Modeling tree mortality in relation to climate, initial planting density, and competition in Chinese fir plantations using a Bayesian logistic multilevel method

Xiongqing Zhang, Quang V. Cao, Aiguo Duan, and Jianguo Zhang

Abstract: Tree mortality models are important tools for simulating forest dynamic processes, and logistic regression is widely used for modeling tree mortality. However, most of the mortality models that have been developed generally ignore the hierarchical structure. In this study, Bayesian logistic multilevel mortality models were developed with the independent variables of initial planting density, competition, site index, and climate factors in Chinese fir (Cunninghamia lanceolata (Lamb.) Hook.) plantations in southern China. The results showed that a Bayesian three-level model was best for describing tree mortality data with multiple sources of unobserved heterogeneity compared to fixed-effects and two-level models. The variance partition coefficient of tree mortality due to the tree level was much larger than that due to the plot level. The initial planting density and site index were positively correlated with mortality and symmetric competition was negatively correlated. For climate variables, the mortality probability decreased with the increasing mean annual temperature and previous summer mean temperature. By contrast, the mortality probability increased with the increasing previous winter mean minimum temperature and annual heat–moisture index. Identifying different sources of variation in tree mortality will help further our understanding of the factors that drive tree mortality during climate change.

Key words: Bayesian logistic multilevel model, climate, initial planting density and competition, tree mortality, variance partition coefficient.

Introduction

Tree mortality, one of the main components of forest succession, is important for the maintenance of biological and structural diversity in forest ecosystems (Bigler and Bugmann 2003; Zhang et al. 2011). A complex of endogenous and exogenous factors acting as induction agents lead to tree death during forest succession (van Mantgem et al. 2009). Among endogenous factors, perhaps the best known cause of increasing tree mortality probabilities is increasing competition for water, nutrients, and light within a stand, resulting from increasing stand density and stand basal area (Franklin et al. 2002). By manipulating the stand basal area and density, initial planting density has a significant effect on intertree competition, thereby affecting tree mortality. Lutz and Halpern (2006) reported that the initial planting density correlates strongly with suppression-induced mortality.

In addition to endogenous factors, the effect of climate on tree mortality was identified as an important exogenous factor (Breshears et al. 2005; Adams et al. 2009). Tree mortality studies were per-
formed primarily in established boreal or temperate areas where temperature is often a limiting factor. Additionally, temperature conditions associated with water availability are an important determining factor affecting tree mortality in these areas (Mueller et al. 2005; van Mantgem et al. 2009; Peng et al. 2011; Zhang et al. 2014). Phillips et al. (2009, 2010) reported that tropical forests also suffer increasing tree mortality probabilities in response to moisture stress. Allen et al. (2010) reviewed the research of drought-induced tree mortality and revealed emerging climate change risks for forests worldwide.

The accurate prediction of tree mortality is an essential feature in any individual forest growth and yield system. Hamilton (1974) introduced logistic regression to modeling tree mortality and found that logistic regression is a good choice for modeling tree mortality. Since then, logistic regression has been used to model tree mortality for a variety of tree species (e.g., Avila and Burkhart 1992; Yang et al. 2003; Boeck et al. 2014) due to the ease of parameter interpretation (Rose et al. 2006). Although logistic regression is the most widely used for modeling tree mortality, other statistical methods have been used for describing and modeling tree mortality, such as the exponential function (Moser et al. 1972), Weibull function (Somers et al. 1980), Richard function (Buford and Hafley 1985), gamma function (Kobe and Coates 1997), lognormal distribution (Presler and Slaughter 1997), and hazards model (Woodall et al. 2005). Vanclay (1995) recommended that logistic regression for tree mortality in tropical forests is appropriate because of its biologically meaningful interpretation, mathematical flexibility, and ease of use.

However, most of the tree mortality models developed generally ignore the heterogeneity that may occur due to repeated measurements nested within a tree and trees nesting within a plot. Multiple sources of heterogeneity occur naturally for most permanent plots used in forestry experiments. Recently, Bayesian multilevel models have been widely used in forestry to account for multiple sources of heterogeneity and to consider prior knowledge of the parameters (e.g., Dietze et al. 2008; Vieilledent et al. 2010; X. Zhang et al. 2015b; Chen et al. 2016).

