NATURAL RESOURCE MODELING
Volume 26, Number 2, May 2013

# MODELING INDIVIDUAL TREE MORTALITY RATES USING MARGINAL AND RANDOM EFFECTS REGRESSION MODELS

ZHIHAI MA
Département des Sciences Biologiques, Institut des Sciences de l'environnement, Université du Québec à Montréal (UQAM), Montréal, H3C3P8, Canada

CHANGHUI PENG*
Département des Sciences Biologiques, Institut des Sciences de l'environnement, Université du Québec à Montréal (UQAM), Montréal, H3C3P8, Canada; Laboratory for Ecological Forecasting and Global Change, College of Forestry, Northwest A & F University, Yangling, Shaanxi, 712100, P. R. China

WEIZHONG LI
Laboratory for Ecological Forecasting and Global Change, College of Forestry, Northwest A & F University, Yangling, Shaanxi, 712100, P. R. China

QIUAN ZHU
Laboratory for Ecological Forecasting and Global Change, College of Forestry, Northwest A & F University, Yangling, Shaanxi 712100, China

WEIFENG WANG
Département des Sciences Biologiques, Institut des Sciences de l'environnement, Université du Québec à Montréal (UQAM), Montréal, H3C3P8, Canada

XINZHANG SONG
Département des Sciences Biologiques, Institut des Sciences de l'environnement, Université du Québec à Montréal (UQAM), Montréal, H3C3P8, Canada; Zhejiang Provincial Key Laboratory of Carbon Cycling and Carbon Sequestration in Forest Ecosystems, Zhejiang Agriculture and Forestry University, Lin'an, 311300, China

JIANWEI LIU
Forestry Branch, Manitoba Conservation, Box 70, 200 Saulteaux Cres, Winnipeg, MB R3J 3W3, Canada

ABSTRACT. Developing models to predict tree mortality using data from long-term repeated measurement data sets can be difficult and challenging due to the nature of mortality as well as the effects of dependence on observations. Marginal (population-averaged) generalized estimating equations (GEE) and random effects (subject-specific) models offer two possible ways to overcome these effects. For this study, standard logistic, marginal logistic based on the GEE approach, and random logistic regression models were fitted and compared. In addition, four model evaluation statistics were calculated by means of $K$-fold cross-valuation. They include the mean prediction error, the mean absolute prediction error, the variance of prediction

*Corresponding author. Dr. Changhui Peng, Department of Biology Sciences, Institute of Environment Sciences, University of Quebec at Montreal, C.P. 8888, Succ. Centre-Ville, Montreal H3C 3P8, Canada, *E-mail*: peng.changhui@uqam.ca
Received by the editors on 5th August 2011. Accepted 12th March 2012.

error, and the mean square error. Results from this study suggest that the random effects model produced the smallest evaluation statistics among the three models. Although marginal logistic regression accommodated for correlations between observations, it did not provide noticeable improvements of model performance compared to the standard logistic regression model that assumed impendence. This study indicates that the random effects model was able to increase the overall accuracy of mortality modeling. Moreover, it was able to ascertain correlation derived from the hierarchal data structure as well as serial correlation generated through repeated measurements.

KEY WORDS: Generalized estimating equations, random effects model, logistic regression, longitudinal data, tree mortality.

**1. Introduction.** The accurate modeling of tree mortality is critical in the development of effective forest management policy. Forest management decisions should be based on scientifically sound information related to the future status of forests, typically predicted by way of forest growth and yield models. One of the most important growth and yield models in use is the mortality model. It is used to estimate tree number reduction due to competition, droughts, disturbances, environmental factors, etc., based on information taken from the current state of individual stands. Traditionally, tree mortality was generally divided into regular mortality owing to suppression and competition and irregular mortality (catastrophic mortality) that is typically caused by fire, wind damage, or natural disturbances such as insect and disease, etc. (Lee [1971]). These two types of mortality may, in fact, not always occur independently. For example, suppression may weaken a tree's defensive capacity that in turn enhances the chance of attacks from disease or insects. Tree mortality is very difficult to predict. It is a complex, gradual, and a highly variable process that results from the interaction of multiple observable or unobservable factors such as insect, disease, competition, climate change, etc. (Lee [1971], Fortin et al. [2008], van Mantgem et al. [2009]). Furthermore, intrinsically non-Gaussian binary responses bounded by 0 and 1 make mortality modeling even more complex because standard ordinary least squares regression is no longer appropriate for use as a means of statistical analysis for discrete, bounded, and binomial-distributed mortality data (Allison, 1999). Multiple statistical methods based on different cumulative distribution functions such as Weibull (Somers et al., 1980) and gamma (Kobe and Coates, 1997) distribution have been developed and used to predict individual tree mortality. The most popular method at the current time is logistic regression model (e.g., Eid and Tuhus, 2001; Zhao et al., 2006).

