An integrated system for modeling tree and stand survival

Quang V. Cao

Abstract: Traditionally, separate models have been used to predict the number of trees per unit area (stand-level survival) and the survival probability of an individual tree (tree-level survival) at a certain age. This study investigated the development of integrated systems in which survival models at different levels of resolution are related in a mathematical structure. Two approaches for modeling tree and stand survival were considered: deriving a stand-level survival model from a tree-level survival model (approach 1) and deriving a tree survival model from a stand survival model (approach 2). Both approaches rely on finding a tree diameter that yields a tree survival probability equal to the stand-level survival probability. The tree and stand survival models from either approach are conceptually compatible with each other but not numerically compatible. Parameters of these models can be estimated either sequentially or simultaneously. Results indicated that approach 2, with parameters estimated sequentially (first from the stand survival model and then from the derived tree survival model), performed best in predicting both tree- and stand-level survival. Although disaggregation did not help improve prediction of tree-level survival, this method can be used when numerical consistency between stand and tree survival is desired.

Key words: individual-tree model, least squares, loblolly pine, logistic regression, maximum likelihood.

Introduction

Growth and yield models play an important role in forest management. These models are often classified based on level of detail, or resolution, of their outputs, which provide information for the entire stand (whole-stand models), each diameter class (size-class models), or each tree (individual-tree models) (Burkhart and Tomé 2012).

Whole-stand survival, defined as the number of trees surviving per unit area, has been predicted from regression models that are either empirical (Zhang et al. 1993; Diéguez-Aranda et al. 2005; Zhao et al. 2007; Gonzalez-Benecke et al. 2012) or derived from biological principles (García 2009, 2011; Tewari et al. 2014; Stankova 2016). Methods to predict individual-tree survival in terms of either survival status or probability include logistic regression (Hamilton 1974; Monserud 1976; Buchman 1979, 1983; Zhang et al. 1997; Monserud and Sterba 1999) and other approaches (Glover and Hool 1979; Amateis et al. 1989; Guan and Gertner 1991a, 1991b). Most of the papers deal with tree mortality before and during the self-thinning stage, without taking into account the effects of factors such as fire, insects, diseases, etc.

With a myriad of available growth and yield models, the user sometimes has to choose one model for reliability of estimates and another for sufficient detail. The problem is that outputs from these models of different resolutions might be inconsistent with one another. Daniels and Burkhart (1988) introduced the concept of developing a unified mathematical structure for modeling tree and stand growth that can be applied at any level of resolution. The result is an integrated system that can provide consistent growth and yield estimates at various levels of resolution. However, consistent sets of whole-stand survival and individual-tree survival models have not been developed to date.

In this study, two approaches were considered: deriving a stand-level survival model from a tree-level survival model (approach 1) and deriving a tree survival model from a stand survival model (approach 2). Parameters of these models can be estimated either simultaneously or sequentially (first from the original model and then from the derived model).
The objectives of this study were to (i) develop an integrated system for predicting stand-level and tree-level survival using each of the two above approaches, (ii) develop methods to estimate model parameters, and (iii) evaluate combinations of the two approaches and two estimation methods.

Data

Data used in this study were from 200 plots randomly selected from the Southwide Seed Source Study, which included 15 loblolly pine (*Pinus taeda*) L. seed sources planted at 13 locations across 10 southern states (Wells and Wakeley 1966). Each 0.0164 ha plot consisted of 49 trees, planted at a 1.8 m × 1.8 m spacing. Included in the data set were measurements of tree diameters and survival at ages 10, 15, 20, and 25. There was a total of 600 growth periods. Effects of seed sources and locations were not considered in this study.

The data were randomly divided into two groups of 100 plots each (Table 1). The twofold evaluation scheme was applied in this study. Parameters of the stand and tree survival models were estimated from group 1 (considered the fit data) and then used to predict for group 2 (considered the validation data). The same procedure was repeated with group 2 being the fit data and group 1 as the validation data. Finally, predictions from both groups were pooled to compute evaluation statistics for the different methods.