Chinese fir (Cunninghamia lanceolata (Lamb.) Hook.), a fast-growing evergreen coniferous tree, is one of the most important tree species for timber production in southern China. As a native species, Chinese fir has been widely planted for over 1000 years (X. Zhang et al. 2015b). The objectives of this study were as follows: (1) develop an individual-tree mortality model using a Bayesian logistic multilevel method considering the full multiple sources of variability of data, (2) examine the effects of climate, initial planting density, and competition on tree mortality, and (3) calculate the variance partition coefficient (VPC) for reporting the variation associated with multiple levels in tree mortality models.

### Materials and methods

#### Data

The Chinese fir stands were established in 1981 using bare-root seedlings in Fenyi County (27°30’N, 114°33’E), Jiangxi Province, in southern China, which has a middle-subtropical climate. A total of 15 plots were planted in a random block arrangement with the following tree spacing: 2 m × 3 m (1667 trees/ha), 2 m × 1.5 m (3333 trees/ha), 2 m × 1 m (5000 trees/ha), 1 m × 1.5 m (6667 trees/ha), and 1 m × 1 m (10 000 trees/ha). Each spacing level was replicated three times. Each plot comprised an area of 20 m × 30 m, and a buffer zone (two lines) consisting of similarly treated trees surrounded each plot. Tree diameter measurements in all plots were conducted after the tree height reached 1.3 m. More than 50 trees, including dominant trees in each plot, were tagged and measured for total height. Measurements were performed each winter from 1983 to 1989 and then every other year from 1989 to 1999. Summary statistics are shown in Table 1.

#### Methods

### Candidate variables

#### Endogenous variables

The selection of appropriate independent variables should not only be based on test statistics but also on better understanding of how forest ecosystems function and how factors contributing to death are expressed (Pedersen 2007; Adame et al. 2010). Hamilton (1986) classified the endogenous factors contributing to mortality into three groups: (i) measures of individual tree size, such as diameter at breast height (Dg) or tree height (ii) measures of tree competition, such as stand basal area (Ba), stand density (N), and the ratio of the diameter of the subject tree to the stand quadratic mean diameter (d/Dg), and (iii) the ratio of the height of the subject tree to the stand dominant height (Hd). Moreover, we normalized the initial planting densities (1667, 3333, 5000, 6667, and 10 000 trees/ha) before modeling. In summary, the following endogenous variables were tested for inclusion in the mortality model: IPD, tree size index (ii), tree competition (iii), mortality (A, d−1), tree competition (d/Dg, Ba), and site quality (dominant height, log[Hd]).

#### Climate variables

ClimateAP (T. Wang et al., unpublished data) was used to generate climate data across the study region. ClimateAP is a climate data downscaling tool that produces directly calculated seasonal and annual climate variables and derives climate variables for specific locations (scale-free) based on longitude, latitude, and elevation. The climatic variables tested in this study are shown in Table 2 and are widely used in tree mortality modeling.
(van Mantgem and Stephenson 2007; van Mantgem et al. 2009; Peng et al. 2011; Zhang et al. 2014). In addition, the annual heat–moisture index (AHM) (Wang et al. 2006) was used to indicate the annual climatic water deficit because it integrates the mean annual temperature (MAT) and annual precipitation (MAP) into a single parameter: AHM = (MAT + 10)/[MAP/1000], which better reflects evapotranspiration and soil moisture content than precipitation and temperature alone (Zhang et al. 2014). Large values of AHM indicate dry conditions due to high evaporative demand relative to the available moisture, whereas low values of AHM represent relatively wet conditions.

Independent variables included in the function were selected as variables through forward stepwise regressions. The set of variables is a combination of different groups (tree size, initial planting density, competition, site quality, and climate), avoiding correlations between the groups.