Although logistic regression is the most common approach used to determine mortality in modeling, it is based on the assumption of independent observations. This is unlikely a reasonable assumption since tree measurements are typically repeated through time (longitudinal) and clustered in plots. This type of data is typically referred to as "clustered longitudinal data." Parameter estimations from standard logistic regressions remain unbiased even without considering the existence of dependence among observations. However, estimates are no longer efficient, and the

estimated standard errors of regression parameters are biased, which may produce misleading statistical tests and thus make the selection of appropriate covariates difficult (Allison, 1999). Two possible methods exist to correct biases induced by dependence and thus obtain accurate standard errors and test statistics. One method is to apply marginal generalized estimating equations (GEE) proposed by Liang and Zeger (1986). A GEE model includes a working correlation structure to accommodate correlation between repeated measurements within subject. This working correlation structure is assumed to be the same for all subjects and reflects the average dependence among repeated measurements for all subjects. For the GEE model, several working correlation structure choices are available. Recently, Kiernan et al. (2009) applied GEE in conjunction with an exchangeable working correlation structure to model tree mortality. Their results show that the GEE approach better predicts tree mortality than does a standard logistic model. Another useful alternative method is to use the random effects model (also referred to as multilevel models, and generalized linear mixed models). This model extends the usage of generalized linear models by including random coefficients to take into account correlation derived from the hierarchical structure of the data during the model fitting process (Calama and Montero, 2005). Data used for modeling tree mortality are hierarchical because they are measured repeatedly through time on trees growing in different plots. For repeated measurements, the within-tree observations are likely to be serial correlation. Moreover, trees within the same plot are also likely correlated in relation to competition and microsite variability such as soil type, topography, geology, and microclimatic factors (Fox et al., 2007). Mailly et al. (2009) recently applied a generalized linear mixed model containing random effects to compare mortality rates between top height and average site trees to correct correlations that may arise from the hierarchical structure of the data. In addition, Jutras et al. (2003) applied multilevel logistic regression models to study mortality of individual trees in drained peatlands in Finland. Finally, Rose et al. (2006) applied a multilevel model to predict individual tree survival rates.

Although both random effects and marginal GEE approaches have been used for modeling clustered longitudinal binary tree mortality data in forest stands, they address the same problems differently in several fundamental ways such as in their assumptions and interpretation of parameter estimates. Comparisons of the two approaches have appeared in statistical literature (Pendergast et al., 1996), biomedical research (Carrière and Bouyer, 2002; Twisk, 2004), and ecological journals (Fieberg et al., 2009; Koper and Manseau, 2009), but to date little research has been published on the quantification of the differences between marginal and random effects model approaches in forest literature. Data used in forestry typically contain certain distinct features separate from other fields. First, data typically contain a large number of subjects/trees (e.g., 139 nests were used by Fieberg et al. (2009) and 18 woodland caribou were used by Koper and Manseau (2009) to compare the two approaches). Second, the time and number of repeated measurements are highly irregular (from one to up to 16 years for this study), and the number of repeated

TABLE 1. Summary statistics of stand variables and a description of data measurements for the 41 plots located in KC.

| Stand variables | Mean | Standard deviation | Minimum | Maximum | Measurement periods (year) |
|---|---|---|---|---|---|
| DBH (cm) | 12.91 | 4.58 | 1.30 | 34.30 | 1,4,5,6,8,9, |
| Trees per ha | 3432 | 1143 | 1069 | 7293 | |
| Plot basal area ($m^2$/ha) | 62.07 | 10.45 | 33.34 | 86.38 | 10,11,12,13,16 |
| BAL ($m^2$/ha) | 37.02 | 8.42 | 0.00 | 85.28 | |

measurement intervals is relatively short (from two to up to seven measurement intervals for this study). Third, trees clustered within the same plot tend to be spatially correlated. It is therefore essential to both compare performance and clarify differences between these two approaches when modeling tree mortality. The objectives of this study were: (1) to describe and compare the primary features of marginal and random effects logistic modeling; (2) to develop individual tree mortality models by applying standard, random effects, and marginal logistic models while identifying differences in model estimations; and (3) to compare overall model performance.

**2. Data and study area.** Data used in this study were obtained from the permanent plots established by Kimberly Clark Limited (KC) (Zhang et al., 2004). All plots used in this study were located within the Longlac-Geraldton area of Ontario. Several criteria were applied when selecting the plots: (1) A plot must have at least three consecutive measurements; (2) a plot must be dominated by black spruce or one in which at least 50% of total basal area is black spruce; and (3) any managed plot such as those thinned or damaged by insect invasion, wind, snow, etc., would be rejected for this study. After selection, a total of 41 plots were chosen from the KC plots. These plots were generated naturally between the years 1790 and 1923. The size of all KC plots was 0.058 ha. A diameter at breast height measurement of 1.30 m (DBH) was measured at each census. Jack pine (*Pinus banksiana*) (31.3% in total number) and black spruce (*Picea mariana*) (48.6% in total number) were the two dominant tree species of the selected plots. Other species included white spruce (*Picea glauca*), balsam fir (*Abies balsamifera*), balsam popular (*Populus balsamifera*), and trembling aspen (*Populus tremuloides)*. A total number of 7987 black spruce trees comprising of 30,825 observations taken from the 41 fixed-area plots were measured between three and seven times for a period from 1 to 16 years (Table 1).

