Forest Ecology and Management 329 (2014) 306–317

Contents lists available at ScienceDirect

Forest Ecology and Management

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

A compatible system of biomass equations for three conifer species in Northeast, China

Lihu Dong [a], Lianjun Zhang [b], Fengri Li [a,*]

[a] Department of Forest Management, School of Forestry, Northeast Forestry University, Harbin, Heilongjiang 150040, PR China
[b] Department of Forest and Natural Resources Management, State University of New York, College of Environmental Science and Forestry (SUNY-ESF), One Forestry Drive, Syracuse, NY 13210, USA

ARTICLE INFO

Article history:
Available online 11 July 2014

Keywords:
Model error structure
Likelihood analysis
Non-linear seemingly unrelated regressions (NSUR)
Anti-logarithmic transformation

ABSTRACT

A compatible system of biomass equations was developed for three major conifer species, Korean spruce (*Picea koraiensis* Nakai), Korean pine (*Pinus koraiensis* Sieb. et Zucc), and Dahurian larch (*Larix gmelinii* Rupr) in northeastern China. The model error structure (additive vs multiplicative) of the power function ($Y = a\ X^b$) was evaluated using a likelihood analysis. The results indicated that the assumption of multiplicative error structure was strongly supported for the biomass equations of total, sub-total, and tree components. Thus, a system of log-transformed biomass equations was developed using nonlinear seemly unrelated regression (NSUR), with three constraints on the structural parameters to account for the cross-equation error correlations between four tree component biomass (roots, stems, branches, and foliage), two sub-total biomass (aboveground and crown), and total tree biomass. The effectiveness of three anti-log correction factors for predicting the expected biomass in original scale was also assessed. Our results indicated that (1) the likelihood analysis can be used as a tool for rigorously evaluating the error structures of tree biomass equations and choosing an appropriate model form for the given biomass data; (2) the additive or compatible system of biomass equations with three constraints can be developed by NSUR to obtain favorable model fitting and prediction performance; and (3) the anti-log correction may not be necessary and sometimes can be ignored.

1. Introduction

Accurate quantification of tree biomass is critical and essential for calculating biomass energy, carbon storage, and sequestration of forests, as well as for studying climate change, forest health, forest productivity, and nutrient cycling (Clark and Murphy, 2011). Up to date, hundreds of biomass equations have been developed for more than 100 tree species around the world (e.g., Chojnacky, 2002; Jenkins et al., 2003; Zianis and Seura, 2005; Woodall et al., 2011). However, there are some noteworthy issues in constructing and applying biomass equations, such as (1) the possibility of some reported biomass models not holding the additivity or compatibility property among tree component equations; (2) the existence of a contentious issue on which model error structure is appropriate for biomass data, i.e., additive error structure ($Y = a\ X^b + \varepsilon$) versus multiplicative error structure ($Y = a\ X^b\ \varepsilon$); (3) deciding what is the best correction factor for reducing or removing the bias due to anti-log transformation if a log-transformed model is applied; and deciding (4) what is the optimal method for validating tree biomass models when the sample size of biomass data is relatively small. Researchers and model developers have been continuously working and debating on these issues over the last several decades.

To develop biomass equations, total tree biomass and tree component biomass (e.g., stems, roots, crown, branches and foliage) are regressed to tree diameter, or both diameter and height using linear or nonlinear regression (e.g., Burrows et al., 2000; Ketterings et al., 2001; Wang, 2006; Wang et al., 2006; Cai et al., 2013; Chan et al., 2013). However, if there is more than one tree component involved, fitting a biomass equation separately to each component ignores the inherent correlations among the biomass of tree components that are measured on the same sample trees. Consequently, the sum of biomass predictions from the separate models of tree components may not equal the biomass prediction from the total tree model (Parresol, 1999). To deal with this issue of in additivity or incompatibility, different model specifications and estimation methods have been suggested in order to ensure the additivity in a system of biomass equations for both linear and non-linear models (Kozak, 1970; Parresol, 1999, 2001). For linear biomass models, the parameter estimation methods include:

* Corresponding author. Tel.: +86 451 82190609(O); fax: +86 451 82190347.
E-mail address: fengrili@126.com (F. Li).

http://dx.doi.org/10.1016/j.foreco.2014.05.050
0378-1127/

simple least squares (SLS) (Kozak, 1970; Cunia and Briggs, 1984), restricted least squares (RLS) (Chiyenda and Kozak, 1984), and seemingly unrelated regressions (SUR) (Parresol, 1999). For nonlinear biomass models, the parameter estimation models include: minimizing loss function (MLF) (Reed and Green, 1985), generalized method of moments (GMM) (Greene, 1999), maximum likelihood (ML) (Parresol, 2001), two-stage error-in-variable model (TSEM) (Tang et al., 2001; Tang and Wang, 2002), and non-linear seemingly unrelated regressions (NSUR) (Parresol, 2001). SUR and NSUR have become more popular in recent years because they are more general and flexible, and permit that each component model may have its own independent variable(s) and each model can use its own weighting function for heteroscedasticity, resulting in a lower variance for the total tree biomass model (Parresol, 2001).

