Environmental Modelling & Software 67 (2015) 138–148

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Data-intensive modeling of forest dynamics

Jean F. Liénard [a], Dominique Gravel [b], Nikolay S. Strigul [a,*]

[a] Department of Mathematics, Washington State University, Vancouver, Washington, USA
[b] Département de Biologie, Universit du Québec à Rimouski, Québec, Canada

ARTICLE INFO

Article history:
Received 12 January 2015
Accepted 19 January 2015
Available online

Keywords:
Data-intensive model
Forest dynamics
Gibbs sampling
Markov chain model
Markov chain Monte Carlo
Patch-mosaic concept
Plant population and community dynamics

ABSTRACT

Forest dynamics are highly dimensional phenomena that are not fully understood theoretically. Forest inventory datasets offer unprecedented opportunities to model these dynamics, but they are analytically challenging due to high dimensionality and sampling irregularities across years. We develop a data-intensive methodology for predicting forest stand dynamics using such datasets. Our methodology involves the following steps: 1) computing stand level characteristics from individual tree measurements, 2) reducing the characteristic dimensionality through analyses of their correlations, 3) parameterizing transition matrices for each uncorrelated dimension using Gibbs sampling, and 4) deriving predictions of forest developments at different timescales. Applying our methodology to a forest inventory database from Quebec, Canada, we discovered that four uncorrelated dimensions were required to describe the stand structure: the biomass, biodiversity, shade tolerance index and stand age. We were able to successfully estimate transition matrices for each of these dimensions. The model predicted substantial short-term increases in biomass and longer-term increases in the average age of trees, biodiversity, and shade intolerant species. Using highly dimensional and irregularly sampled forest inventory data, our original data-intensive methodology provides both descriptions of the short-term dynamics as well as predictions of forest development on a longer timescale. This method can be applied in other contexts such as conservation and silviculture, and can be delivered as an efficient tool for sustainable forest management.

Software and data availability

The software to estimate transition matrices based on forest inventory was implemented by Jean Liénard in R version 2.15.1 (R Core Team, 2012) and is attached as a zip file to the submission.

The database studied in this paper is available upon request to the Quebec provincial forest inventory database (http://www.mffp.gouv.qc.ca/forets/inventaire/). Straightforward modifications of the software allow to use with the USDA Forest Inventory and Analysis program (http://www.fia.fs.fed.us/) .

1. Introduction

Forest ecosystems are complex adaptive systems with hierarchical structures resulting from self-organization in multiple dimensions simultaneously (Levin, 1999). The patch-mosaic concept was actively developed in the second half of the twentieth century after Watt (1947) suggested that ecological systems can be considered a collection of patches at different successional stages. Dynamical equilibria arise at the level of the mosaic of patches rather than at the level of one patch. The classic patch-mosaic methodology assumes that patch dynamics can be represented by changes in macroscopic variables characterizing the state of the patch as a function of time (Levin and Paine, 1974). Forest disturbances are traditionally associated with a loss of biomass; however, Markov chain models based only on biomass do not capture forest succession comprehensively (Strigul et al., 2012). This limitation motivates the need for alternative formulations that are able to consider several forest dimensions instead of only one.

Here we develop a novel statistical methodology for estimating transition probability matrices from forest inventory data and generalize classic patch-mosaic framework to multiple uncorrelated dimensions. In particular, we develop a landscape-scale patch-mosaic model of forest stand dynamics using a Markov chain framework, and validate the model using the Quebec provincial forest inventory data. The novelty of our modeling framework lies in the consideration of forest transitions within multiple

* Corresponding author.
E-mail address: nick.strigul@vancouver.wsu.edu (N.S. Strigul).

http://dx.doi.org/10.1016/j.envsoft.2015.01.010

dimensional space of macroscopic stand-level characteristics (biomass, average age of trees, biodiversity and shade tolerance index) that constitutes a generalization of the one-dimensional model of forest biomass transitions developed earlier (Strigul et al., 2012). Our framework is also substantially distinct from previous models of forest dynamics, where successional stages are ordinated using empirical observations on successional pathways (Curtis and McIntosh, 1951; Kessell and Potter, 1980; Logofet and Lesnaya, 2000).

The Quebec forest inventory (Perron et al., 2011) is one of the extensive forest inventories that have been established in North America, among others led by the Canadian provincial governments and the USDA Forest Inventories and Analysis program in the USA. These inventories provide a representative sample of vegetation across the landscape through a large number of permanent plots that are measured repeatedly. Although they were originally developed for estimating growth and yield, they were rapidly found to be extremely useful to studies in forest ecology, biogeography and landscape dynamics. Each permanent plot consists of individually marked trees that are periodically surveyed and remeasured. Each plot can be considered as a forest stand and then, theoretically, the forest inventories provide empirical data sufficient for parametrization and validations of patch-mosaic models (Strigul et al., 2012). However, practical development of the patch-mosaic forest models (i.e. their parametrization, validation and prediction) is challenging due to the underlying structure of the forest inventory datasets. These datasets are indeed collected at irregular time intervals that are not synchronized across the focal area, and data collection procedures including spatial plot design and tree measurement methods can be different at various survey times and conducted by different surveyors (Strigul et al., 2012).