**3. Methods.** Response variable values were coded as 1 and 0, representing dead and living trees, respectively. Three regression models that include standard logistic (LR), marginal logistic (MLR), and random logistic regression (RLR) models were used to estimate the probability of individual tree mortality. Instead of predicting the annual probability of survival (e.g., Yao et al., 2001) for trees that go through irregular measurement periods, tree mortality probability in this study was directly modeled within a specific period by including the length of census interval as one of the independent dummy variables (Kiernan et al., 2009). The interaction terms between dummy variable of time and other independent variables were also explored (Kiernan et al., 2009). In addition, the initial tree DBH (cm), stand basal area (BA) ($m^2/ha$), basal area of the larger trees compared to subject trees (BAL), and the number of live trees per hectare (Num Trees) were also included in each model. The same set of independent variables was used for each model. The following section briefly discusses the three logistic models introduced.

**3.1. Standard logistic regression (LR) model.** The form of a logistic regression model can be expressed as:

$$logit\, E(y_{ijt}) = g(\mu_{ijt}) = \log\left(\frac{\mu_{ijt}}{1-\mu_{ijt}}\right) = X_{ijt}\beta, \tag{1}$$

where $\mu_{ijt} = \mathrm{E}(y_{ijt})$ is the expected mortality probability of the $t$th measurement of the $i$th tree in the $j$th plot; $logit(.) = g(.)$ is the link function that is used to transform the expected mortality probability of the response variable to a linear form; $X_{ijt}$ is the predictor variable vector; and $\beta$ is the unknown regression parameter vector to be estimated. In equation (1), the ratio of $\frac{\mu_{ijt}}{1-\mu_{ijt}}$ is typically referred to as the "odds." The *logit* transformation is unbounded because the transformation of the probability to the odds removes the upper bound, and the natural logarithm of the odds removes the lower bounds. The $\mu_{ijt}$ can be obtained as following:

$$\mu_{ijt} = \frac{1}{1+\exp(-X_{ijt}\beta)}, \tag{2}$$

where exp(.) is the exponential function. Note that $\mu_{ijt}$ will always be a number between 0 and 1. Generally, the maximum likelihood method will be used to estimate logistic regression models. The estimators of the maximum likelihood approach are consistent, asymptotically efficient, and asymptotically normal if the data meet the independent assumption (Allison, 1999). Once the estimated coefficients are obtained, equation (2) can be used to predict the mortality probability for a given tree over a specific period. Note that in this study $\mu_{ijt}$ is the probability of death for a given tree.

**3.2. Marginal logistic regression (MLR) model.** The GEE approach was used to build a marginal logistic regression model. The term marginal implies that the mean response of the model depends only on the independent variables and does not include any random effect. Thus, the marginal model is also referred to a population-averaged model since the relationship between the response variable and the independent variables is the same for all subjects. It allows correlations within subjects to be estimated while using these estimated correlations to estimate the regression coefficients as well as their standard errors. The marginal models are semiparametric since estimates do not require distributional assumptions to be associated to the observations (Fitzmaurice et al., 2004).

A marginal model ascribed the expected values of response $\mu_{ijt} = E(y_{ijt})$ to a set of independent variables by means of a link function

$$Logit(E(y_{ijt})) = g(\mu_{ijt}) = X_{ijt}\beta, \tag{3}$$

where $g(.)$ is the specified link function; $y_{ijt}$ is the response for the $t$th measurement of the $i$th tree in the $j$th plot; and $\beta$ is the vector of unknown regression coefficients. The variance for each $y_{ijt}$ is assumed to be a function of the mean

$$Var(y_{ijt}) = \phi V(\mu_{ijt}), \tag{4}$$

where $\phi$ is a scale parameter, and $V(\cdot)$ is a known variance function of the mean $\mu_{ijt}$. The *logit* link function was used in this study for binary response and the corresponding $V(\mu_{ijt}) = \mu_{ijt}\ (1 \pm \mu_{ijt})$.

