Forest Ecology and Management 319 (2014) 62–66

Contents lists available at ScienceDirect

Forest Ecology and Management

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

Prediction of tree diameter growth using quantile regression and mixed-effects models

Som B. Bohora [a], Quang V. Cao [b,*]

[a] Department of Biostatistics and Epidemiology, College of Public Health, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA
[b] School of Renewable Natural Resources, Louisiana State University Agricultural Center, Baton Rouge, LA 70803, USA

ARTICLE INFO

Article history:
Received 19 December 2013
Received in revised form 5 February 2014
Accepted 5 February 2014

Keywords:
Calibration
Localization
Projection
Pinus taeda

ABSTRACT

A tree diameter growth function is an important component of an individual-tree model. This function can be considered as a mixed-effects model, in which a diameter measurement can be used to calibrate (or localize) the equation to produce improved diameter predictions for the same tree in the future. Another approach considered in this study involved a system of quantile regressions, in which future diameters can be determined through interpolation, based on a current diameter measurement. The aim of this study was to evaluate the use of quantile regression and mixed-effects models in predicting tree diameter growth. Tree diameter at the end of each growth period was predicted from diameter at the beginning of the period by use of one of the four methods: the mixed-effects model and three quantile regression methods that were based on nine quantiles, five quantiles, and three quantiles. The mixed-effects model performed as well as the three quantile regression methods, based on the mean absolute difference and fit index, but was far superior in terms of the mean difference. The mixed-effects model produced an unbiased prediction of future diameter, up to ten years into the future, when calibrated with a current diameter measurement.

1. Introduction

Forest management decisions are often based on current stand conditions, obtained from inventory data, and future stand conditions, predicted from growth and yield models. A tree diameter growth function is an important component of an individual-tree model, which simulates survival and growth of individual trees in a forest stand (Avery and Burkhart, 2002). It can also be incorporated into a stand-table projection system (Cao, 2007), which predicts future movements of trees among diameter classes.

Recently, there has been a growing interest in applying mixed-effects models to solve forestry regression problems, such as predicting timber volume (Hall and Clutter, 2004), dominant height growth (Fang and Bailey, 2001), tree mortality (Groom et al., 2012), and tree taper (Cao and Wang, 2011), to name a few. A mixed-effects diameter growth model can be developed which contains both fixed-effects parameters that are common to all trees in the sample, and random effects that are specific to each tree. When these random effects are calculated by use of measurement(s) from a particular tree, the model is said to be calibrated (or localized) for that tree. Results are improved predictions for future diameters of that tree.

Quantile regression (Koenker and Bassett, 1978) has recently been employed by scientists from various backgrounds to address different kinds of problems. These include research in medicine (Austin and Schull, 2003), economics (Machado and Mata, 2005), education and policy (Haile and Nguyen, 2008), and natural resource management (Cade et al., 2005). In forestry, quantile regression has been applied to model self-thinning boundary lines (Zhang et al., 2005) and tree diameter percentiles (Mehtätalo et al., 2008), compute stand density index (Ducey and Knapp, 2010), or evaluate the spread rate of forest diseases (Evans and Finkral, 2010).

Individual tree growth from a locality might not follow the path modeled by a published tree-level equation, which gives predictions based on average tree growth. It is well known that a tree model can be localized by use of the mixed model approach. We introduce in this paper a new method for local tree-level predictions that involves quantile regression models. These models constitute a set of quantile diameter curves, similar to a set of site index curves, that can be used to predict tree diameter at a future age from past diameter measurement, assuming that the relative

* Corresponding author. Tel.: +1 225 578 4218; fax: +1 225 578 4227.
E-mail addresses: Som-Bohora@ouhsc.edu (S.B. Bohora), QCAO@lsu.edu (Q.V. Cao).

http://dx.doi.org/10.1016/j.foreco.2014.02.006
0378-1127/

location of the tree does not change over time. The objective of this paper is to evaluate the use of quantile regression and mixed-effects models in predicting tree diameter growth.