**Bayesian logistic multilevel models**

For a given single tree, survival or death can be represented as a binary that has a value of 1 if the tree survives or 0 if it dies over a given time interval. The most widely used link function for binary data is the logit link function, also named the logistic regression model (Vanclay 1995; Zhang et al. 2014). The traditional logistic mortality model is given by

\[
\ln \left( \frac{p}{1-p} \right) = \alpha_i + \beta x_i
\]

where \( x \) is a vector of explanatory variables, including nonclimatic candidate variables, IPD and climatic variables. \( \alpha_i \) is the intercept, and \( \beta \) is a vector of parameters including the intercept.

**Two-level models**

In this study, we developed two two-level models: measurement occasions (level 1) and trees (level 2) called two-level1 and measurement occasions (level 1) and plots (level 2) called two-level2. Let \( i \) \((i = 1, ..., N)\) denote the level-2 units (trees or plots) and \( j \) \((j = 1, ..., m)\) denote the level-1 units (measurement occasions); therefore, a two-level random intercept logistic model is expressed as

\[
\ln \left( \frac{p}{1-p} \right) = \alpha_i + \beta x_{ij} + \mu_i
\]

where \( \mu_i \) is a level-2 random effect assumed to have a normal distribution with mean 0 and variance \( \sigma^2 \).

**Three-level model**

The simple two-level model (eq. 2) may be adequate for analyzing data from the permanent plots because the trees that are measured repeatedly are also nested within plots. Therefore, a three-level random intercept logistic model is needed:

\[
\ln \left( \frac{p}{1-p} \right) = \alpha_i + \beta x_{ij} + \mu_i + v_{ik}
\]

where \( \mu_i \) and \( v_{ik} \) represent the random effects of the ith tree and kth plot, respectively.

In the Bayesian models, we used “noninformative” priors for all parameters, i.e., a normal distribution with zero mean and a large variance (1000). Such priors typically arise in the form of a parametric distribution with large or infinite variance. Notably, the word “noninformative” prior used here is the classical expression but does not necessarily mean that the prior is truly noninformative (Li et al. 2011). The random effect in the intercept is assumed to follow a normal distribution with mean 0, and the variance of the random effects is given a prior distribution by the inverse gamma (0.001,0.001). The Bayesian method is performed through Markov chain Monte Carlo (MCMC) simulation using the SAS procedure PROC MCMC (SAS Institute Inc. 2011), which uses a random walk Metropolis algorithm to obtain posterior samples (Lindsey 2011). The total number of iterations for tree mortality models was 250 000 with a burn-in of 30 000. Additionally, the thinning parameters were all set to 50 to reduce autocorrelation.

**Bayesian model evaluation**

The Bayesian three-level model was compared with Bayesian fixed-effects (nonhierarchical), two-level1, and two-level2 models using the deviance information criterion (DIC). This is very useful in Bayesian model selection (Spiegelhalter et al. 2002), which is given by

\[
\text{DIC} = \text{Dbar} + pD
\]

where Dbar refers to the posterior mean of the deviance and pD is the effective number of parameters in the model. The posterior mean of the deviance is Dbar = \( E_i(-2\log(p(y|i))) \) and pD = Dbar – Dhat. Dhat is a point estimate of deviance given by \( \hat{D} = -2\log(p(y|\hat{\theta})) \). The advantage of DIC over other criteria in Bayesian model selection is that DIC is easily calculated from samples generated by a MCMC simulation. Models with lower DIC values indicate a better fit to the data in which differences ≥5 are regarded as substantial evidence and differences ≥10 are regarded as very strong evidence in favor of the model with the lowest DIC (Hurst et al. 2011).

The value of the area under the receiver operating curve (AUC) is widely used for evaluating tree mortality models (Saveland and Neuenschwander 1990; Zhang et al. 2011) and was also used in this study. The larger the value of the AUC, the better the model performs (Fielding and Bell 1997). A rough guideline for the AUC follows the traditional academic point system as follows: 0.9–1 = excellent (A); 0.8–0.9 = good (B); 0.7–0.8 = fair (C); 0.6–0.7 = poor (D), and 0.5–0.6 = fail (F) (Zhang et al. 2011).