Dependence between repeated responses within subjects was incorporated into the GEE modeling process by a $n \times n$ working correlation matrix $R(\alpha)$ where n is the maximum number of the repeated measurements between subjects, and $\alpha$ is a vector of unknown correlation parameters. The working correlation matrix in the GEE modeling process is assumed to be the same for all subjects and is estimated to reflect the average dependence between the repeated observations over the subjects. Several working correlation structures are available: independent, autoregressive $AR(1)$, exchangeable, and unstructured. The independent working correlation structure assumes no correlation between repeated measurements, and in this particular case the GEE estimation is equivalent to a standard logistic regression model. The $AR(1)$ working correlation structure assumes that the observations are only correlated to their own previous values by way of the first order autoregressive process. The $AR(1)$ is a good choice if measurements are taken repeatedly over time, but correlation decays quickly as intervals of measurements increase (Davis, 2002). The exchangeable working correlation structure assumes that correlations are equal across all time points. The unstructured working correlation structure is completely unspecified and is useful when few observational time points exist (Allison, 1999). The number of parameters to be estimated in vector $\alpha$ may vary depending on

the type of working correlation matrix used (e.g., only one unknown coefficient is present for the exchangeable structure while there are present n(n−1)/2 parameters for the unstructured coefficient to be estimated). A corresponding working covariance can be obtained for a given working correlation structure $R(\alpha)$:

$$V_i = D(V(\mu_{it}))^{1/2} R(\alpha) D(V(\mu_{it}))^{1/2}, \tag{5}$$

where $D$ is a diagonal matrix, and $V(\mu_{it})$ is the variance of the marginal mean $\mu_{it}$. The GEE estimator of $\beta$ is therefore obtained by solving the following GEE:

$$\sum_{i=1}^{n} D_i' V_i^{-1} (y_i - \mu_i) = 0, \tag{6}$$

where $V_i$ is the working covariance matrix, and $D' = \partial \mu_i / \partial \beta$.

The following iterative process is used to estimate model parameters. First, a generalized linear model that assumes independence is fitted to obtain estimates of model parameters and residuals, after which the parameters of the working correlation matrix are all estimated from the residuals. Second, the regression models are refit using the estimates of model parameters and the working correlation from the first set. The above two steps continue until the convergence criteria is attained (Fitzmaurice et al., 2004).

Two types of standard error estimates exist for regression coefficients: the model-based standard error estimate and the robust or "sandwich" standard error estimate. Model-based standard error is only consistent when the specified correlation structure is correct while robust standard error is a consistent estimator even if the correlation structure is not correctly specified or the strength of the association between measurements varies from subject to subject (Liang and Zeger, 1986). Robust standard error estimators are therefore preferred. This is especially true when a large number of clusters must be managed to calculate the confidence intervals as well as carry out hypothesis tests.

**3.3. Random logistic regression (RLR) model.** In contrast to the marginal model approach, the random model approach can be used to estimate subject-specific effects. The random effects model allows the relationship between variables to vary between subjects by including random effects. It may be written as:

$$logit(E(y_{ijt}|\gamma)) = g(\mu_{ijt}|\gamma) = X_{ijt}\beta + Z_{ijt}\gamma, \tag{7}$$

where $\gamma$ is the random effects parameters assumed to be normally distributed with a mean of 0 and a variance matrix of $G$. The random effects parameters are assumed to vary from subject to subject. In other words, random effects parameters reflect the localized (e.g., localized plots or localized per individual tree) relationships between variables. The variance of the random effects parameters exhibits

the unexplained variation between subjects. It must be estimated during the model fitting process. Because a tree is nested in a plot and a plot is nested in a forest stand, both the plot and tree are considered random effects that explain variation between plots and trees. Convergence problems surfaced when this study took into account two random effects at the same time. Due to this, the tree level random effect was finally dropped from the random effects model. Only a random intercept is included in equation (7) to represent the simplest case:

$$logit(E(y_{ijt}|\gamma)) = g(\mu_{ijt}|\gamma) = X_{ijt}\beta + \gamma_{0j} \tag{8}$$

where $\gamma_{0j}$ is the random intercept for plot $j$. In this case each tree has the same set of independent variable slopes but may retain different intercepts, depending on the specific plot the tree originated from. The random effects of plots were used to capture the response dependence that arises from unobserved or unobservable characteristics of the plots.

There are several ways to estimate RLR parameters: penalized quasi-likelihood (PQL), Laplace approximations, and Gauss-Hermite quadrature (GHQ) (Bolker et al., 2009). The PQL approach is the most commonly used method. However, this approach will produce biased parameter estimates when standard deviations of the random effects are large (Bolker et al., 2009). Moreover, PQL only calculates the quasi-likelihood rather than a "true" likelihood. Due to this, a likelihood-based model selection criteria such as AIC (Akaike information criterion) can not be used to compare model performance. Laplace approximations and GHQ methods produce true likelihoods rather than quasi-likelihoods. Since Gauss-Hermite quadrature is more accurate (Bolker et al., 2009), the GHQ method was used for this study.

**4. Model fitting.** Both logistic and marginal logistic regression models were fitted using the GENMOD procedure in SAS (SAS Institute, Inc., 2008). All observations were treated as independent for the logistic regression model. The likelihood estimation method and the Logit link function were selected to estimate regression coefficients. The robust standard errors based on the sandwich estimator were used to evaluate the statistical significance of the regression coefficients. The exchangeable correlation structure was used for the MLR model due to the uneven measurement periods between plots (Allison, 1999; Kiernan et al., 2009). The QIC (**Q**uasi-likelihood under the **I**ndependence model **C**riterion) statistic proposed by Pan (2001) was also calculated for the purpose of choosing the correct correlation structure. Results indicate that the MLR model in conjunction with an exchangeable correlation structure produced the smallest QIC. Due to this, the exchangeable correlation structure was finally selected for this study. The model with the smallest statistic is preferred when using QIC to compare two models.