Our objective in this study is to develop a data-intensive method predicting the dynamics of forest macroscopic characteristics. The idea of a data-intensive modeling approach is to develop and explore a quantitative theory using statistical modeling, in contrast with the hypothesis-driven theoretical approach in which selected mechanisms are used to design and constrain models. We focus here on the development of the modeling framework and illustrate the application of the framework to a large forest inventory dataset spanning 38 years of observations collected in Quebec. To overcome the issue of irregular samplings in time specific to forest inventory data, we develop a Gibbs sampling procedure for augmenting the data and infer the transition probabilities. Our particular use of Gibbs sampling (Pasanisi et al., 2012) has a substantial scientific novelty, as this is the first application of this statistical machinery to overcome the problems of irregularities in the forest inventory sample design. In this paper, we demonstrate the power of this statistical methodology in our application, and deliver it as ready-to-go tool for other applications by explaining every step, providing pseudocode, and original R code. We anticipate that this novel statistical methodology will be broadly used in forest inventory analysis as the issue of irregularities in inventories has previously been a substantial hindrance (e.g. Strigul et al., 2012).

We present in this paper the general methodology and demonstrate each of its steps on the Quebec dataset. In particular, we consider the dimensionality of stand characteristics in this dataset and present evidence that some characteristics are redundant. We apply the method to predict long-term dynamics of Quebec forests, as represented by a subset of macroscopic properties that best represent the variability in the data. We validate the model utilizing two different cross-validation schemes to split the original data, based on survey date (predicting later years using earlier years) and based on a random 2-folds partition of plots (comparing long-term predictions inferred from two independent subsets). We finally discuss the implications of this work, such as the effect of spatial and temporal variability, the independence of most forest variables, the effect of changing external drivers and of feedbacks.

2. Patch-mosaic modeling framework

The goal of this section is to introduce the modeling of patch-mosaic using Markov chains, which is generalized and employed to predict forest dynamics in the main text. The patch-mosaic concept assumes that the vegetation at the landscape level can be represented as a collection of isolated spatial units – patches – where patch development follows a general trajectory and is subject to disturbances (Watt, 1947; Levin and Paine, 1974). Patch-mosaic models are derived using the conservation law, which takes into account patch aging and other changes to macroscopic variables representing succession, growth of patches in space, and disturbances (Levin and Paine, 1974). The same general idea as well as mathematical derivations are broadly used in population dynamics to describe age- and size-structured population dynamics. Patch-mosaic models can be partial differential equations or discrete models depending on whether time and patch stages are assumed to be continuous or discrete (Strigul, 2012). Classic continuous patch-mosaic models are based on the application of the conservation law to continuously evolving patches that can be destroyed with a certain probability, and can be represented by the advection equation (model developed by Levin and Paine, 1974, for fixed-size patches) or equivalently by the Lotka–McKendick–von Foerster model (Strigul et al., 2008). The continuous patch-mosaic models have been used in forest ecology to model the dynamics of individual canopy trees within the stand or forest gap dynamics (Kohyama et al., 2001; Kohyama, 2006).

In the case of patches changing in discrete time, the derivation of the conservation law leads to discrete-type patch-mosaic models. In particular, the advection-equation model (Levin and Paine, 1974) is essentially equivalent to several independently developed discrete models (Leslie, 1945; Feller, 1971; Van Wagner, 1978; Caswell, 2001; Strigul, 2012). These models consider only large scale catastrophic disturbances (patch “death” process), destroying the patch, which then develops along the selected physiological axis until the next catastrophic disturbance Levin and Paine (1974). The stochastic model we are considering here employs a Markov chain framework (Waggoner and Stephens, 1970; Usher, 1979; Facelli and Pickett, 1990; Logofet and Lesnaya, 2000; Caswell, 2001) that is capable of taking into account all possible disturbances.

In a Markov chains model, the next state of a forest stand depends only on the previous state, and the probabilities of going from one state into another are summarized in what is called a transition matrix, denoted T.

We summarize the distribution of states at time t as the row vector X_t, with length equal to the number of discrete classes of patch state and with a sum equal to 1. We can predict $X_{t+\Delta t}$ by multiplying the transition matrix:

$$X_{t+\Delta t} = X_t \cdot T \quad (1)$$

To project an arbitrary number n time steps into the future, one simply multiplies by T^n instead of T. The Perron–Frobenious Theorem guarantees the existence of the long-term equilibrium, which can be practically found as the normalized eigenvector corresponding to the first eigenvalue, or by iterative sequence of state vectors. In this paper we employ the iterative method as it allows to derive forest states at different time steps in the future, for example allowing to make predictions in 10, 20 or 30 years from now. To derive the long-term equilibrium we simply choose an n large enough to satisfy the condition: