

Bayesian calibration of process-based forest models: bridging the gap between models and data

MARCEL VAN OIJEN,^{1,2} JONATHAN ROUGIER³ and RON SMITH¹

¹ CEH-Edinburgh, Bush Estate, Penicuik EH26 0QB, U.K.

² Corresponding author (mvano@ceh.ac.uk)

³ University of Durham, Durham DH1 3LE, U.K.

Received August 6, 2004; accepted October 30, 2004; published online May 2, 2005

Summary Process-based forest models generally have many parameters, multiple outputs of interest and a small underlying empirical database. These characteristics hamper parameterization. Bayesian calibration offers a solution to the calibration problem because it applies to models of any type or size. It provides parameter estimates, with measures of uncertainty and correlation among the parameters. The procedure begins by quantifying the uncertainty about parameter values in the form of a prior probability distribution. Then data on the output variables are used to update the parameter distribution by means of Bayes' Theorem. This yields a posterior calibrated distribution for the parameters, which can be summarized in the form of a mean vector and variance matrix. The predictive uncertainty of the model can be quantified by running it with different parameter settings, sampled from the posterior distribution. In a further step, one may evaluate the posterior probability of the model itself (rather than that of the parameters) and compare that against the probability of other models, to aid in model selection or improvement.

Bayesian calibration of process-based models cannot be performed analytically, so the posterior parameter distribution must be approximated in the form of a representative sample of parameter values. This can be achieved by means of Markov Chain Monte Carlo simulation, which is suitable for process-based models because of its simplicity and because it does not require advance knowledge of the shape of the posterior distribution.

Despite the suitability of Bayesian calibration, the technique has rarely been used in forestry research. We introduce the method, using the example of a typical forest model. Further, we show that reductions in parameter uncertainty, and thus in output uncertainty, can be effected by increasing the variety of data, increasing the accuracy of measurements and increasing the length of time series.

Keywords: *Markov Chain Monte Carlo, tree measurement, parameterization, uncertainty.*

Introduction

Process-based forest models (PBMs) are tools for explaining and predicting the dynamics of forest ecosystems. They simulate forest behavior by integrating information on the underlying processes in trees, soil and atmosphere. Their complexity is a strength, allowing them, in principle, to reproduce the complex dynamics of forest ecosystems in detail, but it is also a weakness because it makes their use and evaluation difficult. Many different PBMs have been published, but in most cases we know little about their reliability. There is a need for a method, applicable to all PBMs, that quantifies output uncertainty, identifies key parameters and variables, and is of use in model improvement and model selection. Such methods tend to be ad-hoc, and the need for improvement has often been noted (Gardner et al. 1990, Mäkelä et al. 2000, Robinson and Ek 2000, Van Oijen 2002).

Our goal in this paper is to show how a Bayesian approach provides a logical method for calibrating forest PBMs. The main characteristic of the Bayesian approach is that it quantifies model inputs and outputs in the form of probability distributions, and applies the rules of probability theory to update the distributions when new data are obtained (Sivia 1996). Bayesian approaches have been much less frequently used in forest research (Ghazoul and McAllister 2003) than in the other environmental sciences (Reckhow 1990, Ellison 1996, 2004). Although a recent review of evaluation methods for forest models used for carbon accounting does not mention Bayesian approaches at all (Prisley and Mortimer 2004), some pioneering efforts in Bayesian parameterization of forest models have been made (Gertner et al. 1999, Green et al. 1999, 2000, Radtke et al. 2002).

Treating everything probabilistically, even if the model has many different inputs and outputs, makes Bayesian calibration computationally demanding. This has hampered its practical use in the past, and no examples exist for forest models with many parameters. However, we show how recent developments in sampling-based evaluation of probability distribu-

tions, in particular Markov Chain Monte Carlo techniques, can help alleviate the computational problem.

A Bayesian framework

Process-based models in forestry tend to be deterministic, so that the values of inputs and parameters determine model output completely. This is not at odds with the probabilistic framework we present in this paper. Determinism in model structure may reflect a modeler's idea that the real world operates according to deterministic physical laws. However, even if this is so, our knowledge of the world is incomplete. A recent review of forest modeling identified a range of biogeochemical processes, already incorporated in models, for which data are lacking (Van Oijen et al. 2004). Our incomplete knowledge constitutes uncertainty that we may represent in the form of probability distributions. Our uncertainties are linked: uncertain parameters imply uncertain predictions and uncertainty about the real world implies uncertainty about model structure and parameterization. Because of these linkages, our methods for model parameterization, uncertainty analysis, sensitivity analysis, prediction, testing and comparison with other models need to be based on a consistent quantification of uncertainty. Bayesian statistics aims to provide such "rules and procedures for disciplined uncertainty accounting" (Bernardo and Smith 1994). In the Bayesian approach, all uncertainties are quantified in terms of probabilities, and the probability calculus is shown to be appropriate for coherent updating of uncertainty using data. More details and further references can be found in Bernardo and Smith (1994, Chapter 2).

The procedure

The procedure we propose begins with quantifying the uncertainty about parameter values in the form of so-called prior probability distributions. Then we use measured data on the output variables to calibrate the model. The calibration, which is done by application of Bayes' Theorem, yields an updated posterior distribution for the parameters. This focus on quantifying uncertainty in the form of probability distributions distinguishes Bayesian calibration from parameter estimation methods that focus on goodness-of-fit, like maximum-likelihood (ML) estimation of parameters. (We consider ML methods to be less useful than Bayesian calibration because ML precludes the use of prior information on parameters and provides no probability distributions for the parameters that express our uncertainty about them in cases where we have limited amounts of data.)

The predictive uncertainty of the model can be determined by running the model with different parameter settings sampled from the posterior parameter distribution. If so desired, we may proceed further and evaluate the posterior probability of the model itself (rather than that of the parameters), comparing it against the probability of other models, to aid in model selection or improvement.

Communicating models and model results

Application of a probabilistic framework affects not only the way we model, but also the way we report our model and mod-

eling results. When communicating model results, we need to present more than just the most likely predictions. Uncertainty intervals should also be generated. When communicating the models themselves, we should not only provide default values of all parameters, but also present the full joint probability distribution of the parameters, including their correlations, as produced by the Bayesian calibration. New model users can use this parameter probability distribution as their prior, and update it by applying Bayes' Theorem to their own measured output data, as long as it is made clear which data have already been used, to avoid double-counting. Nevertheless, some of the information contained in the parameter distribution may be of limited value to new users if they plan to apply the model to different conditions, e.g., a new cultivar.

Approach

The approach taken in this paper is to focus on the simplest and most widely applicable methods for Bayesian modeling, rather than on the most efficient ones. We aim to provide sufficient detail of the method that a reader can initially implement the algorithms without consulting other material. However, we will provide pointers to the literature where methods are presented that may be less simple but more efficient in specific circumstances. We explain our basic method using the realistic forest PBM BASFOR as an example; however, the model is not the focus of this article and any other model could have been used in its place because the techniques presented are applicable generally. We then provide a formal presentation of Bayesian statistics, followed by a description of how the methods are implemented in computer algorithms, and how they work out on BASFOR. Finally, we show how we can determine what kind of data most reduce uncertainty.

Parameterization of forest model BASFOR

General aspects of forest models

Bayesian calibration can be applied to any kind of deterministic model, static or dynamic, simple or complex, analytically solvable or simulation-based. It can also be generalized to stochastic models. Here we demonstrate the use of the technique for deterministic process-based forest models (PBMs) that share a number of characteristics. First, they are dynamic simulation models that tend to be parameter-rich. For example, a recent forest model comparison analyzed models ranging from 44 to 261 parameters, not counting initial constants (Van Oijen et al. 2005). Second, forest PBMs produce many different kinds of output that may be of interest to the user. Third, relatively few measured data tend to be available for the many parameters and output variables of such models (Van Oijen et al. 2004). Examples are rare where this typical combination of parameter richness, multiplicity of outputs and data scarcity has been the object of Bayesian techniques. To demonstrate the feasibility of the approach in such cases, we show how a new generic conifer forest model, BASFOR, was calibrated for specific use with Norway spruce (*Picea abies* (L.) Karst.).

