A method to derive a tree survival model from any existing stand survival model

Quang V. Cao

Abstract: This study addresses a situation in which a forest manager has been using a whole-stand model that seems to predict well for their stands and now wants to derive an individual-tree model from it to form an integrated system that can perform well at both stand and tree levels. A simple method was developed to derive tree survival models from three existing stand-level survival models. The derived tree survival models were based on the difference between the diameter of a given tree and the diameter at which tree and stand survival probabilities are equal. For stand survival prediction, each stand model performed less adequately than its derived tree model, and one of the derived tree survival models was the best overall. For tree survival prediction, the same derived tree model also performed best overall. Even though only three stand-level survival models were considered in this study, the method presented here should be applicable to any stand survival model. When no tree survival data were available, tree survival models derived from stand survival models ranked lowest in terms of performance but produced acceptable evaluation statistics for predicting tree-level survival.

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

Introduction

Contemporary forest managers make decisions partly based on forecasts provided by growth and yield models. Depending on the level of detail, or resolution, of their outputs, these models are often broadly classified as whole-stand models, size-class models, or individual-tree models (Burkhart and Tome 2012); however, outputs from these models of different resolutions might be inconsistent with one another. One model might provide better reliability of estimates, whereas another model might provide more details.

In this paper, stand survival refers to the number of surviving trees per unit area, and tree survival refers to the status (dead or alive) of a given tree. Regression models have been used to predict cumulative survival at the stand level. These models are either empirical (Zhang et al. 1993; Diéguez-Aranda et al. 2005; Zhao et al. 2007; Gonzalez-Benecke et al. 2012) or are derived from biological principles (Garcia 2009, 2011; Tewari et al. 2014; Stankova 2016). In contrast, logistic regression has generally been used to predict survival at the individual-tree level (Hamilton 1974; Monserud 1976; Buchman 1979, 1983, Zhang et al. 1997; Monserud and Sterba 1999), even though other approaches have also been used (Glover and Hool 1979; Amateis et al. 1989; Guan and Gertner 1991).

Growth and yield models have traditionally been modeled separately at stand and tree levels. Daniels and Burkhart (1988) introduced a revolutionary concept of developing a unified mathematical structure for modeling tree and stand growth, which 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. Cao (2017) applied this concept and experimented with two approaches: deriving a stand-level survival model from a tree-level survival model and deriving a tree survival model from a stand survival model. The latter approach delivered satisfactory results.

This study addresses a likely scenario in which a forest manager has been using a whole-stand model that seems to predict well for their stands and now wants to derive an individual-tree model from it to form an integrated system that can perform well at both stand and tree levels. The advantage of this integrated system over an independent tree survival model is that both stand-level and...
Table 1. Stand and tree attributes at the beginning of each growth period by group.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Age of growth period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10–15 years</td>
</tr>
<tr>
<td>Group 1</td>
<td></td>
</tr>
<tr>
<td>Dominant height (m)</td>
<td>9.5 (1.3)</td>
</tr>
<tr>
<td>Density (trees·ha⁻¹)</td>
<td>1875 (386)</td>
</tr>
<tr>
<td>Basal area (m²·ha⁻¹)</td>
<td>23.0 (6.4)</td>
</tr>
<tr>
<td>Tree diameter (cm)</td>
<td>11.9 (2.8)</td>
</tr>
<tr>
<td>Group 2</td>
<td></td>
</tr>
<tr>
<td>Dominant height (m)</td>
<td>8.9 (1.1)</td>
</tr>
<tr>
<td>Density (trees·ha⁻¹)</td>
<td>1912 (720)</td>
</tr>
<tr>
<td>Basal area (m²·ha⁻¹)</td>
<td>19.4 (6.2)</td>
</tr>
<tr>
<td>Tree diameter (cm)</td>
<td>11.0 (2.8)</td>
</tr>
<tr>
<td>Group 3</td>
<td></td>
</tr>
<tr>
<td>Dominant height (m)</td>
<td>9.4 (1.0)</td>
</tr>
<tr>
<td>Density (trees·ha⁻¹)</td>
<td>1857 (653)</td>
</tr>
<tr>
<td>Basal area (m²·ha⁻¹)</td>
<td>21.5 (5.9)</td>
</tr>
<tr>
<td>Tree diameter (cm)</td>
<td>11.8 (2.7)</td>
</tr>
</tbody>
</table>

Note: Values are means, with standard deviations in parentheses. There were 30 plots in each growth period of each group.

tree-level models have the same mathematical structure, even though they might not be numerically compatible. This study, which is an extension of the work by Cao (2017), widens the scope of the original research by allowing the tree survival model to be derived from any existing stand survival model, whether or not tree survival data are available.