The RLR model was fitted using the GLIMMIX procedure in SAS (SAS Institute, Inc., 2008). Plot was treated as random effects in the RLR model. Likelihood ratio

tests were carried out to test if the incorporation of random effects in RLR provided statistically significant improvements with regards to model fittings. The $p$-value of the likelihood ratio tests for models either without or with plot random effects was smaller than 0.0001. The plot random effects were therefore retained in the RLR model.

**5. Model evaluation.** It is not currently possible to directly compare the relative fitting of the three models under investigation. This is true for the following reasons: The standard LR model and the RLR model are based on the maximum likelihood framework for model estimation while MLR is based on the quasi-likelihood theory, and no assumption can be made concerning the distribution of response observations. Although the Akaike information criterion (Burnham and Anderson, 1998), a popular method used for model selection, could be applied to evaluate the goodness of fit between the LR and RLR models, it is not applicable to GEE. Similarly, QIC may be applied to evaluate the goodness of fit of the MLR model, but it is not applicable for the LR and RLR models. No common criterion therefore exists that can be applied to all models. Moreover, the direct comparison of parameter estimates is also unsuitable since parameter estimates for the MLR model are marginal while parameter estimates for the RLR model are conditional.

Instead, the $K$-fold cross-validation method (Efron and Tibshirani, 1993; Koper and Manseau, 2009) was used to compare model performance in its capacity to predict individual tree mortality. $K$ was set to 10 for this study. The 10 is selected based on empirical usage of $K$-fold cross-validation in previous study of McLachlan et al. (2004). For the simplicity, we only select one number for $K$ and the other numbers such as 5, 15 can also be selected. The original data set was randomly partitioned into 10 subsamples. Rather than using 10% of the total observations, approximately 10% of the total numbers of individual trees were selected for each subsample. One subsample was chosen and served as validation data for each cross-validation. The nine remaining subsamples were used as calibration data. This cross-validation process was repeated 10 times. During this process all observations were used for both validation and calibration while each observation was used for the purpose of validation one time only. The model residual was defined as the observed mortality minus the predicted mortality. It was obtained for each observation. The predicted mortality based on the fitted LR and MLR was obtained by $\hat{p}_{ijt} = \frac{1}{1+\exp(-X_{ijt}\hat{\beta})}$ where $X_{ijt}$ is the observed vector of independent variables for the $t$th measurement of the $i$th tree in the $j$th plot, and $\hat{\beta}$ is the vector of estimated fixed effects coefficients. Since the estimated coefficients of RLR may vary between plots, the predicted mortality based on the fitted RLR model was calculated as $\hat{p}_{ijt} = \frac{1}{1+\exp(-(X_{ijt}\hat{\beta}+\hat{\gamma}_{0j}))}$ where $\hat{\beta}$ is the estimated fixed effect coefficients, and $\hat{\gamma}_{0j}$ is the estimated random intercept of plot $j$. Therefore, all trees possessed the same set of estimated fixed regression coefficients while trees between different plots may possess different sets of estimated regression coefficients. Once the residuals

were obtained for the three models at the end of each cross-validation, the four evaluation statistics (Zhang, 1997; Kiernan et al., 2009) were calculated as follows:

1. The mean prediction error ($\bar{e}$) was computed as:

$$\bar{e} = \frac{\sum (p_{ijt} - \hat{p}_{ijt})}{n} \tag{9}$$

where $p_{ijt}$ and $\hat{p}_{ijt}$ are the observed and predicted mortality for the $t$th measurement of the $i$th tree in the $j$th plot, and $n$ is the number of observations in the validation data set. The mean prediction error was used to evaluate the prediction error of each model.

2. The mean absolute prediction error ($|\bar{e}|$) was obtained by:

$$|\bar{e}| = \frac{\sum |p_{ijt} - \hat{p}_{ijt}|}{n} \tag{10}$$

The mean absolute prediction error was used to show the magnitude of the model's prediction error.

3. The standard deviation of the prediction error ($\sqrt{v}$) was used to evaluate model precision. It is defined as:

$$\sqrt{v} = \sqrt{\frac{\sum (e_{ijt} - \bar{e}_{ijt})^2}{n-1}} \tag{11}$$

where $e_{ijt} = (p_{ijt} - \hat{p}_{ijt})$ is the prediction error for the $t$th measurement of the $i$th tree in the $j$th plot.