Forest model BASFOR

The focus of our paper is on Bayesian methodology, not on forest modeling as such, so only a short description of the BASFOR model is given. The calibration procedure is generic and can be applied to any forest model. The BASic FOREst simulator, BASFOR, simulates carbon and nitrogen cycling in trees, soil organic matter and litter. It simulates the response of trees and soil to radiation, temperature, precipitation, humidity, wind speed, atmospheric CO₂ and N-deposition and thinning regime. The model has 11 state variables, representing carbon and nitrogen pools in trees and soil, and 39 parameters that include the initial constants of the state variables (Table 1). Besides time series for the state variables, output may be produced for net primary productivity (NPP), tree height, ground cover, leaf area index (LAI), N mineralization and other tree and soil variables. BASFOR is built from well-known process representations. Light absorption is calculated by Beer's law. Gross primary productivity (GPP) is calculated as light absorption times light-use efficiency (LUE). Net primary productivity is calculated as a fixed ratio of GPP. Light-use efficiency is temperature- and CO₂-dependent and may be reduced

if insufficient nitrogen is taken up by the plants. Potential nitrogen uptake scales with root system area. Actual nitrogen uptake is the minimum of demand, determined by tissue N concentration, and potential uptake. Allocation of assimilates follows allometric rules, but water stress may limit LAI. Turnover of tree and soil components proceeds at constant relative rates.

The model is deterministic and is solved by Euler integration with a time step of one day. Typical run length ranges from 10,000 to 30,000 days (about 30 to 80 years).

Model parameters, inputs and outputs

None of the 39 parameters of BASFOR is known with great precision. Following a literature review of data on parameters, and partly based on the work of Levy et al. (2004), we defined likely ranges for the values of each parameter (Table 2). The ranges were deliberately set wide to allow sufficient scope for subsequent calibration on the basis of measurements of output variables. The main source of output data we used was the report on the CANIF project (Schulze 2000), in particular the data for the Norway spruce site at Skogaby, Sweden (56.22° N,

Table 1. State variables and parameters of the process-based forest model BASFOR.

	Trees	Description (units)	Soil	Description (units)
State variables	C_L, C_R, C_W	Carbon in leaves, roots and wood (kg C m ⁻²)	$C_{LITT}, C_{SOMF}, C_{SOMS}$	Carbon in litter and fast and slowly degrading soil organic matter (kg C m ⁻²)
	N_L	Nitrogen in leaves (kg N m ⁻²)	$N_{LITT}, N_{SOMF}, N_{SOMS}, N_{MIN}$	Nitrogen in litter, fast and slowly degrading SOM and mineral N (kg N m ⁻²)
Parameters	$C_{L,0}, C_{R,0}, C_{W,0}$	Initial values of state variables (kg C m ⁻²)	$C_{LITT,0}, C_{SOMF,0}, C_{SOMS,0}, N_{LITT,0}, N_{SOMF,0}, N_{SOMS,0}, N_{MIN,0}$	Initial values of state variables (kg C or N m ⁻²)
	$\beta, CO_{2,0}$	CO ₂ response (unitless, ppm)	$f_{LITT,SOMF}, f_{SOMF,SOMS}$	Efficiency of litter and organic matter transformation (kg kg ⁻¹)
	$f_{L,max}, f_W$	Allocation (kg kg ⁻¹)	$k_{d,LITT}, k_{d,SOMF}, k_{d,SOMS}$	Rate of litter and organic matter transformation (kg kg ⁻¹ day ⁻¹)
	Γ, LUE_0	Gas exchange (kg kg ⁻¹ , kg C MJ ⁻¹)		
	$k_{CA}, k_{CA,exp}, k_h, k_{h,exp}, SLA$	Morphology (m ² , unitless, m, unitless, m ² kg ⁻¹)		
	$k_{d,L}, k_{d,R}, k_{d,W}$	Turnover rates of leaves, roots and wood (kg kg ⁻¹ day ⁻¹)		
	k_L	Light extinction (m ² m ⁻²)		
	$k_{LAI,MAX}$	H ₂ O response (m ² m ⁻² mm ⁻¹)		
	$k_{N,MIN}, k_{N,upt}$	N uptake (kg N m ⁻² , kg N m ⁻² day ⁻¹)		
	$k_{T,A}, k_{T,B}$	Temperature response (°C ⁻¹ , °C)		
	$N_{L,com,min}, N_{L,con,max}, N_{R,con}, N_{W,con}$	Nitrogen concentrations (kg N kg ⁻¹ C)		

Table 2. Prior and posterior probability distributions of the 39 parameters of BASFOR. Details of parameters are given in Table 1. The prior distribution is multivariate uniform between bounds θ_{\min} and θ_{\max} . The posterior distribution is not analytical and is characterized here by the mean values of the parameters, $\bar{\theta}(i)$, the coefficients of variation (CV(i) = standard deviation/ $\theta(i)$), the parameters with which each parameter is correlated (underlined if negative) at greater absolute value than 0.3, and the maximum a posteriori estimates.

Parameter vector $\theta = [\theta(1) \dots \theta(39)]$			Prior probability distribution		Posterior probability distribution, using data from Skogaby			
$\theta(i)$	Symbol	Unit	$\theta_{\min}(i)$	$\theta_{\max}(i)$	$\bar{\theta}(i)$	CV(i)	Correlated $\{\theta(i)\}$	θ_{MAP}
1	$C_{L,0}$	kg m^{-2}	0.0001	0.010	0.00606	0.41	{6}	0.00651
2	$C_{R,0}$	kg m^{-2}	0.0001	0.010	0.00579	0.49	{25,30}	0.00668
3	$C_{W,0}$	kg m^{-2}	0.0001	0.010	0.00497	0.56		0.00405
4	β	unitless	0.4	0.6	0.50	0.12	{36}	0.41
5	$CO_{2,0}$	ppm	320	380	351.1	0.05	{25}	339.1
6	$f_{L,\max}$	unitless	0.25	0.35	0.305	0.10	{1,18,35,37}	0.333
7	f_W	unitless	0.52	0.62	0.587	0.04		0.614
8	Γ	unitless	0.4	0.6	0.46	0.10	{33}	0.42
9	k_{CA}	m^2	0.17	2.13	1.554	0.26		1.33
10	$k_{CA,\exp}$	m^2	0.33	0.77	0.585	0.20		0.76
11	$k_{d,L}$	day^{-1}	0.00013	0.00507	0.00026	0.44		0.00027
12	$k_{d,R}$	day^{-1}	0.000135	0.00054	0.00035	0.27	{27,33,38}	0.00040
13	$k_{d,W}$	day^{-1}	0.000027	0.000052	0.00004	0.18	{30,35}	0.00004
14	k_h	m	3.00	4.92	4.018	0.13	{15,21,29}	4.085
15	$k_{h,\exp}$	unitless	0.24	0.40	0.312	0.14	{14}	0.300
16	$k_{LAI,MAX}$	$\text{m}^2 \text{m}^{-2} \text{mm}^{-1}$	0.002	0.008	0.0069	0.12	{27}	0.0079
17	$k_{N,MIN}$	kg m^{-2}	0.0005	0.002	0.00121	0.35		0.00094
18	$k_{N,upt}$	$\text{kg m}^{-2} \text{day}^{-1}$	0.0005	0.002	0.00125	0.36	{6}	0.00186
19	$k_{T,A}$	$^{\circ}\text{C}^{-1}$	0.02	0.04	0.036	0.08		0.039
20	$k_{T,B}$	$^{\circ}\text{C}$	10	30	26.7	0.11		29.7
21	k_L	$\text{m}^2 \text{m}^{-2}$	0.43	0.64	0.544	0.11	{14}	0.522
22	LUE_0	kg MJ^{-1}	0.001	0.003	0.0027	0.08		0.0027
23	$N_{L,com,min}$	kg kg^{-1}	0.014	0.021	0.0157	0.09	{28}	0.0144
24	$N_{L,con,max}$	kg kg^{-1}	0.04	0.06	0.045	0.10	{29}	0.057
25	$N_{R,con}$	kg kg^{-1}	0.02	0.04	0.027	0.18	{2,5,38}	0.021
26	$N_{W,con}$	kg kg^{-1}	0.0005	0.002	0.00120	0.34		0.00126
27	SLA	$\text{m}^2 \text{kg}^{-1}$	2.0	28.2	8.06	0.36	{12,16}	7.06
28	$C_{LITT,0}$	kg m^{-2}	0.15	0.60	0.361	0.34	{23,39}	0.324
29	$C_{SOMF,0}$	kg m^{-2}	5	10	6.6	0.19	{14,24,30}	5.6
30	$C_{SOMS,0}$	kg m^{-2}	1	3	1.8	0.30	{2,13,29}	2.1
31	$N_{LITT,0}$	kg m^{-2}	0.005	0.02	0.0128	0.31		0.0084
32	$N_{SOMF,0}$	kg m^{-2}	0.2	0.4	0.27	0.19	{33,38}	0.22
33	$N_{SOMS,0}$	kg m^{-2}	0.05	0.20	0.135	0.29	{8,12,32,38}	0.166
34	$N_{MIN,0}$	kg m^{-2}	0.0001	0.0020	0.00105	0.54		0.00014
35	$f_{LITT,SOMF}$	unitless	0.4	0.8	0.61	0.17	{6,13,37}	0.45
36	$f_{SOMF,SOMS}$	unitless	0.01	0.10	0.055	0.47	{4}	0.041
37	$k_{d,LITT}$	day^{-1}	0.0007	0.0028	0.00165	0.40	{6,35}	0.00112
38	$k_{d,SOMF}$	day^{-1}	0.000028	0.000110	0.000050	0.27	{12,25,32,33}	0.000057
39	$k_{d,SOMS}$	day^{-1}	0.0000028	0.0000110	0.000006	0.39	{28}	0.000004

