Forest Ecology and Management 326 (2014) 142–156

Contents lists available at ScienceDirect

Forest Ecology and Management

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

Predicting site index of plantation loblolly pine from biophysical variables

Charles O. Sabatia *, Harold E. Burkhart

Department of Forest Resources and Environmental Conservation, Virginia Polytechnic Institute and State University, 319 Cheatham Hall, 310 West Campus Drive, Blacksburg, VA 24061-0324, United States

ARTICLE INFO

Article history:
Received 20 January 2014
Received in revised form 17 April 2014
Accepted 18 April 2014
Available online 14 May 2014

Keywords:
Pinus taeda
Random forest
Factor analysis
Climate change

ABSTRACT

Concerns of the effect of climate change on forest productivity have impelled the need to accurately predict forest productivity from climate, physiographic and edaphic variables (biophysical variables). We fitted and evaluated random forest models and nonlinear least squares regression models for predicting plantation loblolly pine (*Pinus taeda* L.) site index from biophysical variables. Tree and stand location data were provided by the Virginia Tech Forest Modeling Research Cooperative. Climate data for each stand location were computed using the Oakridge National Laboratories' daily surface weather prediction models, while soils data were extracted from the USDA Natural Resource Conservation Service SSURGO GIS database using GIS data extraction techniques. Separate models were fitted for non-intensively managed (Non-IMP) and intensively managed (IMP) loblolly pine plantations. Variable selection methods in both modeling approaches showed that the number of biophysical variables that were important in predicting site index of IMP loblolly pine was smaller than the number for Non-IMP stands. The non-parametric random forest models had better fit and prediction statistics than the least squares parametric models but exhibited the potential to give illogical predictions under extrapolation. Site index predictions from both modeling approaches exhibited a regression towards the mean.

1. Introduction

Climate, edaphic, and physiographic factors (biophysical factors) are important drivers of forest productivity. Adverse changes in some of these factors may cause a decline in forest productivity with a possibility of significant impacts on ecological and economic roles of forests. Currently, there are concerns of climate change with projections of warmer temperatures, increased carbon dioxide concentrations, and longer growing seasons (Menzel and Fabian, 1999; Hansen et al., 2012; Peters et al., 2013). The need to accurately predict the effects of climate change on forest productivity has led to the development of different kinds of empirical models that predict forest productivity directly from biophysical predictors. Parametric and non-parametric modeling approaches have been applied. Monserud et al. (2006) developed a linear regression model that predicted lodgepole pine (*Pinus contorta* Dougl. ex Loud.) site index as a function of growing degree days. This model was later used to predict the potential change in lodgepole pine site index under climate change in Alberta, Canada (Monserud et al., 2008). Other linear regression models that predict site index from biophysical factors have also been developed (e.g. Fries et al., 2000; Hamel et al., 2004; Sharma et al., 2012). Crookston et al. (2010) showed how the non-parametric random forest model, that predicted site index as a function of climate variables, can be used to incorporate climate change effects in the US Forest Service Forest Vegetation Simulator (FVS) model. In a similar study, Weiskittel et al. (2011) related site index and forest gross primary productivity to climate variables in western United States using the non-parametric random forest model and used the resulting models to predict the effect of climate change on the productivity of the forests in the study area. In our knowledge, similar modeling approaches have not been applied nor evaluated on plantation loblolly pine (*Pinus taeda* L.) in southern United States.

Loblolly pine plantations form a significant proportion of forest land in the southern United States. In the early 2000s there were over 12 million ha of pine plantations in this region composed mainly of loblolly pine, with projections that the area would increase to 22 million ha by 2040. Forests in southern United States produced 58% of United States and 15.8% of world's timber with the world's proportion projected to increase by about 30% by 2040 (Wear and Greis, 2002). Loblolly pine makes up about 80%

* Corresponding author. Current address: New Zealand Forest Research Institute Limited (Scion), Te Papa Tipu Innovation Park, 49 Sala Street, Private Bag 3020, Rotorua 3046, New Zealand. Tel.: +64 7 343 5899; fax: +64 7 348 0952.

E-mail address: charles.sabatia@scionresearch.com (C.O. Sabatia).

http://dx.doi.org/10.1016/j.foreco.2014.04.019
0378-1127/

of the approximately 1 million ha of forest plantations established annually in southern United States (McKeand et al., 2003). Thus, it is important to be able to accurately predict the potential effects of climate change on this critical forest resource to inform development of appropriate strategies for mitigating the effects of climate change.

The objectives of the current study were to: (1) develop parametric and non-parametric empirical models that relate site index of plantation loblolly pine to biophysical variables, and (2) provide an evaluation of the predictive and application performance of the two modeling approaches. The parametric approach involved the use of factor analysis and a nonlinear model fitted by least squares regression while the non-parametric approach involved the use of the regression trees approach random forests (RF) (Breiman, 2001). A detailed application level description of the random forests method is given in Prasad et al. (2006) and also in Weiskittel et al. (2011).