4. The mean square error ($MS$) for which both prediction bias and precision were considered was obtained by:

$$MS = \bar{e}^2 + v \tag{12}$$

The average of the above four statistics was calculated after 10 cross-validation treatments. To compare model performance across tree size, trees in the validation set were grouped into 3 cm diameter classes. The mean of the prediction error, the mean of the absolute prediction error, and the standard deviation of the prediction error were then computed for each 3 cm diameter class at the point of each cross-validation. The last step was to obtain the average mean of the prediction error, the mean absolute prediction error, and the standard deviation of the prediction error across all diameter classes.

TABLE 2. Parameter estimates and related statistics of the LR model.

| Parameter | Estimate | SE | Wald $\chi^2$ | $P$ |
|---|---|---|---|---|
| Intercept | 18.5897 | 8.0931 | 5.22 | 0.0223 |
| DBH | −0.7358 | 0.3510 | 4.40 | 0.0360 |
| BA | 0.0052 | 0.0039 | 1.75 | 0.1860 |
| BAL | −3.2637 | 1.6359 | 3.98 | 0.0460 |
| Num trees | −0.0056 | 0.0004 | 188.82 | <0.0001 |
| Time 5 | −18.0958 | 8.0960 | 5.00 | 0.0254 |
| Time 6 | −19.1554 | 8.3088 | 5.32 | 0.0211 |
| Time 8 | −19.4374 | 8.2358 | 5.57 | 0.0183 |
| Time 11 | −20.3700 | 8.2379 | 6.11 | 0.0134 |
| DBH*Time 11 | 0.7814 | 0.3559 | 4.82 | 0.0281 |
| BAL*Time 5 | 3.4867 | 1.6372 | 4.54 | 0.0332 |
| BAL*Time 6 | 3.7160 | 1.6749 | 4.92 | 0.0265 |
| BAL*Time 8 | 3.8943 | 1.6643 | 5.48 | 0.0193 |
| BAL*Time 11 | 3.6798 | 1.6749 | 4.83 | 0.0280 |
| BAL*Time 12 | 3.4603 | 1.6434 | 4.43 | 0.0352 |

**6. Results.** To show estimation differences between the three approaches all data were first fitted to the three models. The same set of independent variables was used for all three models. For simplicity, we only showed the significant dummy variable and its significant interaction terms with other independent variables. Tables 2–4 showed that the inclusion of the dummy variable and its interaction terms depends on the specific approach. The BA variable was not statistically significant at $\alpha = 0.05$ in the LR and MLR models (Tables 2 and 3). Generally, the MLR model produced smaller standard error estimates than the LR model (Table 4). The reason was that the MLR model used an exchangeable working correlation structure to accommodate for serial correlation during the model fitting process and, therefore, made corrections to estimate standard errors of model coefficients. There are some changes of the magnitude for the coefficient estimates; however, the differences were relatively small. Moreover, the BAL variable was not statistically significant ($p = 0.6482$) in the RLR model (Table 4). It was expected that the RLR model would produced different coefficient estimates and standard errors in relation to the regression coefficients and in comparison to the MLR model due to the different estimation techniques applied to the RLR and MLR models. The random effects assumptions of normality based on histograms with estimated normal density curves were also assessed (Figure 1) as well as the nonparametric Kolmogorov–Smirnov test for the estimated random effects. Comparisons of the

TABLE 3. Parameter estimates and related statistics of the MLR model using an exchangeable working correlation structure.

| Parameter | Estimate | SE | $Z$-value | $p$ |
|---|---|---|---|---|
| Intercept | 18.5931 | 7.8285 | 2.38 | 0.0175 |
| DBH | −0.7400 | 0.3418 | −2.17 | 0.0304 |
| BA | 0.0052 | 0.0042 | 1.22 | 0.2224 |
| BAL | −3.2787 | 1.5647 | −2.10 | 0.0361 |
| Num trees | −0.0057 | 0.0004 | −12.62 | < 0.0001 |
| Time 5 | −18.1819 | 7.8328 | −2.32 | 0.0203 |
| Time 6 | −19.2333 | 8.0214 | −2.40 | 0.0165 |
| Time 8 | −19.4916 | 7.9697 | −2.45 | 0.0145 |
| Time 9 | −15.8196 | 8.0011 | −1.98 | 0.0480 |
| Time 10 | −15.6287 | 7.8401 | −1.99 | 0.0462 |
| Time 11 | −20.4175 | 7.9857 | −2.56 | 0.0106 |
| DBH*Time 11 | 0.7838 | 0.3473 | 2.26 | 0.0240 |
| BAL*Time 5 | 3.5027 | 1.5665 | 2.24 | 0.0254 |
| BAL*Time 6 | 3.7318 | 1.6022 | 2.33 | 0.0198 |
| BAL*Time 8 | 3.9050 | 1.5931 | 2.45 | 0.0142 |
| BAL*Time 11 | 3.6902 | 1.6060 | 2.30 | 0.0216 |
| BAL*Time 12 | 3.4755 | 1.5722 | 2.21 | 0.0271 |