13.22° E) (Table 3). Weather data for Skogaby was collected on site (Schulze 2000) and uncertainty in these input data is not considered here. Additional data on tree height and biomass were taken from another Norway spruce site at Rajec in the Czech Republic (49.40° N, 16.63° E) (E. Klimo, Mendel University, Brno, Czech Republic, personal communication).

Our aim was to parameterize BASFOR based on the previously available information on possible parameter values taken from the literature, in conjunction with measurements of output variables taken at the Skogaby site. Both types of infor-

mation are uncertain, so a probabilistic calibration method is required for which the underlying theory is explained in the next section.

Bayesian calibration: theory

Quantifying uncertainty rather than maximizing fit

Calibration is inferring the values of parameters from data on model outputs. Unfortunately, information on output variables is never sufficiently accurate and comprehensive to allow ex-

Table 3. Data from Skogaby, Sweden, used for calibrating the model BASFOR. All data were taken from or based on information reported by Schulze (2000). The Skogaby Norway spruce experiment was planted in 1966. Only data from the control treatment (i.e., no fertilization or irrigation) are used here. All areal units refer to ground area unless otherwise indicated.

Variable	Measurement years	Values
Height (m)	1987, 1993	9.76, 11.94
Leaf N/C ratio ($\text{kg N kg}^{-1} \text{C}$)	1987, 1993, 1999	0.0234, 0.0228, 0.0204
Leaf carbon (kg C m^{-2})	1987, 1993, 1999	0.739, 0.975, 1.009
Stem carbon (kg C m^{-2})	1987, 1993, 1999	4.43, 6.48, 8.01
Root carbon (kg C m^{-2})	1987, 1993, 1999	0.919, 1.345, 1.641
Tree carbon (kg C m^{-2})	1987, 1993, 1999	6.09, 8.8, 10.66
Tree nitrogen (kg N m^{-2})	1987, 1993, 1999	0.0395, 0.0542, 0.0588
Soil carbon (kg C m^{-2})	1987, 1993	9.74, 9.04
Soil nitrogen (kg N m^{-2})	1987, 1993	0.446, 0.394
Soil mineral N (kg N m^{-2})	1987, 1993, 1998	0.0006, 0.0006, 0.0008
Leaf area index ($\text{m}^2 \text{ leaf m}^{-2}$)	1995	7
Soil N-mineralization rate ($\text{kg N ha}^{-1} \text{ year}^{-1}$)	1995	81
Net primary productivity ($\text{kg C m}^{-2} \text{ year}^{-1}$)	1995	0.95

act inference of parameter values. The best we can hope to achieve is less uncertainty in our knowledge of the parameters. We therefore seek a calibration procedure that updates probability distributions for parameters conditional on our data on output variables.

The aim of such a calibration procedure is different from parameter “tuning,” i.e., finding the best parameter values by comparing model outputs to data. The parameter vector selected by some such optimization methods may well be only one member of a set of equally well-fitting (or nearly as well-fitting) points, because process-based models tend to be over-parameterized with respect to data availability (Sievänen and Burk 1994, Hopkins and Leipold 1996). In probabilistic or Bayesian calibration, we go beyond identifying the best parameter vector, in whatever way we choose to define this notion, and instead determine a probability distribution over the parameters. Typically, we might summarize the probability distribution in the form of a mean vector and a variance matrix.

Bayes’ Theorem

In the Bayesian approach, the prediction for some uncertain quantity is taken to be the conditional distribution of that quantity based on data. We denote the data as D , which in our case comprise 30 observations on 13 different variables (Table 3). The particular values that these data take in our experiment are denoted as d . The uncertain quantity we wish to predict is denoted θ , which is a vector of 39 parameters (Table 2). Bayesian calibration involves computing the posterior distribution $p_{\theta|D}(t|d)$ for t denoting the possible values of θ . Often we simplify this notation and write just $p(\theta|D=d)$ or even $p(\theta|D)$. Bayes’ Theorem tells us how we can compute the posterior distribution from two other distributions. In the simplified notation, Bayes’ Theorem states:

$$p(\theta|D) = c p(D|\theta)p(\theta) \quad (1)$$

where $c = p(D)^{-1}$. The value c is fixed, and usually it is unnecessary to compute it explicitly. The factor $p(D|\theta)$ is termed

the likelihood function for θ , and the factor $p(\theta)$ the prior distribution for θ . These three terms in Bayes’ Theorem are commonly referred to as “the posterior,” “the prior” and “the likelihood.”

The Bayesian approach is sometimes termed “probabilistic inversion,” because it tells us how to infer the parameter θ from the data D in terms of how D is determined by the parameter θ . Given a candidate value for θ , we can evaluate the process model BASFOR, and so compute the model output at that θ . We can compare this output with the data D . The likelihood function thus “scores” each value for θ according to how well the process model at θ is able to reproduce the data D . A common assumption when choosing the likelihood is that θ , were it to be known, would provide a perfect match between the process model output and the true value from the system. In this case the observed differences between the two are attributable to measurement error in D . We will adopt this assumption below, although it may be generalized to include structural failures in the model (Craig et al. 2001, Kennedy and O’Hagan 2001, Goldstein and Rougier 2004, Higdon et al. 2004). In particular, we will assume that the measurement errors are additive, so that, writing $M(\theta)$ for the outputs of the process model:

$$p(D| \theta) = p(E = D - M(\theta)) \quad (2)$$

where E is the measurement error.

Choices for $p(\theta)$ and $p(E)$ are discussed below.

Bayesian calibration of BASFOR: implementation and results

The conceptual and the computational problem

For models that are analytically solvable, Bayes’ Theorem may, if the prior and likelihoods are sufficiently simple and compatible, lead to an analytical expression for the posterior probability distribution of the parameters. In contrast, process-based forest models like BASFOR are simulation mod-

els, which need to be run if we want to quantify the likelihood. The posterior must therefore be estimated in a non-analytical way. In this case, we summarize the posterior distribution in terms of a sample, from which we can then compute summary statistics like the posterior mean.

In the case of BASFOR, each element in the sample should be a vector of 39 values, representing the parameters of the model. The conceptual problem here is how to sample from (or summarize) the posterior distribution if we know only the density function up to a proportional constant, i.e., when we do not know c . This problem was first solved by Metropolis et al. (1953), who proved that a sample from the posterior can be constructed by an iterative procedure, using only the information embodied in the prior and data-likelihood. The procedure was inherently probabilistic and was therefore referred to as an example of Monte Carlo simulation. Metropolis' algorithm will be described below. However, there is a computational problem here as well, in that any algorithm for sampling representatively from a high dimensional space requires a large sample size. Moreover, calculating the data-likelihood requires evaluating the model at each sampled point in parameter space, and process-based models may be slow to run. These problems have hampered the application of Bayesian approaches to such models, but increases in computing power have made Monte Carlo simulation a feasible tool in recent years (Robert and Casella 1999).