2. Material and methods

2.1. Data

The data set consisted of measurements collected from 340 trees in an unthinned loblolly pine (*Pinus taeda* L.) plantation in Lee Memorial Forest near Bogalusa, Louisiana, USA. The trees were planted in a 2.7 m by 3.7 m spacing. Diameters at breast height were recorded annually from age 2 to age 21 (Fig. 1), totaling 6147 observations.

The data were randomly divided into five groups of 68 trees each (Table 1). The leave-one-out evaluation scheme was applied in this study. The scheme consisted of five stages, each involved validation data from a group and fit data from the remaining four groups. Parameters of the diameter growth model were estimated from the fit data, and then used to predict for the validation data. The predictions from all stages were used to compute evaluation statistics for the different methods.

2.2. Diameter growth model

Several growth functions were investigated in a preliminary analysis. The function that fit the data best had a similar form to that of the height–age function developed by Bailey and Clutter (1974):

$$y(t_{ij}) = \exp(b_1 + b_2 t_{ij}^{b_3}) + \varepsilon_{ij}, \tag{1}$$

where $y(t_{ij})$ is diameter at breast height (cm) of measurement j for tree i at age t_{ij}, and b_k's are regression parameters.

2.3. Mixed-effects model

In the mixed-effects framework, all parameters of Eq. (1) can be expressed as fixed-effects parameters (common to all trees), with certain parameters containing additional random components, which are specific to individual trees. Note that because the data were limited due to lack of multiple plots, a hierarchical model with grouping of plots and trees within plots was not possible. All combinations of the regression parameters (b_1, b_2, and b_3) are candidates for random parameters. Eq. (1) can be written in matrix form as follows:

$$\mathbf{y_i} = \mathbf{f}(\mathbf{b}, \mathbf{u_i}, \mathbf{t_i}) + \boldsymbol{\varepsilon}_\mathbf{i}, \tag{2}$$

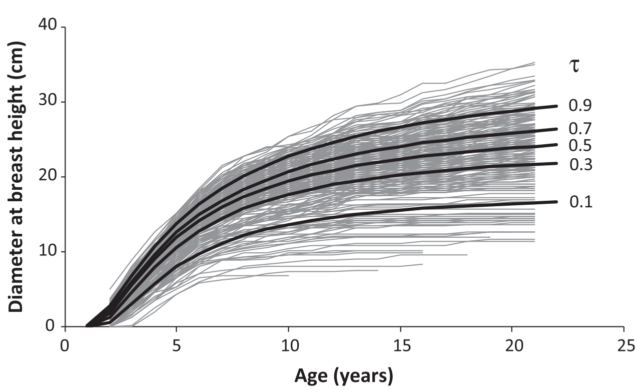

Fig. 1. Graphs of observed tree diameter growth (gray) and curves generated by quantile regressions (black) based on five quantiles.

Table 1
Summary of tree diameters measured at age 21, by group.

Group	N[a]	Mean	SD[b]	Minimum	Maximum
1	50	23.6	4.6	13.0	34.3
2	60	24.4	4.6	14.2	33.0
3	58	23.9	5.3	11.4	35.1
4	53	23.9	4.3	14.2	35.3
5	56	24.4	4.6	12.7	33.5

[a] N = number of surviving trees at age 21.
[b] SD = standard deviation.

where $\mathbf{y_i} = [y(t_{i1}), y(t_{i2}), \ldots, y(t_{i,n_i})]^T$, $\mathbf{t_i} = [t_{i1}, t_{i2}, \ldots, t_{i,n_i}]^T$, $\boldsymbol{\varepsilon}_\mathbf{i} = [\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{i,n_i}]^T$, n_i is number of measurements for tree i, and $\mathbf{b}$ and $\mathbf{u_i}$ are column vectors of fixed- and random-effects parameters, respectively. The assumptions are:

$$\boldsymbol{\varepsilon}_\mathbf{i} \sim N(0, \mathbf{R}), \text{ and}$$

$$\mathbf{u_i} \sim N(0, \mathbf{D}),$$

where $\mathbf{R}$ and $\mathbf{D}$ are diagonal matrices, assuming that the $\boldsymbol{\varepsilon}_\mathbf{i}$ and $\mathbf{u_i}$ are independent. Procedure NLMIXED from SAS (SAS Institute Inc., 2008) was used to obtain fixed- and random-effects parameters of Eq. (2).