TABLE 4. Fixed parameter estimates and related statistics of the RLR model.

| Parameter | Estimate | SE | $t$-value | $p$ |
|---|---|---|---|---|
| Intercept | 5.7856 | 1.2295 | −4.71 | <0.0001 |
| DBH | −0.1432 | 0.0693 | 2.07 | 0.0389 |
| BA | −0.0270 | 0.0064 | 4.247 | <0.0001 |
| BAL | −0.0294 | 0.0645 | −0.46 | 0.6482 |
| Num trees | −0.0117 | 0.0007 | 16.40 | <0.0001 |
| Time 5 | 3.9024 | 1.4324 | 2.72 | 0.0064 |
| Time 9 | 3.3362 | 1.3035 | 2.56 | 0.0105 |
| Time 11 | 2.8212 | 1.2805 | 2.20 | 0.0276 |
| Time 12 | −3.3646 | 1.2349 | −2.72 | 0.0064 |
| DBH*Time 12 | 0.5011 | 0.0800 | 6.26 | <0.0001 |

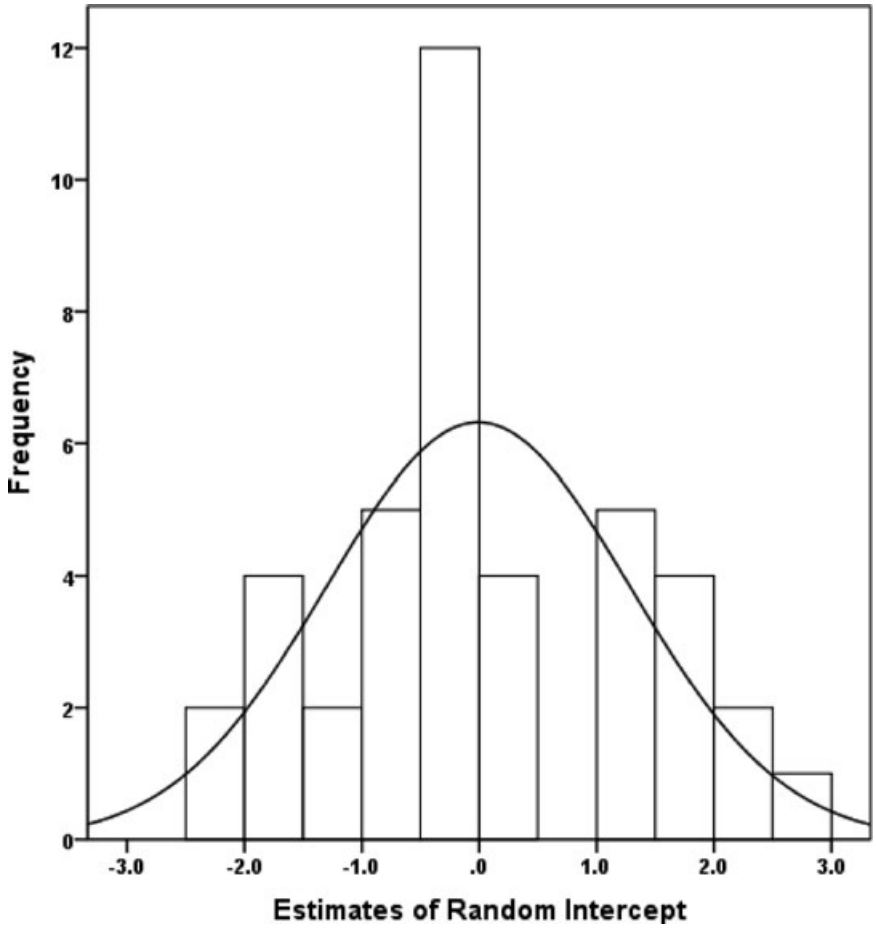


FIGURE 1. A histogram with a normal density curve for the estimated intercepts of the random effects of plots.

frequency of estimated random effects to that which was expected (assuming normal distribution) supports the assumption of normality (Figure 1). Generated $p$-values of the Kolmogorov–Smirnov test were 0.12 for plot random effects, respectively. Thus, normality assumptions at $\alpha = 0.05$ were not rejected.

To evaluate model prediction performance, the 10-fold cross-validation method was used after which the average values of $\bar{e}$, $|\bar{e}|$,$\sqrt{v}$, and $MS$ were computed (see Table 5). The LR model possessed the largest positive$\bar{e}$. This indicates that the LR model tended to under predict the probability of mortality across all diameter classes. Moreover, the LR and MLR models produced similar values for $|\bar{e}|$,$\sqrt{v}$, and

*MS*, suggesting that the MLR model did not provide a better model prediction ability even when serial correlation in the parameter estimation was incorporated. Between the three models, the RLR model produced the smallest$\bar{e}$, $|\bar{e}|$,$\sqrt{v}$, and *MS*.