Methods

Approach 1

In this approach, we begin with a tree-level survival model and attempt to derive a stand-level survival model from it.

Tree-level model

The following logistic regression model was employed to predict tree survival probability ($p_{ij}$) of tree $j$ in plot $i$ during a 5-year growth period:

\[ p_{ij} = \frac{1}{1 + \exp(b_0 + b_1H_{1i} + b_2RS_{ij} + b_3N_{ij}/A_{1i} + b_4A_{1i} + b_5d_{ij})} \]

where $A_{1i}$ is stand age (in years) for plot $i$ at the beginning of the growth period; $H_{1i}$ is the dominant height (in m; calculated as the average height of the upper half of trees in each plot ranked by height) for plot $i$ at age $A_{1i}$; $RS_{ij}$ is the number of trees per hectare for plot $i$ at age $A_{1i}$; $N_{ij}$ is the number of trees per hectare for plot $i$ at age $A_{1i}$; $d_{ij}$ is the diameter at breast height (in cm) of tree $j$ in plot $i$ at age $A_{1i}$; and parameters $b_0$ to $b_5$ are regression coefficients. This model is similar to the stand survival model used by Cao (2006, 2014), with the addition of the last term for tree diameter.

Deriving a stand-level model

Let $p_{0j}$ be the stand survival rate, or $p_{0j} = N_{0j}/N_{1j}$, where $N_{0j}$ is the number of surviving trees per hectare for plot $i$ at age $A_{1i}$. Consider an individual tree with diameter $D_{1ij}$ that, when plugged into eq. 1, would yield a tree survival probability equal to $p_{0j}$. This illustrates the concept of using an attribute (tree survival in this case) of a “typical” tree to expand to a stand-level attribute (stand survival). Replacing $d_{ij}$ in eq. 1 with $D_{1ij}$ results in the following stand survival model:

\[ N_{1j} = \frac{N_{0j}}{1 + \exp(b_0 + b_1H_{1i} + b_2RS_{ij} + b_3N_{ij}/A_{1i} + b_4A_{1i} + b_5D_{1ij})} + \epsilon \]

The relationship between $D_{1ij}$ and $d_{ij}$, the quadratic mean diameter at age $A_{1i}$, can be modeled by a power function. By expressing $D_{1ij}$ as a power function of $d_{ij}$, eq. 2 becomes

\[ N_{1j} = \frac{N_{0j}}{1 + \exp(b_0 + b_1H_{1i} + b_2RS_{ij} + b_3N_{ij}/A_{1i} + b_4A_{1i} + b_5(c_{1ij}d_{ij}^{c_{2ij}}))} + \epsilon \]

where parameters $c_1$ and $c_2$ are regression coefficients; and $\epsilon$ is the random error, assumed to be normally distributed with mean 0 and variance $\sigma^2$.

Parameter estimation

Sequential estimation

In this estimation approach, parameters of eqs. 1 and 3 were estimated sequentially. Parameters $b_1$-$b_5$ in eq. 1 were estimated by use of maximum likelihood procedure. The resulting values for these parameters were also used in eq. 3. The least squares method was then employed to estimate the remaining two parameters of eq. 3, $c_1$ and $c_2$.

Simultaneous estimation

Maximum likelihood technique was used to simultaneously estimate parameters of eqs. 1 and 3. Parameter estimates were obtained by maximizing the following combined log-likelihood function:

\[ \ln L = \frac{\ln L_1}{\ln L_{1_{\text{max}}}} + \frac{\ln L_3}{\ln L_{3_{\text{max}}}} \]

where $\ln L_1 = \sum_i \ln L_1$ is the log-likelihood for eq. 1; $z_i$ equals $p_{ij}$ if tree $j$ in plot $i$ is alive and $(1 - p_{ij})$ if it is dead; $L_{1_{\text{max}}}$ is the maximum value of $L_1$, obtained by fitting only eq. 1; $\ln L_3 = -\frac{1}{2} \ln(2\pi^2) - \frac{1}{2} \sum_i (N_{ij} - \bar{N}_{ij})^2$ is the log-likelihood for eq. 3; and $L_{3_{\text{max}}}$ is the maximum value of $L_3$, obtained by fitting only eq. 3. Note that $\ln L_1$ and $\ln L_3$ have different magnitudes because there

---

Table 1. Means (and standard deviations) of stand and tree attributes by group and age.