The allometric equation of the form $Y = a\,X^b$ (power function) is a mathematical function commonly used for tree biomass modeling, where Y is tree biomass (total or component biomass), X is a tree dimension variable, such as diameter, and a and b are model parameters. It is well known that this allometric model assumes either an additive error structure ($Y = a\,X^b + \varepsilon$) or a multiplicative error structure ($Y = a\,X^b\,\varepsilon$). If the former is assumed, nonlinear regression should be used to directly fit the power function to tree biomass data. If the latter is assumed, logarithmic transformation is usually applied to convert the nonlinear power function to a linear model so that $\log(Y) = \log(a) + b\ \log(X) + \log(\varepsilon) = a^* + b^*\log(X) + \varepsilon^*$, where log is natural logarithm. The continuing debate is focused on the question of which error structure is more appropriate to be assumed for the allometric relationship. Some researchers suggest that data analysis and modeling based on log-transformation are flawed because (1) converting the original data to logarithms is a nonlinear transformation, which fundamentally alters the relationship between Y and X in the power function; (2) consequently, the nonlinear log-transformation produces disproportional weighting on small or large data observations, which leads to misleading estimation for the two parameters of the power function; (3) fitting the log-model to the logarithms is based on the implicit assumption that the power function is adequate to the original data, which may or may not be correct; and (4) the fitted log-model provides the prediction of $\log(Y)$, not Y, and the anti-log transformation from $\log(Y)$ to Y introduces bias so that correction on the bias is necessary. Therefore, data analysis and modeling should be carried out on the original data of measurement using the nonlinear power function (Fattorini, 2007; Packard and Birchard, 2008; Packard, 2009; Bi et al., 2004). Other researchers argue that most biological phenomena are multiplicative in nature, i.e., the biological variation of allometry is proportional rather than absolute across the orders of magnitude. In addition, the log-transformation can also normalize sample variance and reduce the influence of outliers. Thus, the multiplicative error structure is more appropriate and better for the allometric relationship between variables (Gingerich, 2000; Kerkhoff and Enquist, 2009). To facilitate the objective determination on the error structures, Xiao et al. (2011), Ballantyne (2013) outline the approach of likelihood analysis to evaluating and comparing model error structures (additive vs multiplicative) for an allometric relationship between variables with available data, which is considered consistent with core statistical principles. However, the likelihood analysis has not been widely applied in forestry and ecology to verify the error structure of tree biomass data (Lai et al., 2013).

If a log-model is used to fit the log-transformed biomass data, the fitted log-model yields the prediction of $\log(Y)$. To obtain the desired prediction of Y, anti-log transformation is needed to convert $\log(Y)$ to Y. It is well known that this back-transformation process introduces bias into the estimation of Y. Consequently, a correction factor is typically applied to remove or reduce the bias. Several correction factors have been proposed, applied, and compared in the literature over the last several decades (e.g., Finney, 1941; Baskerville, 1972; Flewelling and Pienaar, 1981; Clifford et al., 2013). However, Madgwick and Satoo (1975) finds that anti-log transformation tends to overestimate Y by applying the corrections factor, and suggests that the correction factor may be ignored if the bias from anti-log is relatively small compared to the overall variation in the estimate of biomass.

One of the main objectives of regression analysis is to select an accurate and reliable model for predicting the dependent variable (e.g., tree biomass). The last and perhaps most important step in regression analysis is to carry out a thorough validation of the selected model. Snee (1977) outlines four procedures for validating regression models: (1) comparing model coefficient estimates and prediction with theory; (2) comparing model results with those obtained by theory and simulation; (3) using new data (independent of model development data); and (4) using data splitting or cross validation. Since the use of new data is costly and not feasible in practice, data splitting or cross validation became popular in the literature (Berk, 1984; Shao, 1993; Ronchetti et al., 1997; Zhang, 1997). However, Kozak and Kozak (2003) shows that data splitting provides little, if any, additional information on assessing and evaluating model prediction or performance because the model construction and validation data sets are the subsets of the entire data set. The results of cross validation by data splitting are not as reliable as using the entire data set. Therefore, model validation by the data splitting methods may be unnecessary, while the lack of fit statistics (average bias and its standard error) across the subgroups or classes of tree diameter are superior for the purpose of model validation (Kozak and Kozak, 2003). Others suggest that the most applicable models should be validated using a jackknifing technique, also known as the "leave-one-out" method or Predicted Sum of Squares (PRESS) (Quint and Dech, 2010; Li and Zhao, 2013). Perhaps the prediction bias by jackknifing across the diameter classes would be a adequate way of validating the tree biomass models.

Korean spruce (*Picea koraiensis* Nakai), Korean pine (*Pinus koraiensis* Sieb. et Zucc), and Dahurian larch (*Larix gmelinii* Rupr) are important conifer species in the natural forests of northeastern China. Up to date, there are limited studies in China on the aboveground biomass of these species, while no biomass model is available for crown and root biomass. The objectives of this study were: (1) to examine which model error structure (additive vs multiplicative) is more suitable for the allometric relationships between tree biomass (e.g., total and tree components) and diameter for the three coniferous species; (2) to construct a system of additive biomass equations for the three species; (3) to validate the performance of the biomass models by jackknifing technique across the classes of tree diameter; (4) to analyze and compare three correction factors for the anti-log transformation on estimating tree biomass; and (5) to investigate the sources of prediction errors for total and component models.

2. Data and methods

2.1. Study area and data

The study area is the forest regions of northeastern China. It encompasses Daxing'an Mountains (from 121°12′E to 127°00′E and from 50°10′N to 53°33′N), Xiaoxing'an Mountains (from 127°42′E to 130°14′E and from 46°28′N to 49°21′N), and Changbai Mountains (from 126°40′E to 131°16′E and from 41°35′N to 47°57′N), located in Heilongjiang Province and Jilin Province (Fig. 1). The elevation usually ranges 300–700 m above sea level in Daxing'an Mountains, 600–1000 m in Xiaoxing'an Mountains,