We emphasize that use of the term "Monte Carlo" does not imply that we are calibrating a stochastic model. Although Monte Carlo methods are stochastic procedures for solving mathematical problems, those problems may be of any kind. In our case, the problem is the calibration of process-based forest models, and such models are typically deterministic. Terminological confusion may also arise from the fact that statisticians speak of Monte Carlo as a "simulation" technique, because the probability distributions are not constructed analytically, or by means of standard numerical techniques, but by simulating the sampling process. Monte Carlo as a simulation technique should thus not be confused with "simulation" in its more common meaning among forest modelers, i.e., "simulating" forest ecosystem processes within a PBM.

The Metropolis-Hastings algorithm

For our calibration calculations, we used the version of Markov Chain Monte Carlo (MCMC) known as the Metropolis-Hastings random walk (Robert and Casella 1999, p. 245). The idea is to walk through parameter space in such a way that the collection of all visited points forms the sample we are looking for. The starting point of the chain can be any point in parameter space, but the rule for moving from point to point must obey certain criteria, not discussed here (but see Robert and Casella 1999, Chapter 4). In our implementation of the method, we generate a proposal for a new candidate value for θ based on our existing value, namely:

$$\theta' = \theta_t + \varepsilon \quad (3)$$

where ε is a random vector and it is important that $p(\varepsilon) = p(-\varepsilon)$. Here, θ_t is our existing value, and θ' is the proposal. This proposal is accepted as the new value θ_{t+1} according to the ratio:

$$\beta = \frac{p(\theta' | D)}{p(\theta_t | D)} = \frac{p(D | \theta') p(\theta')}{p(D | \theta_t) p(\theta_t)} \quad (4)$$

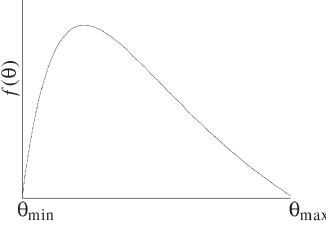
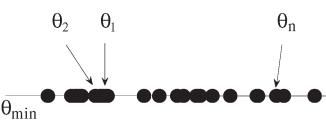
(taking the ratio cancels out the unknown c). We generate a uniform (0,1) random variable u , and we accept θ' as the new value for θ_{t+1} if $u \leq \beta$; otherwise, we set $\theta_{t+1} = \theta_t$. Note that if $\beta \geq 1$, i.e., if the posterior value for θ' is no lower than that for θ_t , we always accept the new point. However, we can also make a "downhill" step. In this respect, the algorithm is much like the simulated annealing approach to maximization. The major difference is that in simulated annealing, we make the acceptance criterion tougher and tougher over the course of the iteration process, and keep only the final value of θ , treating it as the maximum value. What the Metropolis-Hastings random walk shows is that by carefully selecting the acceptance criterion, we can treat the path itself as a sample from the posterior distribution, although it is necessary to discard some initial values (while the Markov chain "burns in").

For ε , we used a Gaussian distribution with a mean of zero. It is common to tune the variance matrix of ε to improve the way in which the sampled values of θ cover the space of possible values. In our specific example, we found that variances equal to the square of 5% of the prior parameter range and zero covariances gave good results. In generating a new candidate value for θ , we need not worry about jumping outside the range of values allowed by the prior lower and upper bounds of the parameters (Table 2), because such candidate points will have zero probability of being accepted. The key differences between MCMC methods of finding a posterior distribution and analytical methods are summarized in Table 4.

The prior probability distribution of the parameters

The prior probability distribution, $p(\theta)$, should reflect our knowledge about the parameters before calibration. This knowledge may be based on observation, literature study and expert opinion. Given little prior knowledge, it is generally better to choose the prior parameter distributions too wide than too narrow. As new data accumulate and are being processed in calibration, the posterior distribution will become increasingly dominated by the data, and the memory of the original prior distribution will fade. However, if the prior was used to set upper and lower bounds to the parameters, the posterior can never include values outside these bounds. Therefore we must not be too restrictive in our definition of the prior. On the other hand, if the data are not very informative (e.g., have a large measurement error), the posterior will not differ much from the prior, and we do not want to ignore information that could have been included in constructing the prior. For forest modeling, databases of parameters exist that can be used in defining the prior (Medlyn and Jarvis 1999).

Table 4. The posterior probability distribution $p(\theta|D)$ of a single, hypothetical parameter, θ , described analytically or approximated by Markov Chain Monte Carlo simulation.

Method	Symbolic representation	Graphical example	Expectation of a function g of the parameters, $E(g(\theta))$
Analytical approach	$p(\theta D) = f(\theta)$ where f is an analytical function, e.g., a Gaussian or Beta function		$\int_{\theta_{\min}}^{\theta_{\max}} g(\theta) f(\theta) d\theta$
Markov Chain Monte Carlo simulation	$p(\theta D) \rightarrow \{\theta_1, \dots, \theta_n\}$ where $\{\dots\}$ is a sample from the posterior distribution		$\frac{1}{n} \sum_{i=1}^n g(\theta_i)$

In the calibration of BASFOR, we use independent uniform distributions for the parameters with the upper and lower bounds as discussed earlier (Table 2), so the joint distribution is equal to their product. More information could have been included in the prior by using other distributions (e.g., beta or gamma); however, the simpler uniform distribution was chosen to emphasize the simplicity and feasibility of the Bayesian approach even for complex forest models.

The data-likelihood function

The likelihood function $p(D|\theta)$ is determined by the distribution of the measurement error, assumed to be additive. Where a multiplicative measurement error seems more appropriate, the outputs of the process model should be log transformed. Typically, this would be necessary where the output corresponds to a quantity that is strictly positive. We make the common assumption that the measurement errors are Gaussian and uncorrelated. Because we have no information about the precision of the measurements of the Skogaby data (Table 3), we choose the standard deviation of each measurement to be 30% of the mean value. This allows us to treat the errors additively, without transforming to logarithms. Consequently the likelihood becomes:

$$p(D|\theta) = p(E = D - M(\theta)) = \prod_{i=1}^{30} \varphi(D_i - M_i(\theta); 0, (0.3D_i)^2) \quad (4)$$

where the i -subscripts index the data and the BASFOR model output, and φ denotes a Gaussian probability density function with a given mean and variance. In the actual computations, all calculations should be performed in logarithms to avoid problems with finite precision arithmetic.

Calibration results for BASFOR

Before calibrating BASFOR, we determined prior predictive uncertainty on the model outputs by random sampling from the prior distribution (Figure 1). Prior uncertainty was large, as shown in the wide probability intervals (Figure 1). Sometimes the standard deviation for NPP and ground cover exceeded the

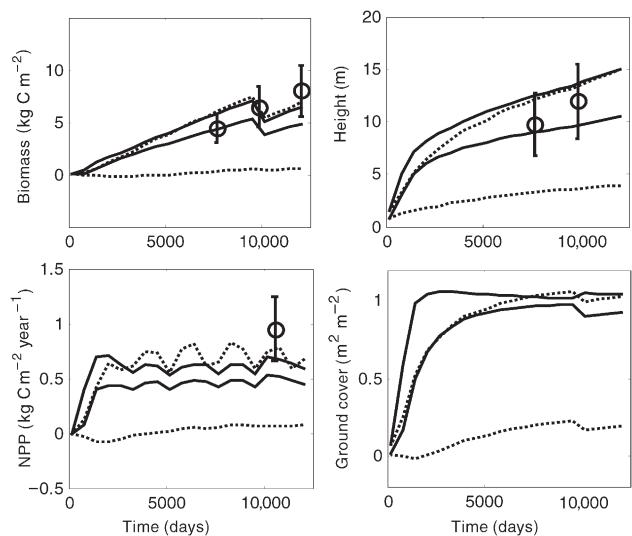


Figure 1. Prior and posterior uncertainty of four output variables of the forest model BASFOR. Pairs of lines show one standard deviation above and below the mean output from 10^5 model runs that simulated 33 years of Norway spruce growth at Skogaby, Sweden. Dotted lines denote prior uncertainty, and solid lines denote posterior uncertainty. Circles with vertical line segments denote data values \pm standard deviation. Posterior uncertainty was determined by Markov Chain Monte Carlo calibration based on data for 13 output variables (Table 3), including the data shown here for biomass, tree height and net primary productivity (NPP). Data were unavailable for the fourth variable, ground cover.

mean, indicating that the distributions were positively skewed. The coefficient of variation at the end of the runs, CV_{end} , averaged 0.68 over all 22 output variables of the model. Note that we were not using CV_{end} as a statistic of goodness-of-fit, but solely as a measure of uncertainty, which is what our calibration procedure aimed to quantify.