The random parameters $\mathbf{u_i}$ for tree i can be computed by use of the first-order Taylor series expansion (Meng and Huang, 2009):

$$\hat{\mathbf{u}}_\mathbf{i}^{\mathbf{k}+1} = \widehat{\mathbf{D}}\mathbf{Z}_\mathbf{i}^\mathbf{T}(\mathbf{Z_i}\widehat{\mathbf{D}}\mathbf{Z}_\mathbf{i}^\mathbf{T} + \widehat{\mathbf{R}})^{-1}[\mathbf{y_i} - \mathbf{f}(\hat{\mathbf{b}}, \hat{\mathbf{u}}_\mathbf{i}^\mathbf{k}, \mathbf{t_i}) + \mathbf{Z_i}\hat{\mathbf{u}}_\mathbf{i}^\mathbf{k}], \tag{3}$$

where $\hat{\mathbf{u}}_\mathbf{i}^\mathbf{k}$ is estimate of the random parameters for tree i at the kth iteration, $\widehat{\mathbf{D}}$ is estimate of $\mathbf{D}$, the variance–covariance matrix for $\mathbf{u_i}$, $\mathbf{Z_i} = \frac{\partial \mathbf{f}(\mathbf{b}, \mathbf{u_i}, \mathbf{t_i})}{\partial \mathbf{u_i}}\Big|_{\hat{\mathbf{b}}, \hat{\mathbf{u}}_\mathbf{i}}$, $\widehat{\mathbf{R}}$ is estimate of $\mathbf{R}$, the variance–covariance matrix for $\boldsymbol{\varepsilon}_\mathbf{i}$, $\mathbf{y_i}$ is the $m \times 1$ vector of observed diameters, and m is number of measurements used in localizing the diameter growth model. For each growth period, the diameter at the beginning of the period was known, therefore $m = 1$. An iterative procedure was needed to estimate $\mathbf{u_i}$, whose starting value was set at zero ($\hat{\mathbf{u}}_\mathbf{i}^0 = \mathbf{0}$). The value for $\hat{\mathbf{u}}_\mathbf{i}$ was then repeatedly updated by means of Eq. (3) until the absolute difference between two successive iterations was smaller than a predetermined tolerance limit. The end result would be the Empirical Best Linear Unbiased Predictor (EBLUP) for random effects.

2.4. Quantile regression model

The same form in Eq. (1) was used to predict the τth diameter quantile:

$$\hat{y}_\tau(t_{ij}) = \exp(b_1 + b_2 t_{ij}^{b_3}), \tag{4}$$

where $\hat{y}_\tau(t_{ij})$ is predicted value of the τth quantile of tree diameter at age t_{ij}.

In contrast to the mean regression technique, which employs the least-squares procedure, parameters from the quantile regression are obtained by minimizing

$$S = \sum_{y(t_{ij}) \geqslant \hat{y}_\tau(t_{ij})} \tau[y(t_{ij}) - \hat{y}_\tau(t_{ij})] + \sum_{y(t_{ij}) < \hat{y}_\tau(t_{ij})} (1-\tau)[\hat{y}_\tau(t_{ij}) - y(t_{ij})]. \tag{5}$$

A set of q quantile regressions was developed for the fit data, by use of SAS procedure NLP (SAS Institute Inc., 2010). For each diameter measurement in the validation data, the goal was to identify either the quantile regression curve that passed through it, or the two closest quantile regression curves.

