Available online at www.sciencedirect.com
SCIENCE DIRECT®
Ecological Modelling 190 (2006) 287–298
ECOLOGICAL MODELLING
www.elsevier.com/locate/ecolmodel

A Bayesian strategy for combining predictions from empirical and process-based models

Philip J. Radtke[a,*], Andrew P. Robinson[b]

[a] Virginia Tech, Department of Forestry, Blacksburg, VA 24061, USA
[b] Department of Forest Resources, University of Idaho, Moscow, ID 83843, USA

Received 28 January 2004; received in revised form 1 April 2005; accepted 6 April 2005
Available online 14 June 2005

Abstract

We present a strategy for using an empirical forest growth model to reduce uncertainty in predictions made with a physiological process-based forest ecosystem model. The uncertainty reduction is carried out via Bayesian melding, in which information from prior knowledge and a deterministic computer model is conditioned on a likelihood function. We used predictions from an empirical forest growth model G-HAT in place of field observations of aboveground net primary productivity (ANPP) in a deciduous temperate forest ecosystem. Using Bayesian melding, priors for the inputs of the process-based forest ecosystem PnET-II were propagated through the model, and likelihoods for the PnET-II output ANPP were calculated using the G-HAT predictions. Posterior distributions for ANPP and many PnET-II inputs obtained using the G-HAT predictions largely matched posteriors obtained using field data. Since empirical growth models are often more readily available than extensive field data sets, the method represents a potential gain in efficiency for reducing the uncertainty of process-based model predictions when reliable empirical models are available but high-quality data are not.

Keywords: Bayesian melding; Bayesian synthesis; Ecosystem modeling; Forest growth and yield modeling; Likelihood; Sampling importance resampling

1. Introduction

Simulation models support a wide variety of activities, ranging from decision-making to scientific research (see, e.g. Robinson and Ek, 2000). Models that accommodate these different objectives embody different values, and are constructed using different principles. A commonly applied classification for simulation models in biology divides them into *empirical* and *process-based* models. It is generally held that empirical models (EM) are statistically oriented, that is, their structure is chosen with a view to optimizing some objective function (e.g. Schabenberger and Pierce, 2002 pp. 95, 195), whilst the structure for

* Corresponding author. Tel.: +1 540 231 8863; fax: +1 540 231 3698.
E-mail address: pradtke@vt.edu (P.J. Radtke).

0304-3800/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2005.04.021

process-based models (PBM) is chosen to explicitly represent known processes in some way (e.g. Thornley and Johnson, 2000). Generally, the statistical properties of model fitting form the basis for inference from EMs, while the theoretical validity of PBMs provides their inference-base.

To the degree that they incorporate cause and effect relationships accurately, the domain of PBMs may extend to conditions outside the range of previous observation. This property makes them an attractive tool for forest management. In many forest systems, climate conditions or management inputs, e.g. tree genetics, irrigation and chemical applications, are changing faster than our ability to acquire data for EM development. Populations of interest are moving targets and EMs are continually aimed behind them. This circumstance provides a compelling argument for the wider use of PBMs. However, a common lament of forest managers is that PBMs embody too many uncertainties and rely on too many poorly known parameters to produce reliable output (Mäkelä et al., 2000). Mäkelä et al. (2000) propose that both types of models can be improved by incorporating the features that give strength to the other type. Attempts have already been made to blend these two approaches, sometimes leading to hybrid-type models. This has generally involved one of the following strategies:

- Using a root mean squared error weighted mean of the predictions as a prediction (Robinson, 1998).
- Embedding an established empirical model inside an established process-based model (Baldwin et al., 1993).
- Creating a new model and allowing both statistical and process criteria to dictate model structure and parameter values (Sievänen and Burk, 1993).
- Tuning or tweaking the parameters of an established process-based model until the output matches a certain set of objectives (Landsberg et al., 2001).
- Constraining process-model predictions using empirical yield tables (Waring and McDowell, 2002).

Here, we demonstrate a method for combining predictions from both empirical and process-based models without necessarily modifying the model structure of either. The method relies on a framework called Bayesian melding (BYSM; Poole and Raftery, 2000) and optimizes PBM predictions based on model inputs or outputs of interest. Optimization is based on marginal or joint Bayesian posterior distributions for inputs or outputs of interest. As such, the PBM inference base is expanded to include both the original cause–effect rationale and the foundations of Bayesian statistics.

While BYSM has been used primarily to assess uncertainty in deterministic PBM predictions (Raftery et al., 1995; Green et al., 1999), we will show how it can be used to link EM and PBM predictions. Previous applications of BYSM used sample data to calculate likelihoods in estimating Bayesian posterior densities (Green et al., 1999; Radtke et al., 2002). We will use the information from EM predictions in lieu of sample observations. The method does not preclude the use of sample data in calculating likelihoods, but allows for an available, often inexpensive source of information (EM predictions) to be used in reducing the uncertainty associated with PBM predictions.

1.1. Bayesian melding

Bayesian melding allows for the assessment of uncertainty in deterministic computer model predictions. Its output includes Bayesian posterior distributions for any model inputs or outputs of interest. The method was developed by Raftery et al. (1995) and Poole and Raftery (2000), and has subsequently been applied to the assessment of accuracy and uncertainty in forest model predictions (Green et al., 1999, 2000; Radtke et al., 2002). BYSM combines, or *melds*, information from three sources in deriving its estimates of Bayesian posterior distributions (Poole and Raftery, 2000). *Direct* information is observed directly on a population of interest. *Indirect* information is gathered from outside sources somehow related to the population of interest. *Model* information is the information contained within a deterministic computer model of interest. In the application discussed here, the model of interest is a forest ecosystem PBM (see Section 2). Direct and indirect information may pertain to either model inputs or outputs. Direct information typically involves a sample of observations made on the population of interest and is used to compute likelihoods. Indirect information is expressed as probability density functions (PDFs) that reflect knowledge and uncertainty about the various model quantities (MacFarlane et al., 2000; Radtke et al., 2001).