Then the MCMC was run, with a chain length of 10^5 , producing a sample of 10^5 parameter vectors from the posterior distribution, with the same number of corresponding simulated time series for each output variable. Figure 2 shows examples of chain trace plots for two of the parameters. Some values at the beginning of the chain—10% is often found to be adequate—need to be discarded: they represent the “burn-in,” while the chain adjusts to the appropriate conditional distribution (Gilks et al. 1996). Note that the trace plots do not converge to specific values of the parameters but maintain an irregular pattern. This is as it should be in MCMC: the parameter values continue to fill out the distribution of possible values rather than approaching a specific optimum. The aim is to quantify the joint probability distribution of the parameters rather than search for the best fit. In Figure 2, histograms of the parameter values in the traces are shown on the right. These represent the marginal distributions of the parameters. Only one of the two parameters shown, SLA, has a sharply peaked distribution. Apparently, the information in the data was insufficient to strongly reduce the uncertainty with respect to the other parameter, the light extinction coefficient, k_L . Posterior predictive uncertainty was much reduced (Figure 1), with an average CV_{end} of 0.22.

Summarizing the posterior

The MCMC procedure for Bayesian calibration does not provide an analytical formula for the posterior probability distribution of the parameters, but a large sample of values from the

posterior. The sample can be used directly to calculate expectation values of model outputs (Table 4, last column on the right), but it is awkward to handle when the posterior needs further updating as additional data are obtained, or when we want to pass on the model and its associated parameter distribution to others. We must therefore summarize the posterior distributions of both the parameters and the model predictions. We can do this by calculating the mean vector and variance matrix, and by presenting the marginal distributions graphically. The latter allows us to assess quickly which components of the parameter vector we have been able to learn about based on the data, D . The mean and variance are unlikely to be sufficient for the posterior distribution of the parameters, unless the quantity of data is large. We therefore suggest adding the maximum a posteriori estimate of θ (θ_{MAP}), which is the value of the parameter vector at which the posterior probability distribution has its maximum. The maximum a posteriori estimate of θ may be thought of as a measure of the single “best” parameter value. This can be approximated from within the MCMC sample or better, the sample approximation could be used as the starting point for a brief numerical optimization of $p(\theta|D)$ over θ . Posterior means, standard deviations and θ_{MAP} for BASFOR are given in Table 2. The full variance matrix is not shown, although the strongest correlations are listed. The differences between means and θ_{MAP} values are mostly small but not insignificant.

If we are dissatisfied with summarizing the posterior in the form of a variance–covariance matrix, more sophisticated methods of multivariate density estimation are available (Robert and Casella 1999, Sköld and Roberts 2003).

Evaluating how well the MCMC represents the posterior

We showed how BASFOR was calibrated by means of MCMC, running a single chain of length 10^5 . The chain length

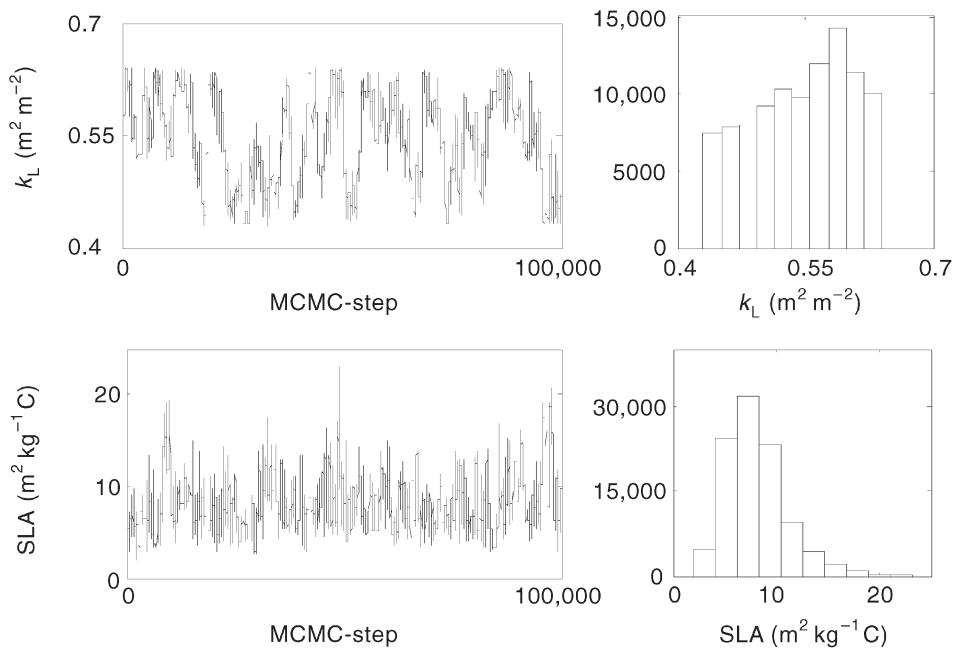


Figure 2. The Markov Chain Monte Carlo trace plots (left) and posterior marginal parameter probability distributions for 2 of the 39 parameters of model BASFOR: light extinction coefficient (k_L) and specific leaf area (SLA).

chosen should be such that the chain as a whole adequately captures the posterior distribution. We decided that the chain was long enough by visually inspecting the parameter trace plots (as those shown in Figure 2) to verify that all parameters had explored their range fully. More formal criteria for diagnosing chains exist, some of which are based on running multiple chains in parallel and verifying that the different chains have reached the same distribution (Robert and Casella 1999). The starting points of these chains should be dispersed around the parameter space. Where lots of chains are envisaged, for example, where multiple processors can be used in parallel, a simple space-filling design such as a Latin Hypercube (Saltelli et al. 2000) can be used to select starting points.

Toward a Bayesian framework of modeling and measurement

Data on parameters versus data on output variables

The procedure of Bayesian calibration that we propose starts by gathering information about the parameters to define their prior probability distribution, followed by using data for the output variables in the actual calibration. Thus, data on parameters are used differently from data on output variables. Many of the parameters in process-based forest models are difficult or impossible to measure, in contrast to output variables, for which we usually have better data (Levy et al. 2004, Van Oijen et al. 2005). Moreover, the appropriate correlation structure among parameters can be difficult to determine from parameter measurement, whereas it is generated automatically by the Bayesian calibration. As the outputs of the model tend to be the variables that we are really interested in, a focus on mea-

suring these variables rather than on parameters is all the more appropriate. By contrast, a pool of common knowledge about the model parameters is likely to accumulate as similar process-based models are calibrated across a range of different instances, e.g., different forest sites.

Identification of the most useful output data to acquire

The simplest way to assess the impact of data on our predictive distribution of model outputs is to perform a statistical experiment in which we vary the data available and their measurement error (in the likelihood function). For example, by making measurement error smaller and then repeating the inferential calculations, we can investigate the benefit of investing in more precise measurements, in terms of reduced uncertainty about our predictions. We investigated the importance of measurements of tree height and NPP for BASFOR in this way (Figures 3 and 4). In each case, we redid the calibration of BASFOR, starting as before from the multivariate uniform distribution of parameters (Table 2), but with less output information being used for calibration.

In the first calibration variant, only the output data for height were used, and the data for the other 12 output variables (Table 3) were disregarded. Somewhat surprisingly, the use of just the two height growth data points changed the parameter distribution noticeably, and posterior predictive uncertainty was greatly reduced compared with prior uncertainty (Figure 3). The coefficient of variation at the end of the runs, CV_{end} , was reduced from its prior average across all 22 output variables of 0.68 to 0.47 based on height growth data only. We already showed previously that the use of all 13 types of output data re-

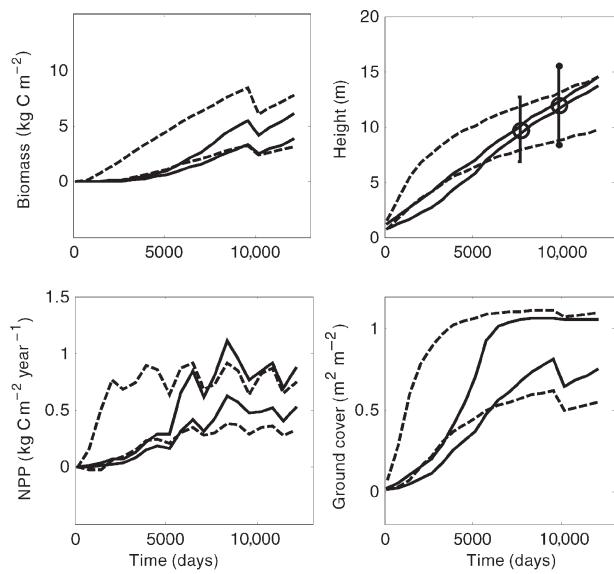


Figure 3. Two examples of calculated posterior uncertainty of the forest model BASFOR, arrived at based on two different sets of tree height data in the calibration. Dashed lines denote the model calibrated with only the height data from Table 3 (indicated by circles in the panel for height). Solid lines denote the model calibrated with the same height data but assuming higher precision of the measurements ($CV = 0.3\%$). Abbreviation: NPP = net primary productivity.

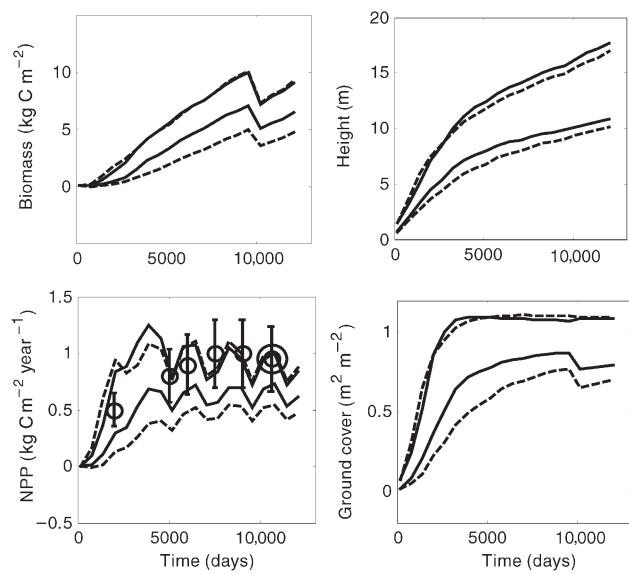


Figure 4. Two examples of calculated posterior uncertainty of the forest model BASFOR, arrived at based on two different sets of net primary productivity (NPP) data in the calibration. Dashed lines denote the model calibrated with only the NPP data from Table 3 (indicated by the large circle in the panel for NPP). Solid lines denote the model calibrated with additional hypothetical data on NPP, i.e., a longer time series of NPP measurement (small circles for NPP).

duced CV_{end} to 0.22. Height growth data are apparently very informative for BASFOR. In a second calibration variant, we again used only the two height data points, but this time assumed that they had been measured at far greater precision ($CV = 0.3\%$). These two precisely measured data points reduced predictive uncertainty almost as much as the use of 30 less precisely measured data points (cf. Figures 1 and 3, solid lines), with CV_{end} being reduced to 0.25.

In the next calibration variant, we used only the output data for NPP, which reduced predictive uncertainty by about as much as did the two height growth data ($CV_{end} = 0.44$; Figure 4). In the next calibration we again assumed that only NPP data were available but now assumed five hypothetical extra data points (Figure 4). This improved posterior predictive uncertainty further, but not by much ($CV_{end} = 0.39$; Figure 4).

Updating the probability distribution as data sets evolve

Bayes' Theorem offers a means for updating probability distributions, and the theorem can be applied again and again whenever new data become available. When a model is calibrated for the first time, the prior probability distribution of the parameters tends to be based on literature data, as we described above for BASFOR. However, after the first calibration, our knowledge about the parameters is captured in the posterior distribution which can function as the prior in the next calibration round, and so on. Bayesian statistics thus offers a natural mechanism for sequential "training" of a model. In this sequential procedure, the order in which data arrive is irrelevant: the posterior will be the same as long as in the end the same

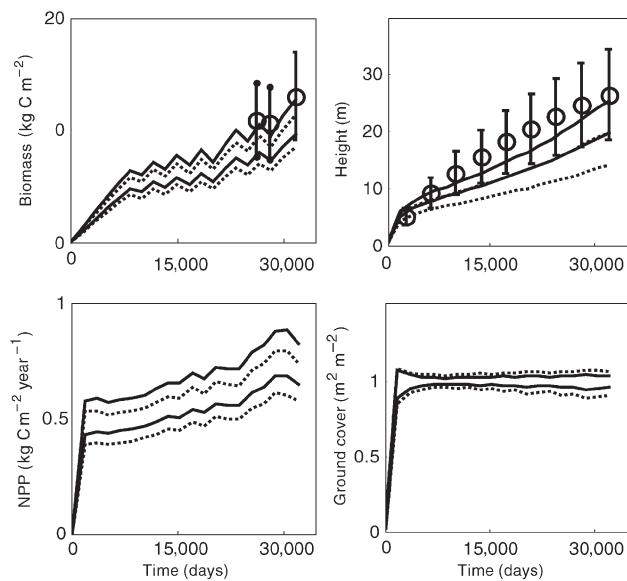


Figure 5. Prior and posterior uncertainty of the model BASFOR when applied to a new site (Rajec, Czech Republic; biomass and height data indicated by circles). Pairs of lines show one standard deviation above and below the mean output from 10^5 model runs that simulated 88 years of spruce growth. Dotted lines denote prior uncertainty, calculated using the parameter distribution previously derived using the data from Skogaby, Sweden (Table 2). Solid lines denote posterior uncertainty after further calibration based on the biomass and height data from Rajec. Abbreviation: NPP = net primary productivity.

data have been processed (Savia 1996, Jaynes 2003). This facility is especially important in the context of forest modeling where long-term monitoring programs generate data from year to year, which can be used immediately for further updating of model parameter distributions.

As an example of sequential calibration, we ran BASFOR, calibrated for Skogaby, Sweden, for a Norway spruce site at Rajec in the Czech Republic (49.40° N, 16.63° E). Without further calibration, the model results underestimated local measurements of height and stem biomass (Figure 5; data provided by E. Klimo, Mendel University, Brno, Czech Republic, personal communication), but after calibration with these data, model performance was improved (Figure 5).

Discussion

Multiple uses of Bayesian calibration

We have shown that Bayesian calibration can be applied to process-based forest models, even if they are of appreciable complexity. Bayesian calibration achieves several goals simultaneously: it provides parameter estimates with measures of uncertainty, and it quantifies the correlations among the parameter estimates. Because the calibration cannot be performed analytically, the method involves producing a large representative sample from the joint posterior probability distribution for the parameters. We showed, apparently for the first time in forestry research (Ghazoul and McAllister 2003), how Markov Chain Monte Carlo (MCMC) techniques can be used to produce the sample. The MCMC allows us to represent any posterior distribution whatever its shape, improving on older methods that only subdivide the parameter space into accepted and rejected regions (Mäkelä 1988). We chose the simplest MCMC algorithm (Metropolis et al. 1953). In our application with the BASFOR model, we found that 10^5 samples from the MCMC procedure gave adequate summaries of the posterior distribution of the parameters and the predicted model outputs. More advanced treatments are available (Robert and Casella 1999) that allow more efficient sampling of the parameter space, which would be important in cases where the simulator takes longer to evaluate. Whatever method is selected, we need to keep in mind that, because process-based models cannot be studied analytically, no sampling method can provide an exhaustive description of model behavior. The only reassurance we have is that adding extra parameters to a model does not imply that many more runs need to be made in the MCMC: as long as the number of parameters that strongly affect variables for which there are data remains the same, the required number of runs will also stay about the same (Jansen and Hagenaars 2004).

