate of 0.005, with the predicted biomass for different mortality rates, we can infer the original value as well as a statistical uncertainty, without having to define a statistical model.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Approximate bayesian calibration of a forest model

F. Hartig et al.

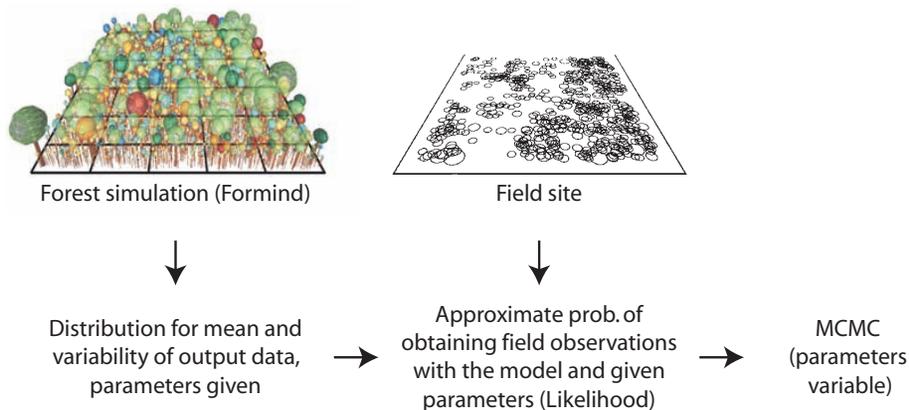


Fig. 2. Parameter estimation: at the top left, a conceptual illustration of the FORMIND model. Different colors represent different PFTs. The model is compared to the field data (middle) by fitting a distribution to the stochastic model output, and calculating the approximate probability of observing the field data from this distribution. This approximate likelihood value feeds into the conventional Bayesian analysis.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Approximate bayesian calibration of a forest model

F. Hartig et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

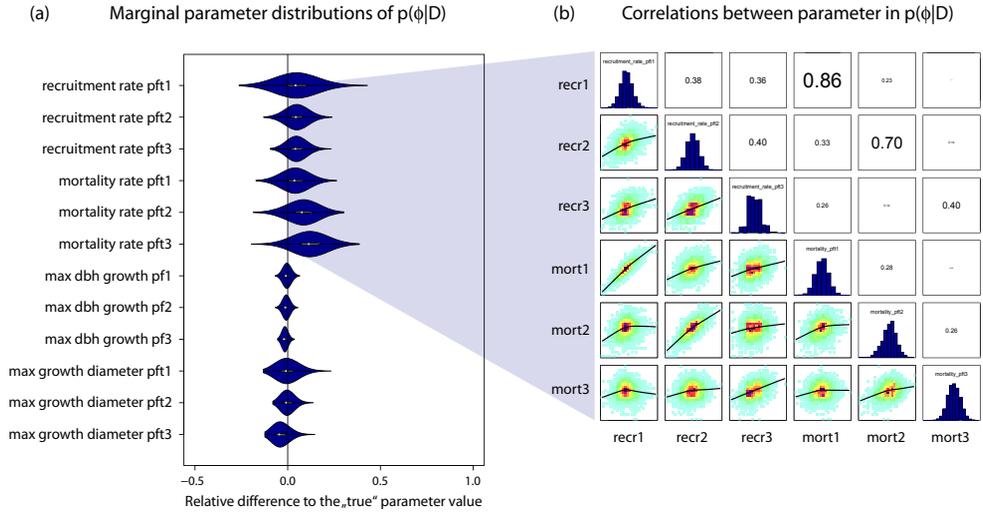


Fig. 3. Summaries of the estimated parameter values (shown as probability distributions) after fitting the model to the virtual field data (case V1 in Table 2). The distributions in (a) correspond to the marginal posterior density $p(\phi|D)$ for each parameter, scaled relative to the “true” values that were used to create the synthetic data (see Table 2 in the Supplement for true values and units). The dot within each distribution denotes the median value. Panels in (b) visualize correlations between recruitment and mortality parameters in the posterior sample (recr1 refers to the recruitment rate of PFT1, mort2 refers to mortality of PFT2 and so on). The diagonal shows the marginal distributions displayed in (a). The lower triangle shows the correlation density between the parameters on the diagonal (red values denoting higher density) and a nonlinear fit of the correlation (black line). The upper triangle shows Spearman’s rank correlation coefficients for the correlations in the lower triangle.