If the diameter measurement of tree i at age t_{ij} was encompassed by the mth and $(m+1)$st quantile regressions, i.e.

$\hat{y}_m(t_{ij}) \leqslant y(t_{ij}) \leqslant \hat{y}_{m+1}(t_{ij})$, a modified diameter growth curve that passed through this point was generated by interpolation:

$$\tilde{y}(t_{ik}) = \alpha\hat{y}_m(t_{ik}) + (1-\alpha)\hat{y}_{m+1}(t_{ik}), \tag{6}$$

where $\tilde{y}(t_{ik})$ is the predicted diameter of tree i at age t_{ik}, and $\alpha = \frac{\hat{y}_{m+1}(t_{ij})-y(t_{ij})}{\hat{y}_{m+1}(t_{ij})-\hat{y}_m(t_{ij})}$ is the interpolation ratio.

If the tree diameter was above the highest (qth) quantile regression curve, Eq. (6) was still appropriate, with $\hat{y}_m$ defined as $\hat{y}_{q-1}$ and $\hat{y}_{m+1}$ as $\hat{y}_q$. Similarly, if the tree diameter was below the lowest (1st) quantile regression curve, $\hat{y}_m$ and $\hat{y}_{m+1}$ in Eq. (6) were defined as $\hat{y}_1$ and $\hat{y}_2$, respectively. The method became extrapolation in nature.

2.5. Evaluation

The mixed-effects model was evaluated against three quantile regression methods that were based on three quantiles (0.1, 0.5, and 0.9), five quantiles (0.1, 0.3, 0.5, 0.7, and 0.9), and nine quantiles (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9). Tree diameter at the end of each growth period was predicted from diameter at the beginning of the period by use of the above four methods. The evaluation statistics included mean difference (MD) between observed and predicted diameters, mean absolute difference (MAD), and a fit index, which is similar to R^2.

3. Results

The mixed-effects combination of b_1 and b_2 produced the lowest values of Akaike's information criterion (AIC) and Bayesian information criterion (BIC) among various combinations of mixed parameters from Eq. (1). The resulting mixed model is:

$$y(t_{ij}) = \exp\left[(b_1+u_1) + (b_2+u_2)t_{ij}^{b_3}\right] + \varepsilon_{ij}, \tag{7}$$

where u_1 and u_2 are random-effects parameters.

Table 2 presents estimates of the parameters of the mixed-effects regression model and quantile regressions at nine quantiles for the complete data. The resulting curves from five quantile regressions are shown in Fig. 1.

Table 3 shows that bias from the mixed-effects model was clearly insignificant compared to the quantile regression approach. Its mean difference fluctuated about zero, and had the lowest absolute values for most projection lengths. On the other hand, all three quantile regression models suffered from overprediction. The bias was more pronounced as projection length increased (Fig. 2a).

For short projections (five years or less), the quantile regression methods were generally better than the mixed model in terms of mean absolute difference and fit index, whereas the reverse was true for projections of more than five years (Table 2). The differences, however, were slight; curves from all methods were almost indistinguishable from one another (Figs. 2b and 2c).

Table 3
Evaluation statistics[a] for the four methods, by projection length.

