Combining a generic process-based productivity model and a statistical classification method to predict the presence and absence of tree species in the Pacific Northwest, U.S.A.

Nicholas C. Coops [a,*], Richard H. Waring [b], Todd A. Schroeder [a]

[a] *Department of Forest Resource Management, 2424 Main Mall, University of British Columbia, Vancouver, Canada V6T 1Z4*
[b] *College of Forestry, Oregon State University, Corvallis, OR 97331, United States*

ARTICLE INFO

ABSTRACT

Although long-lived tree species experience considerable environmental variation over their life spans, their geographical distributions reflect sensitivity mainly to mean monthly climatic conditions. We introduce an approach that incorporates a physiologically based growth model to illustrate how a half-dozen tree species differ in their responses to monthly variation in four climatic-related variables: water availability, deviations from an optimum temperature, atmospheric humidity deficits, and the frequency of frost. Rather than use climatic data directly to correlate with a species' distribution, we assess the relative constraints of each of the four variables as they affect predicted monthly photosynthesis for Douglas-fir, the most widely distributed species in the region. We apply an automated regression-tree analysis to create a suite of rules, which differentially rank the relative importance of the four climatic modifiers for each species, and provide a basis for predicting a species' presence or absence on 3737 uniformly distributed U.S. Forest Services' Forest Inventory and Analysis (FIA) field survey plots. Results of this generalized rule-based approach were encouraging, with weighted accuracy, which combines the correct prediction of both presence and absence on FIA survey plots, averaging 87%. A wider sampling of climatic conditions throughout the full range of a species' distribution should improve the basis for creating rules and the possibility of predicting future shifts in the geographic distribution of species.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

A region's flora and fauna reflect the interplay of dispersal, colonization, and competition for resources under a specific range of environments. Within the Pacific Northwest (PNW) region of the United States, the distribution of the flora is well described by Franklin and Dyrness (1973) in terms of temperature and precipitation patterns, physiography, and associated plant communities. These descriptions, however, lack predictive power, and this deficiency, in a region where the climate may already be changing (Mote et al., 2005; Westerling et al., 2006), makes plans for conservation, as well as exploitation of natural resources, highly uncertain. In the extreme, climatically induced disturbance might cause major structural transformations from one type of vegetation to another, and through changes in the energy balance, further alter the region's climate (Pielke et al., 1998).

At present, there are two divergent approaches that incorporate climatic information to predict the distribution of species. One approach relies on empirical correlations while the other attempts to acquire a mechanistic understanding on which to base predictions. The first, and most widely accepted approach, consists of "niche" or "bioclimatic envelope" models (Austin, 1985; Iverson and Prasad, 1998; McKenzie et al., 2003; Thuiller et al., 2008). Such models usually relate presence/absence data empirically to environmental variables, most often climate (but sometimes including soil and physiographic features), using an array of statistical methods including multiple regression techniques, neutral networks, and regression-tree analysis (Iverson and Prasad, 2001). The capacity of these empirical models to provide accurate predictions of species' distributions under future, possibly novel climatic combinations is unclear (Williams et al., 2007).

At the other extreme are mechanistic models that predict the growth of individual species or even clones under any specified environment (Sands et al., 2000; Rodriguez et al., 2002; Almeida et al., 2004; Dye et al., 2004). The advantage of such mechanistic, process-based models is that they identify the relevant environmental constraints on growth and other processes. Such models are specifically designed to be able to predict performance of a species outside its present natural range (Waring, 2000; Coops et al., 2005; Waring et al., 2008). Their disadvantage is that detailed

* Corresponding author. Tel.: +1 604 822 6452; fax: +1 604 822 9106.
E-mail address: nicholas.coops@ubc.ca (N.C. Coops).

information is required by these models to define a species' tolerance and response to deviations from optimum temperature, frost, drought, and atmospheric humidity deficits, and in how resources are partitioned to leaves, roots, stems, and branches.

We questioned whether it might be possible to combine the automated, statistically sophisticated component of empirical models with the process-based understanding imbedded in the mechanistic type models. To address this question, we recognize that we must first simplify the mechanistic approach by referencing environmental responses of any number of species to one that is widely distributed. We also know that climatic data must be extrapolated across landscapes in an appropriate form and at a spatial resolution that match model requirements and the availability of biological information recorded on ground-based field survey plots. To automate the process of seeking rules to define the distribution of different tree species, we chose regression-tree analysis because of its efficiency and transparency in recognizing those physiological variables and their thresholds that separate one tree species from another in its adaptation to environment. Based on the analysis of a half-dozen species, the results of this hybrid approach were sufficiently encouraging to share, although we recognize the need to expand the analysis to include the full environmental range that each species now occupies.