The objectives of this study were to (i) develop a method to derive an individual-tree survival model from any existing stand-level survival model and (ii) evaluate the tree survival models derived from such stand survival models.

Methods
Data
Data used in this study were from 270 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 states in the southern United States (Wells and Wakeley 1966). Each plot had an area of 0.0164 ha and consisted of 49 trees (ap-approximately 3000 trees·ha⁻¹) planted at a spacing of 1.8 m x 1.8 m. Tree survival varied greatly from plot to plot, ranging from 610 to 2929 trees·ha⁻¹ at age 10 years, with a mean of 2102 trees·ha⁻¹.

The data were randomly divided into three groups of 90 plots each (Table 1). For each group, tree diameters and survival were measured from 30 plots for each of three growth periods: 10–15, 15–20, and 20–25 years.

Stand-level survival models
Three stand survival models were selected for this study because of their successful applications in the past. Their coefficients were estimated from the data in Table 1.

Clutter and Jones (1980)

\[
\hat{N}_{2i} = 100 \left( \frac{N_{ui}}{100} \right)^{a_i} + a_i \left( \frac{A_{ui}}{10} - \frac{A_{ui}}{10} \right)^{\mu_i}
\]

Baldwin and Feduccia (1987)

\[
\hat{N}_{2i} = 100 \left( \frac{N_{ui}}{100} \right)^{a_i} + \left( \frac{A_{ui}}{10} + \frac{A_{ui}}{10} \right)^{\mu_i}
\]

Cao (2006)

\[
\hat{N}_{2i} = 1 + \exp \left[ a_i + a_i H_{ui} + a_i R_{ui} + a_i (N_{ui} A_{ui}) + (a_{5i} A_{ui}) \right]
\]

where \( A_{ui} \) and \( A_{ui} \) are stand ages (in years) at the beginning and end of the growth period, respectively, for plot \( i \); \( N_{ui} \), \( H_{ui} \), and \( R_{ui} \) are stand density (in trees per hectare), dominant height (in metres), and relative spacing at age \( A_{ui} \) (\( R_{ui} = \frac{10000(N_{ui})^0.5}{H_{ui}} \) (Wilson 1946)), respectively; \( i \) is the site index (in metres) at base age 25 years; \( a_i \) is the predicted stand density at age \( A_{ui} \); and \( a_i \) is the regression coefficient.

Tree-level survival models
In this study, two scenarios that depend on the presence or absence of tree survival data were considered.

With tree survival data
The tree survival data were used to estimate coefficients of the individual-tree survival model.

Method 1
The logistic regression model by Cao (2017) was employed to predict tree survival probability \( p_i \) of tree \( j \) in plot \( i \) during a 5-year growth period:

\[
p_i = \frac{1}{1 + \exp(b_0 + b_1 H_i + b_2 R_{ui} + b_3 (N_{ui} A_{ui}) + b_4 A_{ui} + b_5 d_{ui})}
\]

where \( d_{ui} \) is diameter at breast height (DBH, in centimetres; breast height = 1.30 m) of tree \( j \) in plot \( i \) at age \( A_{ui} \), and \( b_k \) is the regression coefficient.

The disaggregation method (Cao 2010, 2014) was also applied to compute the adjusted tree survival probability \( \hat{p}_{ij} \) as follows:

\[
\hat{p}_{ij} = \frac{p_{ij}}{p_{ij} + \alpha}
\]

where \( \alpha \) is the adjustment coefficient used to match the sum of the adjusted tree survival probabilities to predictions from each of the three stand survival models (eqs. 1–3).