Projection length	Quantile regression			Mixed model
	3 Quantiles	5 Quantiles	9 Quantiles	
Mean difference (MD)				
1	***−0.019***	−0.022	−0.022	−0.020
2	−0.040	−0.046	−0.045	***−0.013***
3	−0.043	−0.050	−0.048	***0.020***
4	−0.063	−0.070	−0.067	***0.041***
5	−0.141	−0.147	−0.145	***0.008***
6	−0.212	−0.218	−0.216	***−0.018***
7	−0.298	−0.304	−0.300	***−0.055***
8	−0.350	−0.356	−0.353	***-0.058***
9	−0.361	−0.367	−0.364	***−0.018***
10	−0.384	−0.392	−0.388	***0.011***
Mean absolute difference (MAD)				
1	***0.382***	0.383	0.385	0.405
2	***0.578***	0.580	0.584	0.601
3	0.734	***0.732***	0.738	0.748
4	0.870	***0.864***	0.873	0.880
5	0.995	***0.991***	1.001	1.004
6	1.145	***1.144***	1.155	1.154
7	1.303	***1.300***	1.313	***1.299***
8	1.476	1.472	1.485	***1.459***
9	1.640	1.637	1.653	***1.616***
10	1.781	1.778	1.795	***1.752***
Fit index (R^2)[b]				
1	***0.994***	***0.994***	***0.994***	0.993
2	***0.983***	***0.983***	0.982	0.982
3	***0.967***	***0.967***	0.966	0.966
4	***0.948***	***0.948***	0.946	***0.948***
5	***0.929***	***0.929***	0.926	***0.929***
6	0.903	0.903	0.900	***0.904***
7	0.873	0.873	0.870	***0.876***
8	0.837	0.837	0.833	***0.842***
9	0.796	0.796	0.792	***0.805***
10	0.761	0.761	0.756	***0.771***

[a] A bold, italic number denotes the best method for a projection length.
[b] $R^2 = 1 - \sum[y(t_{ij}) - \hat{y}(t_{ij})]^2 / \sum[y(t_{ij}) - \bar{y}]^2$.

4. Discussion

4.1. Number of quantiles

Results indicated that the 9-quantile regression was consistently worse than the 3- and 5-quantile regression models (Table 3). This might be the case of "overfitting," where the extra curves (from the 0.2, 0.4, 0.6, and 0.8 quantiles) were not helpful and actually caused a small decrease in performance against the simpler quantile regression approaches. The curves from the

Table 2
Estimates of parameters of the mixed-effects regression model and quantile regressions at nine quantiles (τ) for the complete data.

Type	b_1	b_2	b_3	σ^2	$\sigma^2_{u_1}$	$\sigma^2_{u_2}$
Mixed-effects model	3.2441	−6.6949	−1.2873	0.3586	0.0626	1.0612
Quantile regression (τ)						
0.1	2.8566	−10.5419	−1.5824			
0.2	3.1093	−8.0830	−1.3857			
0.3	3.1946	−7.6810	−1.3818			
0.4	3.2642	−6.9093	−1.3097			
0.5	3.3361	−6.1054	−1.2102			
0.6	3.3701	−6.0182	−1.2014			
0.7	3.4334	−5.8307	−1.1625			
0.8	3.4899	−5.5846	−1.1266			
0.9	3.5585	−5.4314	−1.1016			