**Table 2. Summary statistics of climate variables.**

<table>
<thead>
<tr>
<th>Climate variable</th>
<th>Description</th>
<th>Mean</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAT (°C)</td>
<td>Mean annual temperature</td>
<td>18.18</td>
<td>17.3</td>
<td>19.2</td>
</tr>
<tr>
<td>MAP (mm)</td>
<td>Mean annual precipitation</td>
<td>1697.29</td>
<td>1334</td>
<td>1975</td>
</tr>
<tr>
<td>MCMT (°C)</td>
<td>Mean coldest month temperature</td>
<td>6.96</td>
<td>4.1</td>
<td>9.1</td>
</tr>
<tr>
<td>AHM</td>
<td>Annual heat–moisture index</td>
<td>16.82</td>
<td>14.4</td>
<td>21.1</td>
</tr>
<tr>
<td>DD_0</td>
<td>Degree-days below 0 °C, chilling degree-days</td>
<td>3</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>NFFD</td>
<td>Number of frost-free days</td>
<td>349</td>
<td>334</td>
<td>355</td>
</tr>
<tr>
<td>SMMT (°C)</td>
<td>Summer mean maximum temperature</td>
<td>32.02</td>
<td>30.7</td>
<td>33.3</td>
</tr>
<tr>
<td>WMMT (°C)</td>
<td>Winter mean minimum temperature</td>
<td>4.28</td>
<td>2.3</td>
<td>6</td>
</tr>
<tr>
<td>SMT (°C)</td>
<td>Summer (June–August) mean temperature</td>
<td>27.27</td>
<td>26.4</td>
<td>28</td>
</tr>
</tbody>
</table>
VPC in multilevel logistic models

Multilevel models, also known as hierarchical models, assume multiple sources of unobserved heterogeneity and recognize units at one level as grouped (nested) in the next higher level. In multilevel models, the residual variance is split up into components that are attributed to the various levels in the data (Goldstein 1995). Partitioning the variance in binary models is complex because the level-1 variance (Bernoulli) is measured on a different scale compared to the level-2 variance. Rodriguez and Goldman (2001) said that if the relationship between the dependent and independent variables is nonlinear (e.g., logit link function), then ignoring grouping (nested) structures can result in large biases in estimating parameters. Additionally, note that for the logistic multilevel model, the level-1 variance is not identifiable from the likelihood. The classically reported fixed variance pertains to the latent continuous scale and is the variance of \( \logit^{-1}(p) \) for a given \( p \). For a given \( p \), a tree with a larger \( D_g \) will have a lower mortality probability (Table 3). The positive estimated coefficient for dominant height in the mortality model indicates that mortality is higher on better sites.

Effects of initial planting density and competition on tree mortality

In the Bayesian multilevel mortality model, the IPD was significantly positive with tree morality (Table 3). The mortality probability of a tree is influenced by its relative position (competitive status) within the stand, which was calculated the competition index using the ratio of \( d \) to \( D_g \). For a given \( D_g \), a tree with a larger diameter has lower mortality probability (Table 3). The positive estimated coefficient for dominant height in the mortality model indicates that mortality is higher on better sites.

Climate effects on tree mortality

A summary of the parameter estimates obtained in the climate sensitive mortality models is provided in Table 3. Four climatic variables, including MAT, AHM, WMMT, and SMT, are significantly correlated with tree mortality. In the mortality models, MAT had a negative effect on the mortality of Chinese fir. A negative significant correlation with SMT was also found. By contrast, the tree mortality probability had a positive significant correlation with WMMT. Additionally, AHM was significantly positively correlated with tree mortality.
correlated with mortality probabilities (Table 3), which suggests that climate change induced drought increases tree mortality. Although the climate effects, including MAT, AHM, WMMT, and SMT, on mortality were significant (Table 3), the effects were relatively small (Fig. 1).

Discussion

Bayesian multilevel methods in tree mortality

A Bayesian multilevel model is an alternative method that accounts for variation at multiple clustering levels of data, such as the tree or plot level in the tree mortality analysis. An advantage of Bayesian methods over classical methods when fitting models is that independent, prior information, if available, can be incorporated into the model through prior distributions of unknown parameters (Masuda and Stone 2015). In this paper, we do not intend to argue that the Bayesian multilevel model is superior to the classical mixed-effects model. We believe that both classical and Bayesian methods have their own features. In particular, when the Bayesian prior is uninformative, the results would be very close to those obtained with classical mixed-effect models (Pinheiro and Bates 2000; McCarthy 2007).