Approximate
bayesian calibration
of a forest model

F. Hartig et al.

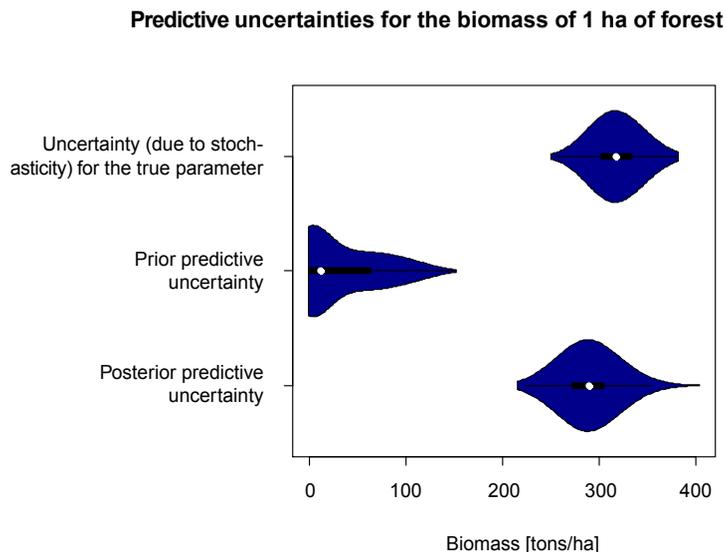


Fig. 4. True, prior and posterior predictive uncertainty. Each distribution is created from 1000 model runs, observing the biomass on a 1 ha forest plot after 2000 yr. The upper distribution shows biomass values from model runs with the same, “true” parameters (Table 2, Supplement), and thereby gives an estimate of the stochastic uncertainty of the model. For the middle distribution, model parameters were drawn from the prior distribution (resulting in what is called the prior predictive distribution). For the lower distribution, model parameters were drawn from the posterior (posterior predictive uncertainty).

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Approximate bayesian calibration of a forest model

F. Hartig et al.

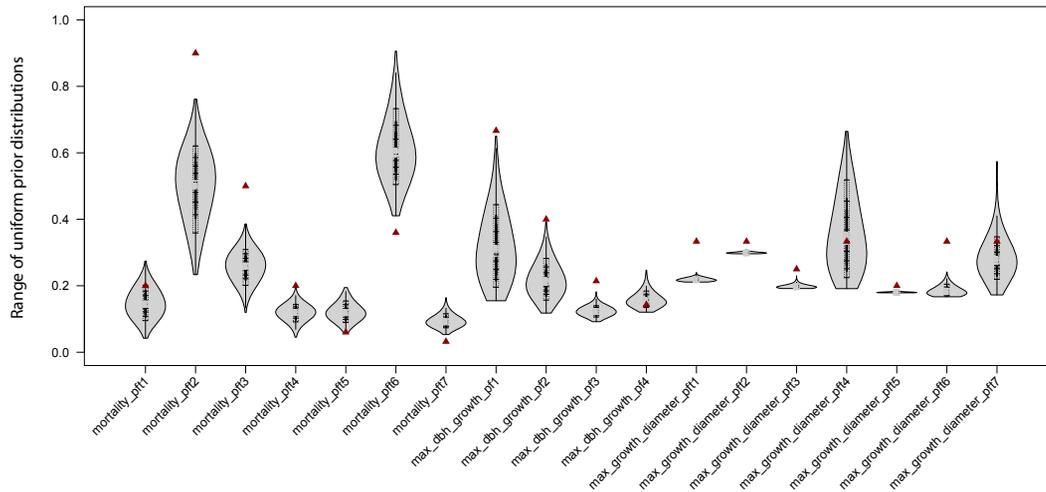


Fig. 5. Marginal posterior probabilities for the model parameters after fitting the model to field data from Ecuador, scaled relative to the uniform prior distributions (see Table 3 in the Supplement for prior values and units). Values used by Dislich et al. (2009) are marked as dark red triangles. An unscaled version of these distributions and correlations are provided in Figs. 12 and 13 of the Supplement.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