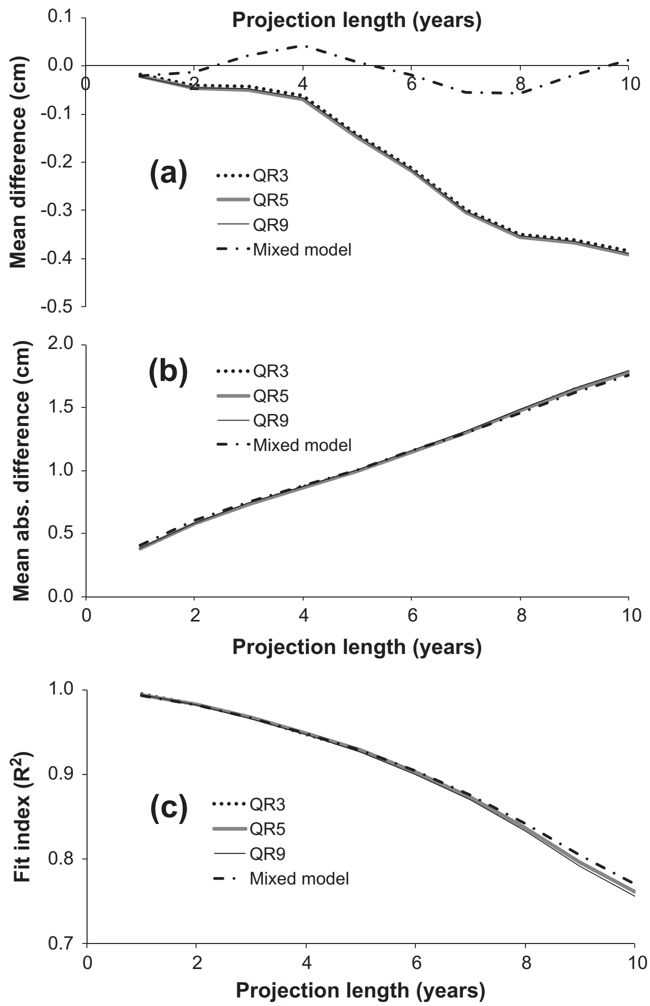

Fig. 2. Graphs of evaluation statistics versus projection length for the quantile regressions with 3, 5, and 9 quantiles (QR3, QR5, and QR9, respectively), and mixed-effects model. The statistics are (a) mean difference, (b) mean absolute difference, and (c) fit index.

quantile regression method based on different number of quantiles (Fig. 2) were very similar and sometimes indistinguishable. It appears from these results that the quantile regression based only on 3 quantiles was adequate in predicting diameter growth.

4.2. Quantile regression versus mixed-effects models

Fig. 2b and c shows that there was virtually little difference among the four methods, based on mean absolute difference and fit index. However, the three quantile regression methods were clearly biased, producing all negative mean differences for various projection lengths (Fig. 2a), or in other words, over-predicted future diameters on average.

On the contrary, the mixed-effects model appeared to be unbiased for these data. The mean difference for the mixed model fluctuated between −0.058 and 0.041 cm, with positive and negative values almost equally represented (Fig. 2a).

The bias from the quantile regression approach stemmed from the implicit assumption that a tree keeps its relative position among the trees over time. The quantile regression does not identify trees, and just assumes that the smallest (or largest) trees at a young age will remain as smallest (or largest) in the future. In the mixed-effects model, on the other hand, the tree number is used as grouping variable, which allows tree-specific predictions to cross one another.

4.3. Effects of projection length

Performance of the four methods should deteriorate as projection length increases. As expected, the mean absolute differences increased and fit indices decreased with increasing projection length. The mean differences for the quantile regression methods also showed the same trend, getting worse as projection length increased (Fig. 2). However, the mean difference of the mixed-effects model did not follow this trend, and failed to show any effect from increasing projection length.

5. Conclusions

The quantile regression curves based on different quantiles have various shapes that can add flexibility to the growth curve. Only three quantile regressions were needed for data from this study, making the system relatively simple to use. The quantile regression method is therefore a straightforward, intuitive method that deserves to be considered as a practical method for projecting tree diameter, height, or volume.

The mixed-effects model performed as well as the three quantile regression methods, based on the mean absolute difference and fit index, but was far superior in terms of the mean difference. When calibrated with a current diameter measurement, this method appeared to produce an unbiased prediction of diameter, up to ten years into the future.

Current individual tree models consist of predicting equations from current tree age, dominant height (or site index), and stand density (Cao, 2000; Zhao et al., 2004, Sánchez-González et al., 2006, Adame et al., 2008). This approach was not possible here because of the data limitation. Future research can focus on potential differences among the current tree model approach, the quantile regression approach, and the hierarchical mixed model approach for multi-level grouping of plots and trees within plot.