Bayesian logistic models with fixed effects only and with multilevel random effects added to the intercept were fitted, including both endogenous and exogenous variables. One two-level model (two-level1) was fitted to account for the correlation in measurement occasions (level 1) from a tree (level 2). The other two-level model (two-level2) was also fitted to account for the correlation in measurement occasions (level 1) of a tree in a plot (level 2) as well as the three-level logistic model accounting for the correlation in repeated measurements (level 1) from trees (level 2) nested in a plot (level 3). The multilevel models were much better than the fixed-effects only model according to DIC, and the three-level model also performed best compared to the two-level models (Table 3). Ma et al. (2013) reported that the multilevel logistic model incorporating both the plot and measurement random effects performed the best compared to standard logistic and marginal logistic models based on the generalized estimating equations. Thapa (2014) modeled loblolly pine (Pinus taeda L.) mortality using multilevel mixed-effects logistic regression and found that the model accounting for measurement, tree, and plot three levels compared to the two levels of measurement and plot was consistent with our study. Groom et al. (2012) observed that inclusion of a random intercept in multilevel tree mortality models for Douglas-fir (Pseudotsuga menziesii (Mirb.) Franco) stands significantly reduced model bias compared to the fixed-effects model. The VPC attributed to tree level was 74.58% larger than that due to plot level. This may be because a large part of the variance is due to the direct environment of the tree, which cannot be fully described with the symmetric competition index. Some researches reported that the variation of seeds used for seedlings also resulted in different mortality probabilities (Bonfil 1998; Khan and Shankar 2001; Westoby et al. 2002).

Endogenous factors affecting tree mortality

A major mortality agent is intertree competition, which can be either symmetric or asymmetric (Yang et al. 2003; Adame et al. 2010). In symmetric competition, larger trees in a plot have competitive advantages over smaller trees and neighbors do not affect the growth (Cannel et al. 1984). Hamilton (1986) reported that the variable d/Dg was a good symmetric competition index (Avila and Burkhart 1992; Zhang et al. 1997). A positive effect of d/Dg on tree survival is shown in Table 3, indicating that tree death tended to occur in strong symmetric competition stands, which was consistent with the result obtained by Laarmann et al. (2009). For long-term competition, if a given tree size was small compared with a mean tree size, the tree must have strong competition in the stand. In an undisturbed stand, intense competition for nutrients, water, and light particularly affects small trees and induces a higher mortality probability (Coomes and Allen 2007). In asymmetric competition, all trees in a plot impose some competition on their neighboring trees, regardless of their size (Cannel et al. 1984), and this can be described by stand density and basal area. Yang et al. (2003) found that the stand basal area is a good measure of stand crowding and could adequately capture asymmetric competition because it combines both tree size and density. However, here, the two variables were excluded in the mortality model according to the stepwise analysis.

Stand density management, using different initial planting densities, was a key factor driving the mortality process. For a given age and site index, competition-induced mortality was higher at higher planting density (Williams 1994; Zhao et al. 2007), which was consistent with the result of this case (Table 3). The increase in crown ratio and crown length led to a reduction in the relative risk of death among the surviving trees (Thapa 2014). The live crown length and crown ratio decreased with increasing
planting density due to competition from neighbors (McClain et al. 1994; Akers et al. 2013).

Site index has frequently been introduced into mortality models. For a given age and initial planting density, the mortality probability of Chinese fir increased with the site index (Table 3). Our finding is consistent with the conclusion of Eid and Tuhus (2001), Yao et al. (2001), and X. Zhang et al. (2015) that higher mortality was related to better productivity. Empirical evidence suggests that density-dependent mortality in plantations becomes apparent earlier in better sites and increases with site productivity (Diéguez-Aranda et al. 2005). However, opposite results were obtained in some studies. Woolions (1998) reported that mortality decreased with increasing site index. Jutras et al. (2003) found that site index had a significant effect on mortality, with a higher mortality probability for Scots pine (Pinus sylvestris L.) but lower mortality probability for pubescent birch (Betula pubescens Ehrh.). Zhao et al. (2007) also found that site productivity affects mortality in an opposite way in the Piedmont/Upper Coastal Plain and Lower Coastal Plain of the southern United States as follows: mortality increases with increasing productivity in the Piedmont/Upper Coastal Plain, but in the Lower Coastal Plain, higher mortality is related to lower productivity. The different effects of site index on mortality may be related to the availability of certain key nutrients (Jutras et al. 2003).