2. Methods

2.1. Hybrid model

All ecosystem process-based models are simplified versions of reality with the choice of which model to utilize dependent a number of factors including the minimum spatial and temporal units of operation and the number and type of output parameters (Nightingale et al., 2004). Likewise the scale at which the model operates (leaf–tree, plot–stand, regional and ecosystem levels) is also critical, with model complexity generally decreasing as the time step and spatial extent of model operation increases (Wulder et al., 2007). Given the need to predict species distribution over large spatial extents we believe a monthly time step, stand-level, process-based model is an appropriate choice for our analysis. Within this specification a number of process-based models exist (Nightingale et al., 2004) including HYBRID (Friend et al., 1997), FOREST-BGC (Running and Coughlan, 1988), BIOME-BGC (Running and Hunt, 1993) amongst others.

The 3-PG model (physiological principles predicting growth) was selected as a basis for the test because it contains a number of simplifying assumptions that have emerged from studies conducted over a wide range of forests (Landsberg et al., 2003). These include:

- Climatic data can be summarized at monthly intervals with little loss in the accuracy of model predictions.
- Each month, the most limiting climatic variable on photosynthesis is selected, based on departure from conditions that are optimum (expressed as unity) or completely limited (expressed as zero).
- Maximum canopy stomatal conductance approaches a plateau above a leaf area index (LAI) of 3.0.
- The ratio of actual/potential photosynthesis decrease in proportion to the reductions in the most limiting environmental factor.
- The fraction of production not allocated to roots can be partitioned among foliage, stem and branches based on allometric relationships and knowledge of annual leaf turnover.

In the model, absorbed photosynthetically active radiation (APAR) is estimated from global solar radiation and LAI; the utilized portion, APARu, is calculated by reducing APAR by an amount

determined by a series of modifiers that take values between 0 (system 'shutdown') and 1 (no constraint) to limit gas exchange via canopy stomatal conductance (Landsberg and Waring, 1997). The modifiers include: (a) high averaged day-time D; (b) the frequency of subfreezing conditions, (c) soil drought and (d) temperature. Limitations on APARu are imposed each month by the modifier with the lowest value. Drought limitations are imposed as a function of soil texture when the total monthly precipitation and soil water supply are significantly less than transpiration estimated with the Penman–Monteith equation (Coops et al., 2005). Gross primary production (PG) is calculated by multiplying APARu by a canopy quantum efficiency coefficient, with a maximum value set by the soil fertility ranking and reduced monthly when mean temperatures are suboptimal for photosynthesis and growth. A major simplification in the 3-PG model is that it does not require detailed calculation of respiration from knowledge of root turnover, but rather assumes that autotrophic respiration (Ra) and total net primary production (PN) in temperate forests are approximately fixed fractions (0.53 and 0.47, SE ± 0.04) of PG (Landsberg and Waring, 1997; Waring et al., 1998; Law et al., 2001). The model partitions PN into root and aboveground biomass. Under more favorable climatic conditions, the fraction of photosynthate allocated to roots increases with infertility of the soil (Landsberg and Waring, 1997).

We further simplified the approach by selecting Douglas-fir (*Pseudotsuga menziesii*), the most widely distributed species in the region, to characterize the importance of climatic constraints on photosynthesis and growth across all forested environments, as we have done previously for other purposes (Swenson et al., 2005; Waring et al., 2005; Coops et al., 2007). Rather than utilizing climatic data directly, we use 3-PG to assess the implications of seasonal limitations of water availability, deviations from an optimum temperature of 20 °C, frost frequency, and atmospheric humidity deficits on photosynthesis and growth. The link to photosynthesis is critical because the potential varies seasonally. The upper limits are set by the amount of light absorbed by the canopy's foliage. Although we recognize that soil fertility and soil water storage capacity vary considerably across the region (Swenson et al., 2005), in this paper we chose to keep soil properties constant to simplify the analysis of the effects of climatic variation on tree distributions. We did this by setting the maximum available soil water storage capacity at 200 mm and giving a moderately high rank to a soil fertility index (0.7), which generates a maximum photosynthetic quantum efficiency of 0.05 mol C mol photon^{-1} (2.75 g C MJ^{-1} of absorbed photosynthetically active radiation).

We used parameters for equations describing the physiological responses of Douglas-fir reported in a previous publication (Coops et al., 2007). The extent that different species encounter environments that would impose restrictions on the performance of Douglas-fir is incorporated through an automated regression-tree analysis, described in more detail below. This statistical procedure generates a suite of rules for each species that differentiates the relative importance of the four climatic modifiers (maximum impact imposed through the year by: water availability, deviations from an optimum temperature of 20 °C, frost frequency, and atmospheric humidity deficits).

2.2. Climatic data

Monthly mean climatic data, registered at a resolution of 1 km^2, were obtained for precipitation, minimum and maximum temperature, frost occurrence, and short wave radiation over the 18-year period from 1980 to 1997 from the DAYMET US climatological database (Thornton et al., 1997; Thornton and Running, 1999).[1]

---

[1] URL: (http://www.daymet.org).