During the construction of the sample of the posterior parameter distribution by MCMC, the forest model needs to be run at each visited point in parameter space, so a corresponding sample of model outputs is constructed at the same time. The variation in this sample of model outputs represents model output uncertainty. In short, the Bayesian calibration procedure simultaneously quantifies the uncertainty in model parameters and in model predictions. These advantages of Baye-

sian calibration for process-based modeling have been outlined before (e.g., Jansen 1999, Jansen and Hagenaars 2004), in the context of crop modeling. Our contribution has been to demonstrate how the methodology can be implemented simply by means of MCMC, and to test it with a realistic forest model and real data.

Prior and posterior predictive uncertainty

The probability distribution of the parameters represents the information we have about them. Good information implies a narrow peaked probability distribution. The prior probability distribution for parameters, based only on the literature or direct measurement of the parameters, tends to be uninformative. Levy et al. (2004) searched the literature for information about the parameters of three process-based forest models (Century, BGC and Hybrid), but this led to predominantly wide, uninformative prior distributions for the parameters, and no information about correlations. They then sampled extensively from the prior distribution and ran the models with the sampled parameter values. As expected, the uncertainty in model predictions—for carbon sequestration in this case—was high (Levy et al. 2004). Because Levy et al. (2004) sampled from the prior probability distribution for the parameters, what they quantified was the “prior predictive uncertainty” of the models. Quantifying prior predictive uncertainty is relatively common, but it generally overstates uncertainty because it does not use all of the available information for parameterization. By contrast, Bayesian calibration allows use of all information, generally leading to narrower posterior parameter distributions and associated reduced “posterior predictive uncertainty,” as we have shown using BASFOR as a model example.

Identification of informative measurements

We have outlined several ways of identifying key parameters from the MCMC output, but conclude that, in parameterization, it may be more efficient to improve the availability of data on output variables rather than on the parameters themselves. The Bayesian approach allows us to quantify the improvements in prediction, measured, for example, in terms of a reduction in standard deviations, that follow from different types of investment in data collection. Having just a little data on tree height or NPP already reduced posterior predictive uncertainty significantly. Moreover, having data of greater precision, or longer data time series, gave further improvement. In general, these results are likely to apply widely across forest models, but the exact impact of data precision and time series length on model calibration will depend on the specific model used. The importance of data is a relative concept that can be defined only with respect to a given model.

Our demonstration that the calibration of model BASFOR was strongly affected by tree height data was at first surprising because of the marginal role that height plays in the model. In BASFOR, tree height is calculated by an allometric function of tree stem biomass. However, tree height itself affects none of the simulated processes in the model. Tree height growth was incorporated in BASFOR only because height is a frequently measured variable in forestry. Although height is

therefore only an output variable in the model, calibration based on precise height data affected the marginal probability distributions for almost all the parameters in the model, and thereby decreased the predictive uncertainty of most variables, not just height itself (see Figure 3). This demonstrates the strength of Bayesian inference: if we have good information about any output variable, we may infer much about the factors that directly or indirectly determine its behavior. The finding that precise height measurements were especially informative may be important given that different tree height measurement methods exist, with a trade-off between precision and expense (Hasenauer and Monserud 1997).

In conclusion, we suggest that variables like height, or any other frequently measured variables, be added to forest models for the benefit of model parameterization. With respect to tree measurement and forest monitoring programs, it may be worth increasing the precision of some measurements, despite the cost, when the added precision enhances the usefulness of the models.

Model comparison and model complexity

Process-based forest models tend to be complex, simulating many different processes and feedback mechanisms. The merits of models have often been debated in terms of their complexity, with some arguing that nature is complex, so models have to be too, and others arguing that model complexity may be disproportionate to the data available for parameterization and testing. These debates have often been sterile because of the absence of a criterion for model plausibility that is quantifiable, as opposed to the subjective concept of complexity (Jaynes 2003). Forest model comparisons have tended to focus on model structure while ignoring model performance (Perruchoud and Fischlin 1995). However, model performance can be evaluated using Bayes’ Theorem. Bayes’ Theorem is no more than a means for updating probability distributions and, as we have shown, can be applied to the problem of parameter probability. Furthermore, it can be applied in any context, including that of model comparison. The theorem may help determine which of a set of models is most probable in view of the data and prior information, and how strongly it is supported relative to the alternative models in the set (Bretthorst 1988). In such a case, we need to specify the prior probability $p(M_j)$ for each model in the set of models under consideration. We must then specify the likelihood of the data given a model, $p(D|M_j)$. Having done so, Bayes’ Theorem allows us to calculate the posterior probability of model M_j , $p(M_j|D)$. Bayesian calibration of models thus progresses in a natural way to Bayesian evaluation of models, where the main criterion is the posterior probability of models given the data. Bayes factors can be quantified for pairwise comparison of the posterior probability of models (Kass and Raftery 1995). Bayes factors penalize models if, given their parameter distribution, mean data-likelihood is low. Complexity only plays a role in that high-dimensional parameter distributions tend to be poorly supported by data, allowing many parameter values that lead to unrealistic model behavior and low data-likelihood. However, this is a matter for quantification, not debate.

We do not suggest that Bayesian comparison should be used to select one forest model from those currently available and reject the alternatives. As long as ecology has not come up with hard laws, forest model development will continue, and will benefit from the existence of multiple models to learn from. However, we do suggest that Bayesian comparison be used to decide, at any given time, which of the currently available models is most likely to provide good answers to particular forestry-related questions.

Prediction

When forest models are used for prediction, we need to take into account that environmental conditions, such as atmospheric CO₂ concentration, may change over the long lifetime of trees. This poses problems for parameterization if no data have been gathered under relevant conditions. A model might, for example, have a CO₂-dependent multiplier of NPP that at current CO₂ concentrations has the value unity, but with a doubling in CO₂ concentration may be expected to be anywhere between 1.0 and 2.0. Bayesian calibration cannot reduce the uncertainty in such a parameter without appropriate data, in which case the posterior distribution for the parameter will be the same as the prior. We might therefore decide to replace the parameter with a more mechanistic sub-model, but in that case, we will have to find probability distributions for all new parameters involved. This may or may not reduce predictive uncertainty; however, this can be determined by the Bayesian approach.

Communicating model results and models

In the environmental sciences, quantifying the uncertainty associated with model predictions has become a central issue (Clark et al. 2001). Bayesian calibration offers a logical means of quantifying uncertainty, grounded in the basic rules of probability theory. In this paper, we focused on showing how a forest modeler can work within a Bayesian framework. Other advantages of the use of Bayesian statistics in forestry, such as the more intuitive meaning of Bayesian probability intervals compared with classical statistical confidence intervals, have been explained by Ghazoul and McAllister (2003), who focused on the way Bayesian statistics can aid in communicating the results of forest research to a wider audience.

A consequence of the probabilistic way of working is that we should always provide probability intervals when we report model predictions. Moreover, when passing our models to colleagues, we should not only provide them with default values of each parameter, but with the full joint probability distribution of the parameters that we use. Only in this way is parameter uncertainty effectively communicated.

Conclusions

1. Bayesian calibration generates the posterior probability distribution of model parameters, and can be applied to all types of process-based forest models.

2. Bayesian calibration is highly informative: it quantifies parameter uncertainty, correlations between parameters and output uncertainty.
3. Bayesian calibration can be implemented easily and efficiently in the form of Markov Chain Monte Carlo (MCMC) algorithms, as demonstrated using a realistic forest model with real data.
4. Key output variables to be measured can be identified by recalibrating with different data.
5. In the case of forest model BASFOR, key output data included precisely measured tree heights and long time series of NPP.
6. Bayesian calibration is part of a larger framework of probabilistic modeling, in which not only the probability of parameters can be assessed, but also that of models.

