Available online at www.sciencedirect.com
SCIENCE DIRECT®
Forest Ecology and Management 214 (2005) 201–211
Forest Ecology and Management
www.elsevier.com/locate/foreco

Evaluation of spatial predictions of site index obtained by parametric and nonparametric methods—A case study of lodgepole pine productivity

Yonghe Wang [a,*], Frédéric Raulier [b], Chhun-Huor Ung [c]

[a] Canadian Forest Service, Northern Forestry Centre, 5320-122 Street, Edmonton, Alta., Canada T6H 3S5
[b] Faculty of Forestry and Geomatics, Laval University, Sainte-Foy, Que., Canada G1K 7P4
[c] Canadian Forest Service, Laurentian Forestry Centre, Sainte-Foy, Que., Canada G1V 4C7

Received 1 October 2004; received in revised form 12 April 2005; accepted 12 April 2005

Abstract

We demonstrate the potential of using least-squares regression, generalized additive model, tree-based model, and neural network model on layers of environmental data grids for mapping site index in a case study. Grids of numerical environmental variables represented layered data, and a sparse site index plot network was located in the grids. Site index data were based on stem analysis (observed height at the index age of 50 years) of 431 lodgepole pine trees in 88 sample plots. The plots were established in a 17,460 km^2 boreal mixedwood forest of Alberta, Canada dominated by mature and over-mature stands. The generalized additive model presented a better fit and better adaptability to extreme data (i.e., mature stands) than the least squares nonlinear and other nonparametric techniques, such as the tree-based model and neural network model. Among the four models tested, nonlinear regression is of the data modeling culture, which assumes a stochastic data to relate productivity to environmental variables, and such models are optimized for estimation. Other three models belong to the algorithm modeling culture, which treat the relationship between productivity and independent variables as an unknown black box and try to find a function between them; therefore, these models are more suitable for prediction purpose. Implications for biophysical site index modelling with extreme data are discussed.

Keywords: Generalized additive model; Tree-based model; Neural network model; Lodgepole pine; Nonlinear regression; Mature stands; Over-mature stand

1. Introduction

Although there has been a paradigm shift from simple, even-aged silviculture and growth modelling to gap-based silviculture and uneven-aged growth modelling, a quantified surrogate of potential wood production is still required for forest management decision-making. Because of its operational importance, site index (SI), defined as the average height of a certain number of the largest trees per hectare at a particular reference age, is a broadly accepted surrogate of

* Corresponding author. Tel.: +780 435 7237; fax: +780 435 7359.
E-mail address: ywang@nrcan.gc.ca (Y. Wang).

0378-1127/$ – see front matter.
doi:10.1016/j.foreco.2005.04.025

potential productivity. There are many cases where site index cannot be adequately estimated because it is a species-specific and phytometric index of site productivity. For example, it cannot be estimated in treeless areas or in stands where the concerned species is not present. Its estimation in young stands is difficult because a slight error in estimation can result in a large error on predicted yields.

Various studies have thus attempted to relate SI to biophysical factors (Hunter and Gibson, 1984; Kabzems and Klinka, 1987; Ung et al., 2001; McKenney and Pedlar, 2003), but observed correlations are generally low, and the relationships differ among studies because of no definitive connection between SI and the biophysical variables. Almost all of these studies have used a parametric approach that implied solving two problems: the relationship definition between SI and the biophysical variables and collinearity existing between the explanatory variables in the regression model. Given that a relationship needs to be determined before parameter estimation and that no definite relationship exists, the use of the least square regression technique implies an unknown source of error that could be bypassed with a nonparametric approach. Also, collinearity often exists among biophysical variables when some of these can be expressed as linear combinations of other predictor variables. For example, climate moisture index, which is used in this study, is derived directly from temperature and precipitation, which in turn are related to latitude and elevation. Collinearity affects the statistical estimation of the parameters as it inflates the variance of at least one of the estimated regression coefficients, and consequently also inflates the estimation of the confidence interval around the predicted values (Belsley et al., 1980).

To avoid these problems, alternative methods must be identified and assessed. For instance, McKenney and Pedlar (2003) have successfully applied a tree-based regression method (TREE) for relating soil, topographic, and climatic attributes to site productivity. TREE has the potential of producing discontinuous SI values; other nonparametric methods such as the generalized additive model (GAM) and the neural network model (NNT) could generate satisfactory results because their outputs are continuous. TREE, GAM, and NNT are considered nonparametric because no functional structure between predictor and response variables is pre-specified. GAM and NNT have been used as exploratory tools in the analysis of species distribution with respect to climatic factors in a landscape study (Yee and Mitchell, 1991) and as tree growth and mortality models (Guan and Gertner, 1991a,b; Sironen et al., 2003), but, as far as we know, they have not yet been used to predict SI in spatial terms.

The objective of this study was to assess the usefulness of the TREE, GAM, and NNT nonparametric techniques, and to compare them with the least-squares nonlinear regression model (NLIN) in developing a spatial SI model for mature stands of lodgepole pine (*Pinus contorta* var. *latifolia*) in Alberta, Canada.

2. Materials and methods

2.1. Study site and data sources

The Wapiti region of Alberta, Canada, encompasses an area of approximately 17,460 km^2 between 118°W and 120°W and between 54°N and 55°N. It is a mosaic of four natural ecological subregions: Boreal Mixedwood, Lower Foothills, Upper Foothills, and Sub Alpine (Beckingham et al., 1996). Lodgepole pine is the dominant timber species across the region (Corns, 1978). The nine predictor variables used in this study consisted of grid data from various sources. They include three geographical factors: the easting (m) grid (x_1) and the northing (m) (x_2) of the Universal Transverse Mercator (UTM) ordination, along with elevation (m) (x_3). In addition, six biophysical factors are considered to influence variability in site conditions within the study region: climate moisture index (cm) (x_4), growing degree days (°C) (x_5), annual precipitation (mm) (x_6), soil sand fraction (%) (x_7), January monthly mean temperature (°C) (x_8), and July monthly mean temperature (°C) (x_9).

Grid x_3 was created from the National Topographic Database 1:50,000 contour lines projected to UTM Zone 11, NAD 83. Grid x_4 was obtained as the difference between x_6 and the annual sum of monthly potential evapotranspiration (PE), which was computed from relevant data grids in the NatGRID database (Hogg, 1994; McKenney et al., 1996). Monthly PE was determined by means of the