<table>
<thead>
<tr>
<th>Group</th>
<th>Attribute</th>
<th>Stand age (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>1 (100 plots)</td>
<td>Dominant height (m)</td>
<td>9.0 (1.2)</td>
</tr>
<tr>
<td></td>
<td>No. of trees per hectare</td>
<td>1987 (642)</td>
</tr>
<tr>
<td></td>
<td>Basal area (m²·ha⁻¹)</td>
<td>21.3 (6.2)</td>
</tr>
<tr>
<td></td>
<td>Tree diameter (cm)</td>
<td>11.7 (3.1)</td>
</tr>
<tr>
<td>2 (100 plots)</td>
<td>Dominant height (m)</td>
<td>9.2 (1.2)</td>
</tr>
<tr>
<td></td>
<td>No. of trees per hectare</td>
<td>1976 (629)</td>
</tr>
<tr>
<td></td>
<td>Basal area (m²·ha⁻¹)</td>
<td>22.1 (6.0)</td>
</tr>
<tr>
<td></td>
<td>Tree diameter (cm)</td>
<td>11.6 (2.9)</td>
</tr>
</tbody>
</table>
are many more tree-level observations than plot-level observations. They are therefore scaled by $L_{\text{max}}$ and $L_{\text{max}}$, respectively, before being combined in eq. 4.

**Approach 2**

In this approach, which is opposite to approach 1, a tree-level survival model is derived from a stand-level survival model.

**Stand-level model**

The following stand-level model (Cao 2006, 2014) was used to predict stand survival after a 5-year growth period:

$$N_i = \frac{N_u}{1 + \exp(b_0 + b_1H_i + b_2RS_i + b_3N_i/A_i + b_4(A_i) + e_i)}$$

(5)

**Deriving a tree-level model**

If $D_{3G}$ is the diameter of a tree having the same survival probability as the stand survival rate ($N_3/N_0$), then a tree-level survival model can be derived from eq. 5 as follows:

$$p_i = \frac{1}{1 + \exp\{b_0 + b_1H_i + b_2RS_i + b_3N_i/A_i + b_4(A_i) + b_5(d_i - D_{3G})\}}$$

(6)

Note that when $d_i = D_{3G}$, predictions for tree-level and stand-level survival rates are identical. Using the same relationship between $D_{3G}$ and $D_{QG}$ as earlier assumed, eq. 6 can be rewritten:

$$p_i = \frac{1}{1 + \exp\{b_0 + b_1H_i + b_2RS_i + b_3N_i/A_i + b_4(A_i) + b_2(d_i - c_2D_{QG})\}}$$

(7)

**Parameter estimation**

**Sequential estimation**

In this estimation approach, parameters of eqs. 5 and 7 were estimated sequentially. The least squares method was employed to estimate parameters $b_0$-$b_4$ in eq. 5. The resulting values for these parameters were also used in eq. 7, and the remaining three parameters $b_5$, $c_2$, and $c_2$ were estimated by using the maximum likelihood procedure.

**Simultaneous estimation**

Similar to approach 1, the maximum likelihood technique was used to simultaneously estimate parameters of eqs. 5 and 7. Parameter estimates were obtained by maximizing the following combined log-likelihood function:

$$\ln L = \frac{\ln L_5}{\ln L_{\text{max}}} + \frac{\ln L_7}{\ln L_{\text{max}}}$$

(8)

where $\ln L_5$ is the log-likelihood for eq. 5; $L_{\text{max}}$ is the maximum value of $L_5$, obtained by fitting only eq. 5; ln $L_7$ is the log-likelihood for eq. 7; and ln $L_{\text{max}}$ is the maximum value of $L_7$, obtained by fitting only eq. 7.