Appendix A: SAS program for quantile regression

```
data one;
  input group tree age dbh;
  if group = 1 then delete;               *Withhold group 1;
proc nlp data = one tech = nmsimp;
  min f;                                  *f is the objective
                                          function to be
                                          minimized;
  decvar a = 2, b = −6, c = −1;
  tau = 0.1;
  dhat = exp(a + b * age ** c);
  if dbh ge dhat
    then f = tau * (dbh − dhat);
    else f = (1 − tau) * (dhat − dbh);
title '10% quantile regression for diameter-age curve';
```

References

Adame, P., Hynynen, J., Cañellas, I., del Río, M., 2008. Individual-tree diameter growth model for rebollo oak (*Quercus pyrenaica* Willd.) coppices. Forest Ecol. Mgt. 255, 1011–1022.
Austin, P.C., Schull, M.J., 2003. Quantile regression: a statistical tool for out-of-hospital research. Acad. Emerg. Med. 10, 789–797.
Avery, T.A., Burkhart, H.E., 2002. Forest Measurements. McGraw-Hill, New York.
Bailey, R.L., Clutter, J.L., 1974. Base-age invariant polymorphic site curves. Forest Sci. 20, 155–159.
Cade, B.S., Noon, B.R., Flather, C.H., 2005. Quantile regression reveals hidden bias and uncertainty in habitat models. Ecology 86, 786–800.
Cao, Q.V., 2000. Prediction of annual diameter growth and survival for individual trees from periodic measurements. Forest Sci. 46, 127–131.

Cao, Q.V., 2007. Incorporating whole-stand and individual-tree models in a stand-table projection system. Forest Sci. 53, 45–49.
Cao, Q.V., Wang, J., 2011. Calibrating fixed- and mixed-effects taper equations. Forest Ecol. Manage. 262, 671–673.
Ducey, M.J., Knapp, R.A., 2010. A stand density index for complex mixed species forests in the northeastern United States. Forest Ecol. Manage. 260, 1613–1622.
Evans, A.M., Finkral, A.J., 2010. A new look at spread rates of exotic diseases in North American forests. Forest Sci. 56, 453–459.
Fang, Z., Bailey, R.L., 2001. Nonlinear mixed effects modeling for slash pine dominant height growth following intensive silvicultural treatments. Forest Sci. 47, 287–300.
Groom, J.D., Hann, D.W., Temesgen, H., 2012. Evaluation of mixed-effects models for predicting Douglas-fir mortality. Forest Ecol. Manage. 276, 139–145.
Haile, G.A., Nguyen, A.N., 2008. Determinants of academic attainment in the United States: a quantile regression analysis of test scores. Educ. Econ. 16, 29–57.
Hall, D.B., Clutter, M., 2004. Multivariate multilevel nonlinear mixed effects models for timber yield predictions. Biometrics 60, 16–24.
Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46, 33–50.
Machado, J.A.F., Mata, J., 2005. Counterfactual decomposition of changes in wage distributions using quantile regression. J. Appl. Econ. 20, 445–465.
Mehtätalo, L., Gregoire, T.G., Burkhart, H.E., 2008. Comparing strategies for modeling tree diameter percentiles from remeasured plots. Environmetrics 19, 529–548.
Meng, S.X., Huang, S., 2009. Improved calibration of nonlinear mixed-effects models demonstrated on a height growth function. Forest Sci. 55, 238–248.
Sánchez-González, M., del Río, M., Cañellas, I., Montero, G., 2006. Distance independent tree diameter growth model for cork oak stands. Forest Ecol. Mgt. 225, 262–270.
SAS Institute Inc., 2008. SAS/STAT 9.2 User's Guide. SAS Institute Inc., Cary, NC.
SAS Institute Inc., 2010. SAS/OR 9.22 User's Guide: Mathematical Programming. SAS Institute Inc., Cary NC.
Zhang, L., Bi, H., Gove, J.H., Heath, L.S., 2005. A comparison of alternative methods for estimating the self-thinning boundary line. Can. J. Forest Res. 35, 1507–1514.
Zhao, D., Border, B., Wilson, M., 2004. Individual-tree diameter growth and mortality models for bottomland mixed-species hardwood stands in the lower Mississippi alluvial valley. Forest Ecol. Mgt. 199, 307–322.