Hereafter, methods 1a and 1b denote the unadjusted and disaggregated models, respectively.

Method 2
The tree survival model from the integrated system by Cao (2017) is

\[
p_i = \frac{1}{1 + \exp(c_0 + c_1 H_i + c_2 R_{ui} + (c_3 N_{ui} A_{ui}) + (c_4 A_{ui} + c_5 d_{ui} - D_{ui})}
\]

where \( D_{ui} \) is the tree diameter that would yield a tree survival probability equal to that of the stand survival prediction, and \( c_k \) is the regression coefficient, or \( p_y = \frac{N_{ui}}{N_{ui}} \) (the proportion of surviving and starting number of trees). Equation 6 can be simplified as follows:

\[
\hat{p}_{ij} = \frac{1}{1 + \exp(b_0 + b_1 d_{ui} - D_{ui})}
\]

where \( D_{ui} \) is expressed as \( D_{ui} = b_1 d_{ui}^{b_1} \), with \( D_{ui} \) being the quadratic mean diameter of plot \( i \) at the beginning of the growth period. In other words, when \( d_{ui} = D_{ui} \),

Combining eqs. 7 and 8 gives

Published by NRC Research Press
Equation 9 is therefore a tree survival model derived from \( \hat{N}_{i,j} \), which is the output of a stand survival model for plot \( i \). \( \hat{N}_{i,j} \) can be obtained from one of the previously mentioned stand survival models (by use of eqs. 1–3) or any other stand-level survival model. Although the derivation method is demonstrated only for these three stand models, it should be applicable to any existing stand survival model. The previously mentioned disaggregation method was then used to adjust the tree survival model. The resulting survival model. The range for AUC is between 0.5 (worst) and 1 (best). The AUC values from all three groups were then pooled to compute evaluation statistics. MD and MAD are measures of accuracy and precision, respectively, whereas FI is a measure of both.

**Tree survival prediction**

Tree-level survival predictions were evaluated from MD, MAD, and area under the receiving operating characteristic (ROC) curve (AUC).

The relative ranking system, introduced by Poudel and Cao (2013), was used in this study to determine the relative position of the tree survival models, and (iii) evaluating the models. Predicted values from all three groups were then pooled to compute evaluation statistics.

**Stand survival prediction**

Mean difference (MD), mean absolute difference (MAD), and fit index (FI) were computed for stand-level evaluation:

\[
\text{MD} = \frac{1}{m} \sum_{i} (N_{i,j} - \hat{N}_{i,j})
\]

\[
\text{MAD} = \frac{1}{m} \sum_{i} |N_{i,j} - \hat{N}_{i,j}|
\]

\[
\text{FI} = 1 - \frac{\sum_{i} (N_{i,j} - \hat{N}_{i,j})^2}{\sum_{i} (N_{i,j} - \bar{N}_{i,j})^2}
\]

where \( m \) is the number of plots; \( N_{i,j} \) and \( \hat{N}_{i,j} \) are the observed and predicted numbers, respectively, of surviving trees per hectare for plot \( i \) at age \( A_{i,j} \); \( \bar{N}_{i,j} \) is the mean number of surviving trees per hectare at age \( A_{i,j} \); and \( \Sigma \) denotes the sum for \( i \) from 1 to \( m \). For tree-level models, stand survival prediction is the sum of predicted tree survival probabilities, scaled by plot area.

**Evaluation**

A three-fold evaluation scheme (Table 2) was used in this study. This scheme simulated the case of deriving a tree model from an existing stand survival model by ensuring that separate data were used for (i) developing the stand survival models, (ii) deriving the stand-level survival model, obtained either from one of eqs. 1–3 or from any existing stand-level survival model.

The range for AUC is between 0.5 (worst) and 1 (best). The AUC has been used as a measure of predictive accuracy of survival models in medical research (Heagerty and Zheng 2005) and forestry research (Weiskittel et al. 2016).
Results and discussion

Table 3 shows evaluation statistics for stand-level survival prediction. The tree model derived from Cao (2006) (method 2a) performed best in predicting stand survival, whereas the stand-level model by Baldwin and Feduccia (1987) ranked last among the seven alternatives.