**Evaluation**

For each of the two approaches, coefficients obtained from one group were used to predict for the other group. Predicted values from both groups were then used to compute evaluation statistics.

**Stand survival prediction**

The following statistics were computed for stand-level evaluation:

$$\text{Mean difference (MD)} = \frac{1}{m} \sum (N_i - \hat{N}_i)$$

(9)

$$\text{Mean absolute difference (MAD)} = \frac{1}{m} \sum j |N_{ij} - \hat{N}_{ij}|$$

(10)

$$\text{Fit index (FI)} = 1 - \frac{\sum_j (N_{ij} - \hat{N}_{ij})^2}{\sum_j (N_{ij} - \Sigma_i)^2}$$

(11)

where $m$ is the number of plots; $\hat{N}_{ij}$ is the predicted number of surviving trees per hectare for plot $i$ at age $A_{ij}$; $\Sigma_i$ is the average number of surviving trees per hectare at age $A_{ij}$; and $\Sigma_i$ denotes the sum for $i$ from 1 to $m$.

**Tree survival prediction**

Tree-level survival predictions were evaluated from

$$\text{Mean difference (MD)} = \frac{\sum_i \sum_j |y_{ij} - p_{ij}|}{\sum_i n_i}$$

(12)

where $y_{ij}$ equals 1 if tree $j$ in plot $i$ was alive and 0 if it was dead; $\Sigma_i$ denotes the sum for $i$ from 1 to $m$; $\Sigma_j$ denotes the sum for $j$ from 1 to $n_i$; and $n_i$ is the number of trees in plot $i$ at age $A_i$.

$$\text{Mean absolute difference (MAD)} = \frac{\sum_i \sum_j |y_{ij} - p_{ij}|}{\sum_i n_i}$$

(13)

AUC is the area under the receiving operating characteristic (ROC) curve. The range for AUC is between 0.5 and 1: the higher the AUC value, the better the fit.

To display the relative position of each method, Poudel and Cao (2013) introduced the relative rank, defined as

$$R_i = 1 + \frac{(k - 1)(S_i - S_{\text{min}})}{S_{\text{max}} - S_{\text{min}}}$$

(14)

for the minimization objective and

$$R_i = k - \frac{(k - 1)(S_i - S_{\text{min}})}{S_{\text{max}} - S_{\text{min}}}$$

(15)

for the maximization objective, where $R_i$ is the relative rank of method $i$ ($i = 1, 2, ..., k$), $k$ is the number of methods evaluated, $S_i$ is the evaluation statistic produced by method $i$, $S_{\text{min}}$ is the minimum value of $S_i$, and $S_{\text{max}}$ is the maximum value of $S_i$. Note that $R_i$ is a real number rather than an integer, and for either the minimization or the maximization objective, the best method receives a rank of 1 and the worst method receives a rank of $k$. After a relative rank was computed separately for each statistic of each method, a final rank was calculated based on the sum of all ranks for each method.

**Results and discussion**

Table 2 shows parameter estimates from two approaches, by group and by estimation method. Evaluation statistics are shown for survival prediction at the tree level (Table 3). A relative rank for tree survival prediction was computed for each combination of approach and estimation method. A final rank was then calculated based on the sum of all ranks for each method. Table 4 presents evaluation statistics for stand-level survival prediction, with rankings computed in a similar manner as in the previous table.
Table 2. Parameter estimates from two approaches, by group and by estimation method.