Acknowledgments

We thank Emil Klimo (Mendel University, Brno, Czech Republic) for the use of data from his Norway spruce site at Rajec. Ronnie Milne (CEH-Edinburgh, U.K.) is acknowledged for his comments on an earlier version of this paper.

References

- Bernardo, J.M. and A.F.M. Smith. 1994. Bayesian theory. Wiley, Chichester, U.K., 608 p.
- Brethorst, G.L. 1988. Bayesian spectrum analysis and parameter estimation. *In* Lecture Notes in Statistics. Springer-Verlag, Berlin, 209 p.
- Clark, J.S., S.R. Carpenter, M. Barber et al. 2001. Ecological forecasts: an emerging imperative. *Science* 293:657–660.
- Craig, P.S., M. Goldstein, J.C. Rougier and A.H. Seheult. 2001. Bayesian forecasting for complex systems using computer simulators. *J. Am. Stat. Assoc.* 96:717–729.
- Ellison, A.M. 1996. An introduction to Bayesian inference for ecological research and environmental decision-making. *Ecol Appl.* 6:1036–1046.
- Ellison, A.M. 2004. Bayesian inference in ecology. *Ecol. Lett.* 7: 509–520.
- Gardner, R.H., V.H. Dale and R.V. O'Neill. 1990. Error propagation and uncertainty in process modeling. *In* Process Modeling of Forest Growth Responses to Environmental Stress. Eds. R.K. Dixon, R.S. Meldahl, G.A. Ruark and W.G. Warren. Timber Press, Portland, OR, pp 208–219.
- Gertner, G.Z., S.F. Fang and J.P. Skovsgaard. 1999. A Bayesian approach for estimating the parameters of a forest process model based on long-term growth data. *Ecol. Model.* 119:249–265.
- Ghazoul, J. and M. McAllister. 2003. Communicating complexity and uncertainty in decision-making contexts: Bayesian approaches to forest research. *Int. For. Rev.* 5:9–19.
- Gilks, W.R., S. Richardson and D.J. Spiegelhalter. 1996. Introducing Markov chain Monte Carlo. *In* Markov Chain Monte Carlo in Practice. Eds. W.R. Gilks, S. Richardson and D.J. Spiegelhalter. Chapman & Hall, London, pp 1–19.
- Goldstein, M. and J.C. Rougier. 2004. Probabilistic formulations for transferring inferences from mathematical models to physical systems. *SIAM J. Sci. Comput.* 26:467–487.
- Green, E.J., D.W. MacFarlane, H.T. Valentine and W.E. Strawderman. 1999. Assessing uncertainty in a stand growth model by Bayesian synthesis. *For. Sci.* 45:528–538.

- Green, E.J., D.W. MacFarlane and H.T. Valentine. 2000. Bayesian synthesis for quantifying uncertainty in predictions from process models. *Tree Physiol.* 20:415–419.
- Hasenauer, H. and R.A. Monserud. 1997. Biased predictions for tree height increment models developed from smoothed “data.” *Ecol. Model.* 98:13–22.
- Higdon, D., M. Kennedy, J. Cavendish, J. Cafeo and R.D. Ryne. 2004. Combining field data and computer simulations for calibration and prediction. *SIAM J. Sci. Comput.* 26:448–466.
- Hopkins, J.C. and R.J. Leipold. 1996. On the dangers of adjusting the parameter values of mechanism-based mathematical models. *J. Theor. Biol.* 183:417–427.
- Jansen, M.J.W. 1999. Data use and Bayesian statistics for model calibration. In *Data and Models in Action*. Eds. A. Stein and F.W.T. Penning de Vries. Kluwer, Dordrecht, pp 69–80.
- Jansen, M.J.W. and T.J. Hagenaars. 2004. Calibration in a Bayesian modelling framework. In *Bayesian Statistics and Quality Modelling in the Agro-Food Production Chain*. Eds. M.A.J.S. Van Boekel, A. Stein and A.H.C. Van Bruggen. Kluwer, Dordrecht, pp 47–55.
- Jaynes, E.T. 2003. Probability theory: the logic of science. Cambridge University Press, Cambridge, 758 p.
- Kass, R.E. and A.E. Raftery. 1995. Bayes factors. *J. Am. Stat. Assoc.* 90:773–795.
- Kennedy, M.C. and A. O’Hagan. 2001. Bayesian calibration of computer models. *J. Roy. Stat. Soc. B Met.* 63:425–450.
- Levy, P.E., R. Wendler, M. Van Oijen, M.G.R. Cannell and P. Millard. 2004. The effects of nitrogen enrichment on the carbon sink in coniferous forests: uncertainty and sensitivity analyses of three ecosystem models. *Water Air Soil Pollut.* 4:67–74.
- Mäkelä, A. 1988. Performance analysis of a process-based stand growth model using Monte Carlo techniques. *Scand. J. For. Res.* 3:315–331.
- Mäkelä, A., J. Landsberg, A.R. Ek, T.E. Burk, M. Ter-Mikaelian, G.I. Ågren, C.D. Oliver and P. Puttonen. 2000. Process-based models for forest ecosystem management: current state of the art and challenges for practical implementation. *Tree Physiol.* 20: 289–298.
- Medlyn, B.E. and P.G. Jarvis. 1999. Design and use of a database of model parameters from elevated CO₂ experiments. *Ecol. Model.* 124:69–83.
- Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller. 1953. Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21:1087–1092.
- Perruchoud, D.O. and A. Fischlin. 1995. The response of the carbon cycle in undisturbed forest ecosystems to climate change: a review of plant–soil models. *J. Biogeog.* 22:759–774.
- Prisley, S.P. and M.J. Mortimer. 2004. A synthesis of literature on evaluation of models for policy applications, with implications for forest carbon accounting. *For. Ecol. Manage.* 198:89–103.
- Radtke, P.J., T.E. Burk and P.V. Bolstad. 2002. Bayesian melding of a forest ecosystem model with correlated inputs. *For. Sci.* 48: 701–711.
- Reckhow, K.H. 1990. Bayesian inference in non-replicated ecological studies. *Ecology* 71:2053–2059.
- Robert, C.P. and G. Casella. 1999. Monte Carlo statistical methods. Springer-Verlag, New York, 507 p.
- Robinson, A.P. and A.R. Ek. 2000. The consequences of hierarchy for modeling in forest ecosystems. *Can. J. For. Res.* 30:1837–1846.
- Saltelli, A., K. Chan and E.M. Scott. 2000. Sensitivity analysis. Wiley, Chichester, U.K., 475 p.
- Schulze, E.-D. 2000. Carbon and nitrogen cycling in European forest ecosystems. *Ecol. Stud.* 142. Springer-Verlag, Berlin, 500 p.
- Sievänen, R. and T.E. Burk. 1994. Fitting process-based models with stand growth data—problems and experiences. *For. Ecol. Manage.* 69:145–156.
- Sivia, D.S. 1996. Data analysis: a Bayesian tutorial. Oxford University Press, Oxford, 189 p.
- Sköld, M. and G.O. Roberts. 2003. Density estimation for the Metropolis-Hastings algorithm. *Scand. J. Stat.* 30:699–718.
- Van Oijen, M. 2002. On the use of specific publication criteria for papers on process-based modelling in plant science. *Field Crops Res.* 74:197–205.
- Van Oijen, M., G.I. Ågren, O. Chertov, S. Kellomäki, A. Komarov, D.C. Mobbs and M.B. Murray. 2005. Application of process-based models to explain and predict changes in European forest growth. In *Causes and Consequences of Forest Growth Trends in Europe—Results of the RECOGNITION Project*. Eds. T. Karjalainen and A. Schuck. Brill, Leiden, Boston, Köln. In press.
- Van Oijen, M., M.G.R. Cannell and P.E. Levy. 2004. Modelling biogeochemical cycles in forests: state of the art and perspectives. In *Towards the Sustainable Use of Europe’s Forests—Forest Ecosystem and Landscape Research: Scientific Challenges and Opportunities*. Eds. F. Andersson, Y. Birot and R. Päivinen. European For. Inst., Joensuu, Finland, pp 157–169.