Climate variables affecting tree mortality

In addition to endogenous variables, climate variables were also directly related to tree mortality and may be helpful for exploring the effects of climate change on forest dynamics under future climate scenarios. However, the effects of these four climate variables were relatively small (Fig. 1). This may be because the forest stand structure and age mediated the climate effects on tree mortality (Ruíz-Benito et al. 2013; Bell et al. 2014). In general, an increasing mean annual temperature decreases tree mortality probabilities physiologically, when water availability is not limiting. Temperature exerts effects on tree growth and mortality by altering rates of photosynthesis, respiration, cell division and elongation, chlorophyll synthesis, enzymatic activity, water uptake, and transpiration (Ricker et al. 2007). Additionally, an increasing mean annual temperature prolongs the nonfrost growing season during the entire year and increases the CO2 sequestration rate (Bergh et al. 2003). The tree mortality of Chinese fir decreased with increasing MAT (Table 3; Fig. 1). However, our finding was different from other studies showing that temperature was positively related to tree mortality (van Mantgem et al. 2009; Ruiz-Benito et al. 2013; Zhang et al. 2014). Chinese fir is a shade-intolerant tree species, the growth of which tends to increase with high temperature, thus decreasing the mortality probability (Wu 1984). In addition, a warm late summer can extend the growing season, increasing radial growth and consequently affecting tree mortality in the following year, which is supported by the negative relationship between the late-summer mean temperature and tree mortality probabilities observed in this study.

For the AHM, higher mortality was related to a higher AHM (Table 3). This result confirmed the findings of Breshears et al. (2005), Peng et al. (2011), Zhou et al. (2013), and Zhang et al. (2014), who showed that climate change induced drought stress increased tree mortality. Zhang et al. (2014) reported that drought-induced mortality may be primarily caused by three factors: (i) carbon starvation, which halts most photosynthesis, thus failing to support metabolism and carbon balance, (ii) hydraulic failure, which increases water deficits and drought stress on trees, and (iii) outbreaks of biotic agents, such as the growth and reproduction of insects and pathogens that attack trees (van Mantgem et al. 2009). Generally, climate change induced drought was an inciting tree mortality factor (Bigler et al. 2006; Allen et al. 2010).

Conclusions

This study investigated tree mortality models with the inclusion of initial planting density, competition, and climate variables in Chinese fir plantations in southern China using Bayesian logistic multilevel methods. The Bayesian three-level model was the best to describe tree mortality data with multiple sources of unobserved heterogeneity. The VPC attributed to the tree level was much larger than that due to the plot level.

The endogenous factors of IPD, d/Dg, and log(Hd) were significantly related to mortality. Tree mortality increased with increasing IPD and Hd and with a decreasing symmetric competition index d/Dg. Four climate variables were also related to tree mortality. In the mortality model, MAT and SMT had negative effects on the mortality of Chinese fir. By contrast, the tree mortality probability was positively significantly correlated with WMMT and AHM.

The effect of the initial planting density has implications for forest management because high planting density results in high mortality. Inclusion of climate variables in mortality models can facilitate the projection of tree mortality under future climate change conditions. The Bayesian multilevel model accounting for the uncertainty in parameter estimates shown by the posterior distribution and the heterogeneity that may occur due to the measurement occasions nesting within a tree (i.e., repeated measurements) and trees nesting within a plot was a good method for modeling the tree mortality of Chinese fir plantations.

Acknowledgements

The authors express their appreciation to the Fundamental Research Funds for the Central Non-profit Research Institute of CAF (CAFYBB2017ZX001-2), the National Natural Science Foundation of China (No. 31670634), and the Scientific and Technological Task in China (No. 2016YFD060302-1).

References