<table>
<thead>
<tr>
<th>Group</th>
<th>Approach</th>
<th>Method</th>
<th>( b_0 )</th>
<th>( b_1 )</th>
<th>( b_2 )</th>
<th>( b_3 )</th>
<th>( b_4 )</th>
<th>( b_5 )</th>
<th>( c_1 )</th>
<th>( c_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Sequential</td>
<td>14.7540</td>
<td>-0.2922</td>
<td>-36.6331</td>
<td>-0.0197</td>
<td>32.8651</td>
<td>-0.3828</td>
<td>1.0999</td>
<td>0.9367</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Simultaneous</td>
<td>15.4257</td>
<td>-0.3278</td>
<td>-39.3029</td>
<td>-0.0206</td>
<td>37.8676</td>
<td>-0.3720</td>
<td>1.3058</td>
<td>0.8766</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>Sequential</td>
<td>21.8423</td>
<td>-0.9268</td>
<td>-60.9684</td>
<td>-0.0295</td>
<td>56.7828</td>
<td>-0.4248</td>
<td>0.9099</td>
<td>1.0116</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Simultaneous</td>
<td>15.1406</td>
<td>-0.6255</td>
<td>-43.2750</td>
<td>-0.0190</td>
<td>30.5898</td>
<td>-0.4092</td>
<td>1.3994</td>
<td>0.8440</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>Sequential</td>
<td>12.6249</td>
<td>-0.2252</td>
<td>-33.9138</td>
<td>-0.0187</td>
<td>37.6765</td>
<td>-0.3492</td>
<td>1.0594</td>
<td>0.9532</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Simultaneous</td>
<td>13.8856</td>
<td>-0.2876</td>
<td>-36.9750</td>
<td>-0.0198</td>
<td>39.5395</td>
<td>-0.3359</td>
<td>1.3025</td>
<td>0.8766</td>
</tr>
<tr>
<td>2</td>
<td>Sequential</td>
<td>23.0569</td>
<td>-0.9647</td>
<td>-64.7751</td>
<td>-0.0317</td>
<td>62.3571</td>
<td>-0.4116</td>
<td>0.6527</td>
<td>1.1398</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Simultaneous</td>
<td>16.3241</td>
<td>-0.6964</td>
<td>-45.7509</td>
<td>-0.0211</td>
<td>37.4917</td>
<td>-0.3966</td>
<td>0.8699</td>
<td>1.0236</td>
</tr>
</tbody>
</table>

Table 3. Evaluation statistics for tree-level survival prediction.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Estimation method</th>
<th>MD</th>
<th>MAD</th>
<th>AUC</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sequential</td>
<td>0.0000</td>
<td>0.2253</td>
<td>0.8036</td>
<td>1.54</td>
</tr>
<tr>
<td></td>
<td>Simultaneous</td>
<td>0.0064</td>
<td>0.2292</td>
<td>0.8010</td>
<td>4.00</td>
</tr>
<tr>
<td>2</td>
<td>Sequential</td>
<td>-0.0006</td>
<td>0.2172</td>
<td>0.8024</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Simultaneous</td>
<td>-0.0091</td>
<td>0.2169</td>
<td>0.8058</td>
<td>1.32</td>
</tr>
</tbody>
</table>

Note: Bold, italic numbers denote the best method for each criterion; underlined numbers denote the worst method.

Table 4. Evaluation statistics for stand-level survival prediction.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Evaluation method</th>
<th>MD</th>
<th>MAD</th>
<th>Fit index</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sequential</td>
<td>-6.82</td>
<td>171.58</td>
<td>0.8070</td>
<td>4.00</td>
</tr>
<tr>
<td></td>
<td>Simultaneous</td>
<td>-0.32</td>
<td>171.72</td>
<td>0.8078</td>
<td>3.27</td>
</tr>
<tr>
<td>2</td>
<td>Sequential</td>
<td>-22.96</td>
<td>157.83</td>
<td>0.8330</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Simultaneous</td>
<td>-20.24</td>
<td>165.32</td>
<td>0.8259</td>
<td>2.63</td>
</tr>
</tbody>
</table>

Note: Bold, italic numbers denote the best method for each criterion; underlined numbers denote the worst method.