Method 2a derived from Cao (2006) also performed best for tree survival prediction (Table 4), closely followed by the unadjusted tree model (method 1a, rank of 1.11). Method 3b, derived from different stand-level models using no tree survival data, produced ranks ranging from 10.53 to 13.00. The models from method 3b formed a group at the bottom of the rankings, separate from the remaining methods.

These results show that the tree survival model derived from Cao’s (2006) stand survival model (method 2a) provided the best prediction not only at the tree level, but also at the stand level, surpassing the performance of the stand model from which it was derived. It is impressive that the summation of individual tree probabilities from method 2a outperformed the direct prediction of stand survival from Cao’s (2006) model.

Table 5 shows parameter estimates and their standard errors from Cao’s (2006) stand survival model and its derived tree survival model (method 2a), based on the pooled data.

Stand-level survival prediction

Stand-level survival models

Among the three stand-level survival models, the Cao (2006) model received the best overall ranking, mainly based on better values of MAD and FI (Table 3). The Baldwin and Feduccia (1987) model, which was derived from the Clutter and Jones (1980) model (modified by the addition of a site index variable), turned out to be slightly inferior to the Clutter and Jones (1980) model.

Stand-level models versus tree-level survival model

The sum of predicted tree survival probabilities from the tree-level model (method 1a) were used for stand survival prediction. Surprisingly, this tree model ranked better than the Cao (2006) stand-level model (rankings of 2.57 and 5.27, respectively), even though it produced worse values of MAD and FI than the Cao (2006) model (Table 3). These results suggest that stand-level models are not necessarily better than tree-level models in predicting...
stand survival, contrary to some findings in the literature (Qin and Cao 2006; Cao 2014; Hevia et al. 2015), which favor the direct prediction of stand models over the extra summation step of tree models. Consequently, disaggregation, which adjusts outputs from tree-level models to match those from stand-level models, is not justified in these cases.

**Stand-level models versus derived tree models**

Method 2a involved deriving a tree survival model from each of the three stand-level models. The sum of predicted tree survival probabilities from each tree model provided better stand survival prediction in terms of MAD and FI than the stand-level model from which it was derived. These findings were true for all three stand survival models (Table 3). Stand survival predictions from the tree models derived from Clutter and Jones (1980) and Baldwin and Feduccia (1987) were more biased (higher absolute MD values) than their stand-level model counterparts, but method 2a derived from Cao (2006) yielded a better MD value than the original Cao (2006) model, which produced the worst MD value of −19.71 trees·ha⁻¹.

Each stand-level model and its derived tree model (method 2a) were conceptually compatible because they formed an integrated system of equations; however, they were not numerically compatible. The fact that a derived tree model actually produced better prediction of stand survival than its corresponding stand model was an unexpected result because stand survival models have been found to outperform tree models in predicting stand survival (Qin and Cao 2006; Cao 2014; Hevia et al. 2015). A plausible reason for this result is that the tree survival model from method 2 consists of Cao’s (2016) stand model with the least squares parameter estimates and extra tree information that, because of the Dᵢ constraint, appeared to help improve stand survival prediction.

**Tree-level survival prediction**

**Without disaggregation**

The tree survival model (method 1a) produced better MD values than the three derived tree models (method 2a); however, the tree model derived from Cao (2006) performed better in terms of MAD and AUC (Table 4).

Among the three stand survival models, the best stand model (Cao 2006) produced the best derived tree model (overall rank of 1.00), which edged out method 1a (rank of 1.11) in terms of predicting tree-level survival.

**With disaggregation**

**Method 1a versus method 1b**

The unadjusted tree survival model (method 1a) produced the best AUC value (0.7624) as compared with the three disaggregated models (method 1b; AUC values ranging from 0.7309 to 0.7447). Its MAD and MD values were in the middle of the statistics from the three disaggregated models. Overall, the unadjusted tree survival model (method 1a) produced a much higher rank (1.11) than the tree disaggregated models (ranks varied between 4.31 and 6.86).