2. Methods

2.1. Study area and site index data

The study area encompassed the loblolly pine natural range across the 13 states that make up southern United States (Fig. 1). In this area, the Forest Modeling Research Cooperative at Virginia Tech established permanent research plots in planted loblolly pine stands between 1980 and 1982 (186 plots in non-intensively managed genetically unimproved stands) and also between 1996 and 2000 (170 plots in intensively managed genetically improved stands) to study the effect of thinning on loblolly pine stand development. The non-intensively managed (Non-IMP) stands ranged in age between 8 and 25 years old at the time of plot establishment and were measured at 3-year intervals over a 21-year measurement period. A more detailed description of the study protocol can be found in Burkhart et al. (1985) and also in Amateis et al. (2006). The intensively managed (IMP) stands were 3–8 years old at the time of plot establishment and were measured at 2-year intervals, and later at 3-year intervals. This study was still ongoing as of the time of the current analysis and had been measured over a period of between 11 and 14 years. A more detailed description of this study can be found in Amateis et al. (2006). Each study location had three 0.04–0.08-ha plots, an unthinned control plot and two thinning treatment plots. For the current study, only tree total height and age data from stands in the control plots were considered. For each plot and measurement age, dominant height was computed as the average height of all the trees in the plot that had been identified, by visual inspection, as dominants or codominants.

The stands at the different study locations exhibited different dominant height growth trajectories (Fig. 2). Locations that exhibited unusual height growth patterns, e.g. those that exhibited a non-increasing height growth trajectory over several re-measurements, and those that had only one or fewer re-measurements were excluded from the analysis. Consequently, data used in the current analysis were from 172 locations in the Non-IMP stands and 136 locations in the IMP stands. Site index estimates were computed for each of the locations by fitting, separately to the IMP and Non-IMP height and age data, the generalized algebraic difference approach (GADA) dynamic self-referencing model

$$H_{dt} = \frac{\alpha + X_0}{1 + \frac{\beta}{X_0} \times t^{\gamma}} + \varepsilon \qquad (1)$$

where $X_0 = 0.5\left(\sum_{i=1}^{n}\rho_i D_i - \alpha + \sqrt{\left(\sum_{i=1}^{n}\rho_i D_i - \alpha\right)^2 + 4 \times \beta \sum_{i=1}^{n}\rho_i D_i t_I^{\gamma}}\right)$; H_{dt} is the dominant height of plot i at age t; t_I is the index age, which

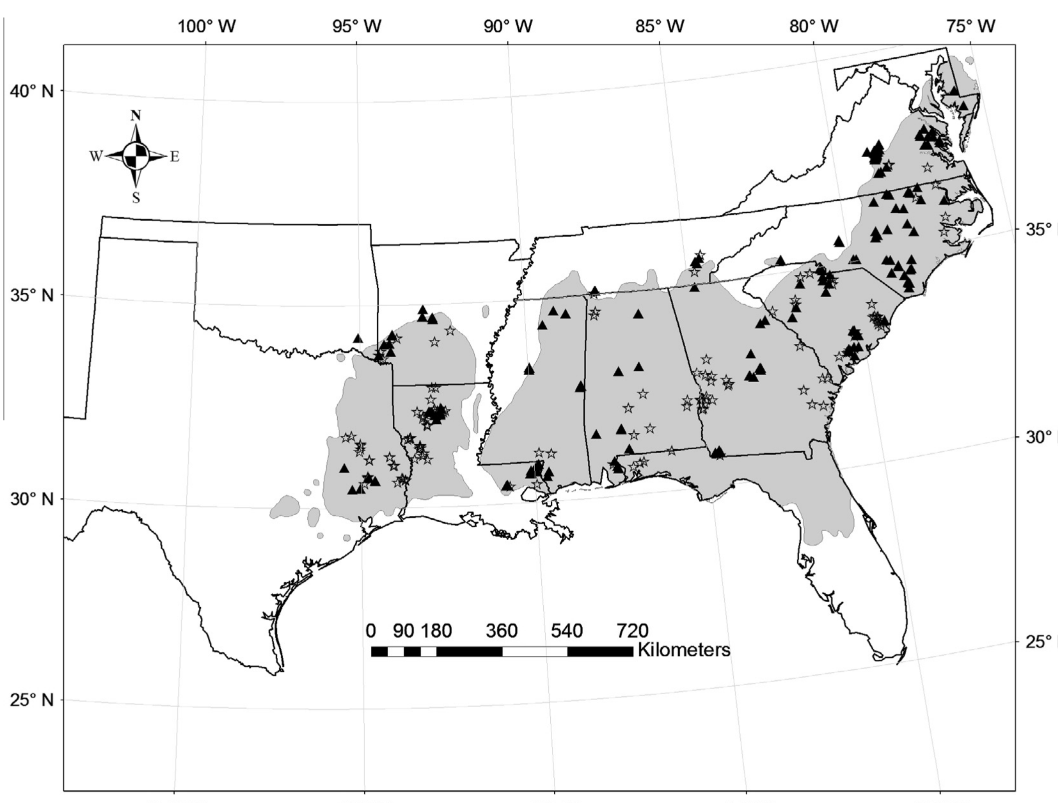

Fig. 1. Location of the 186 Non-IMP (triangles) and 170 IMP (stars) Virginia Tech Forest Modeling Research Cooperative loblolly pine permanent research plots. The portion of the map shaded in gray is the natural range of loblolly pine in southern United States (Little, 1971).