Tree survival prediction

The tree survival model from approach 1 with the sequential estimation method ranked first in terms of MD, whereas the derived tree model from approach 2 (simultaneous estimation method) ranked first based on MAD and AUC (Table 3). The overall winner for tree survival prediction based on all three statistics, however, was the derived tree model from approach 2, with parameters estimated sequentially.

Stand survival prediction

Whereas the derived stand survival model from approach 1 (simultaneous estimation method) ranked first in terms of MD, the stand model from approach 2 with parameters estimated by use of the sequential estimation method ranked first based on MAD and fit index and was also the overall winner in predicting stand survival (Table 4).

Approach 1 vs. approach 2

For each approach, a rank total was obtained by summing up four relative ranks (for each estimation method at tree and stand levels). Compared with approach 1, approach 2 was clearly superior, having consistently lower (better) relative ranks that yielded a rank total of 5.95 versus 12.81. For the sequential estimation method, the superior performance of approach 2 in predicting stand-level survival makes sense. Parameters of the stand survival model were optimized exclusively for stand survival prediction in approach 2, but only partially in approach 1. Results from predicting tree-level survival, also for the sequential estimation method, were more intriguing. Approach 2 puts a lot of emphasis on stand-level prediction and still did better than approach 1, which optimizes tree-level survival. In a sense, this is similar to a disaggregative model that was adjusted to match stand survival prediction and still performed better than the unadjusted tree survival model (Cao 2006, 2014; Hevia et al. 2015). One explanation might be that tree survival is variable and its prediction could be improved by linking with stand survival, which is more stable.

Sequential vs. simultaneous estimation

The total of ranks for each estimation method from Tables 3 and 4 revealed that the sequential estimation method was better than the simultaneous estimation method (7.54 versus 11.22). The sequential method applied to approach 2 was best at both tree and stand levels. On the other hand, results were worst at the stand level when the sequential method was applied to approach 1.

Results for approach 2 suggest that better performance was obtained when the integrated system was optimized for stand-level survival (sequential estimation method) rather than optimized equally for both stand- and tree-level survival (simultaneous estimation method). It seems that the uncertainty in predicting tree survival might “dilute” the accuracy and precision of predicting stand survival in the simultaneous estimation method.

Compatibility

Either approach results in a system of tree- and stand-level survival models that are conceptually compatible because one model is derived from the other. However, they are not numerically compatible in the sense that the sum of predicted tree survival does not equal the predicted stand survival.

The tree survival eq. 7 from approach 2 was then adjusted to match the stand survival prediction from eq. 5. Equation 16 shows the method of disaggregation (Cao 2014) that yields the adjusted tree survival probability (\( \hat{p}_i \)) by use of an adjustment coefficient for each plot (\( \alpha_i \)) to ensure that stand survival predictions from both tree- and stand-level models are numerically compatible:

\[
\hat{p}_i = p_i^n
\]

Results were mixed: the disaggregation method improved the MAD value (0.2128 versus 0.2169) while producing worse values for MD (~0.0137 versus ~0.0006) and AUC (0.7910 versus 0.8058). Apparently, disaggregation did not help improve prediction of tree-level survival in this case, contrary to findings by Cao (2006, 2014) and Hevia et al. (2015). The approach is useful, however, when numerical consistency between stand-level and tree-level survival is preferable.

Summary and conclusions

In this study two approaches and two estimation methods were evaluated. Approach 1 starts with a tree survival model from which a stand-level survival model is derived. Conversely, in approach 2, a tree survival model is derived from a stand survival model. Parameters of the models from both approaches can be estimated either simultaneously or sequentially, first from the original model and then from the derived model. Results indicated that the combination of approach 2 and the sequential estimation method performed best in predicting survival at both tree and stand levels. The tree- and stand-level survival models from this integrated system are conceptually but not numerically com-
patible. Although disaggregation did not help improve prediction of tree-level survival, this method can be used when numerical consistency between stand and tree levels is desired.

**Acknowledgement**

This project was supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, McIntire-Stennis project LAB94223.

**References**