**Method 2a versus method 2b**

For the derived tree survival models, the unadjusted models (method 2a) ranked better than the disaggregated models (method 2b) in predicting tree survival, except for the tree model derived from Clutter and Jones (1980) (Table 4). Zhang et al. (2011) showed that disaggregation from observed stand survival resulted in much better tree survival prediction than that from an unadjusted tree model. The logical conclusion was that a tree model disaggregated from an excellent stand model should perform well. Because the tree models outperformed the stand models in predicting stand survival as previously discussed, it makes sense that disaggregation was not only unnecessary, but also detrimental in terms of predictive accuracy.

**Disaggregation: with tree survival data versus without tree survival data**

When it is desired that a tree survival model yields the same stand survival as predicted from a given stand model, disaggregation can be carried out either from a tree survival model with parameters estimated from a tree-survival data set (methods 1b and 2b) or from a simple function that requires no data (method 3b). Even though all methods (1b, 2b, and 3b) yielded identical MD values for each stand survival model, methods 1b and 2b consistently produced better MAD and AUC values than those of method 3b, as was expected (Table 4). Because the Cao (2006) model was the best among the three stand survival models for this data set, the tree model derived from Cao (2006) was also the best tree survival model among the three models from method 3b. Considering that method 3b requires no survival data, it is surprising that this tree model from method 3b approached method 1b in terms of AUC (0.7212 and 0.7447, respectively) and MAD (0.2772 and 0.2534, respectively). The reason for this might be that the negative exponential CDF used in method 3b was sufficient to model the general trend of tree survival, whereas the tree survival models in other methods were more sensitive to plot variations.

**Conclusions**

In this study, methods were developed to derive tree survival models from three existing stand-level survival models. For stand survival prediction, each stand model performed worse than its derived tree model, and the tree survival model derived from Cao (2006) was the best overall. For tree survival prediction, the same tree survival model derived from Cao (2006) also performed best overall. Even though only three stand-level survival models were considered in this study, the method presented here should be applicable to any stand survival model.

### Table 5. Parameter estimates (and their standard errors (SE)) of the stand- and tree-level survival models, based on all observations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stand-level model</td>
<td>a1</td>
<td>13.7321</td>
</tr>
<tr>
<td></td>
<td>a2</td>
<td>−0.5943</td>
</tr>
<tr>
<td></td>
<td>a3</td>
<td>−37.5397</td>
</tr>
<tr>
<td></td>
<td>a4</td>
<td>−0.0205</td>
</tr>
<tr>
<td></td>
<td>a5</td>
<td>36.5848</td>
</tr>
<tr>
<td>Tree-level model</td>
<td>b1</td>
<td>0.7902</td>
</tr>
<tr>
<td></td>
<td>b2</td>
<td>1.0637</td>
</tr>
<tr>
<td></td>
<td>b3</td>
<td>−3.7471</td>
</tr>
</tbody>
</table>

Note: The stand-level model is $N_{i} = N_{i0} \times \exp \left( a_{1}H_{i} + a_{2}R_{S, i} + a_{3}N_{i0}A_{i} + a_{4}N_{i0}D_{i} + a_{5}N_{i0}A_{i}D_{i} \right)$, and the tree-level model is $p_{ij} = \frac{1 + \exp \left( \frac{b_{1}d_{i} + b_{2}D_{i} + b_{3}D_{i}^{2}}{b_{4}D_{i} + b_{5}D_{i}^{2}} \right)}{1}$, for definitions of variables.
Without tree survival data, the derived tree models ranked worst in terms of performance but produced decent evaluation statistics for predicting tree-level survival, as was expected. The fact that the actual magnitudes of the error metrics for method 3b are not consequentially different from those for other methods shows that this could be a viable approach if a tree survival model is needed when no tree survival data are available.

This study shows that the integrated system of stand- and tree-level survival models is a better alternative to separate stand- and tree-level models. Furthermore, this study suggests a reasonable method of constructing a tree survival model, even when no tree survival data are available.

Acknowledgements

This project was supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, McIntire-Stennis project LAB94379.

References