Predicted mortality between the diameter classes of the three models under investigation and the observed mortality rates were compared (Figure 2). The LR and MLR models exhibited similar predicted mortality rates between diameter classes. The RLR model, however, was able to provide better predictive capacity for both small and large trees, although it over predicted tree mortality of medium size trees. Model evaluation statistics of the mean prediction error ($\bar{e}$), the average absolute prediction error ($|\bar{e}|$), and the variance prediction error ($\sqrt{v}$) were also calculated and plotted in 3 cm classes (Figure 3). The RLR model consistently produced the smallest $\bar{e}$, $|\bar{e}|$, and $\sqrt{v}$ between the three models across all diameter classes while the LR and MLR models behaved in a similar manner across all diameter classes.

**7. Discussions.** For this study standard logistic (LR), marginal GEE logistic (MLR), and random logistic (RLR) regression models were fitted to assess their capacity to model individual tree mortality rates. Results were then compared.

A working correlation structure that reflects the average dependence over all subjects must be chosen for the MLR model. Although inferences on parameter estimates along with working correlations in the MLR model are asymptotically correct (Liang and Zeger, 1986), i.e., even if the correlation structure within subjects is not known and incorrectly specified, results of the analysis will be more or less the same. It is better to chose a structure based on the characteristics of the data. For this study, the exchangeable correlation structure was selected due to uneven measurement periods between plots (Allison, 1999; Kiernan et al., 2009) and also because it is based on the QIC criterion (Pan, 2001). Similarly, the number of random effects must first be chosen for the RLR model since hierarchical structure such as tree, plot, and stand levels are typically treated as random effects in forest research (Jutras et al., 2003; Fortin et al., 2008). Convergence could not be reached with all three levels of random effects at once. Ultimately, the estimation converged when the tree level random effect was dropped. Tests carried out on the likelihood ratio showed that the inclusion of plot random effects significantly improved model fitting. Due to this, plot random effect was kept within the model. Furthermore, the variances and covariance structures (typically referred to as R-sides effects) of the tree level random effects was also taken into consideration to remove possible serial correlation within trees since tree level random effects were dropped. However, model estimation did not converge, indicating that plot random effect may be effective enough on their own to remove both serial correlation derived from the repeated measurements of trees and the correlation derived from the hierarchal data structure. Plot was therefore kept as random effects in the final RLR model.

TABLE 5. Evaluation statistics for all three models taken from the 10-fold cross-validation data sets.

| Statistic | LR | MLR | RLR |
|---|---|---|---|
| $\bar{e}$ | 0.0100 | 0.0083 | −0.0002 |
| $\lvert\bar{e}\rvert$ | 0.1899 | 0.1896 | 0.1800 |
| $\sqrt{v}$ | 0.3142 | 0.3141 | 0.3001 |
| *MS* | 0.0988 | 0.0987 | 0.0900 |

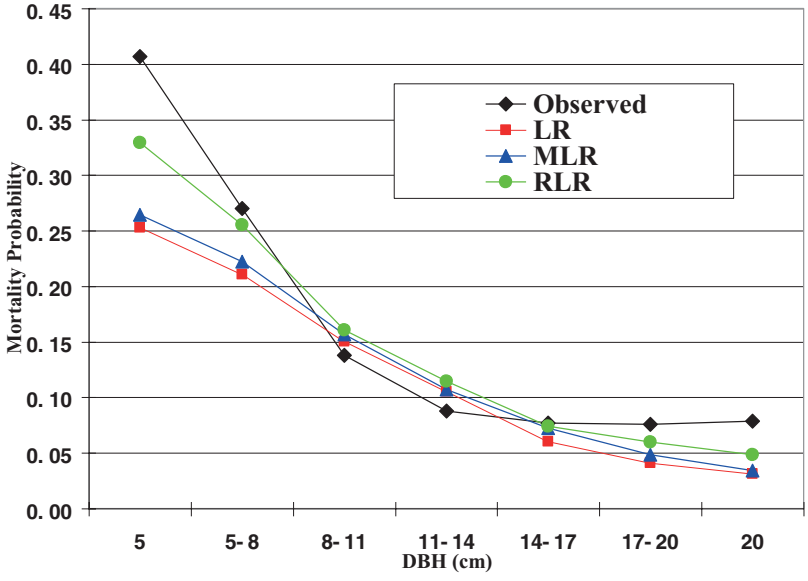


FIGURE 2. Observed mortality and predicted mortality rates using all three models across 3 cm diameter classes.

The Gauss-Hermite quadrature approach was applied in this study to estimate RLR model performance since it provided more accurate approximations compared to the PQL or Laplace methods (Pinheiro and Chao, 2006; Bolker et al., 2009). In addition, Gauss-Hermite quadrature approximated the true likelihood rather than quasi-likelihood. AIC based on likelihood, therefore, can also be used to select random effects.

The interpretation of the regression coefficients of the MLR and RLR models were not straightforward. The MLR (GEE) approach proposed by Liang and Zeger

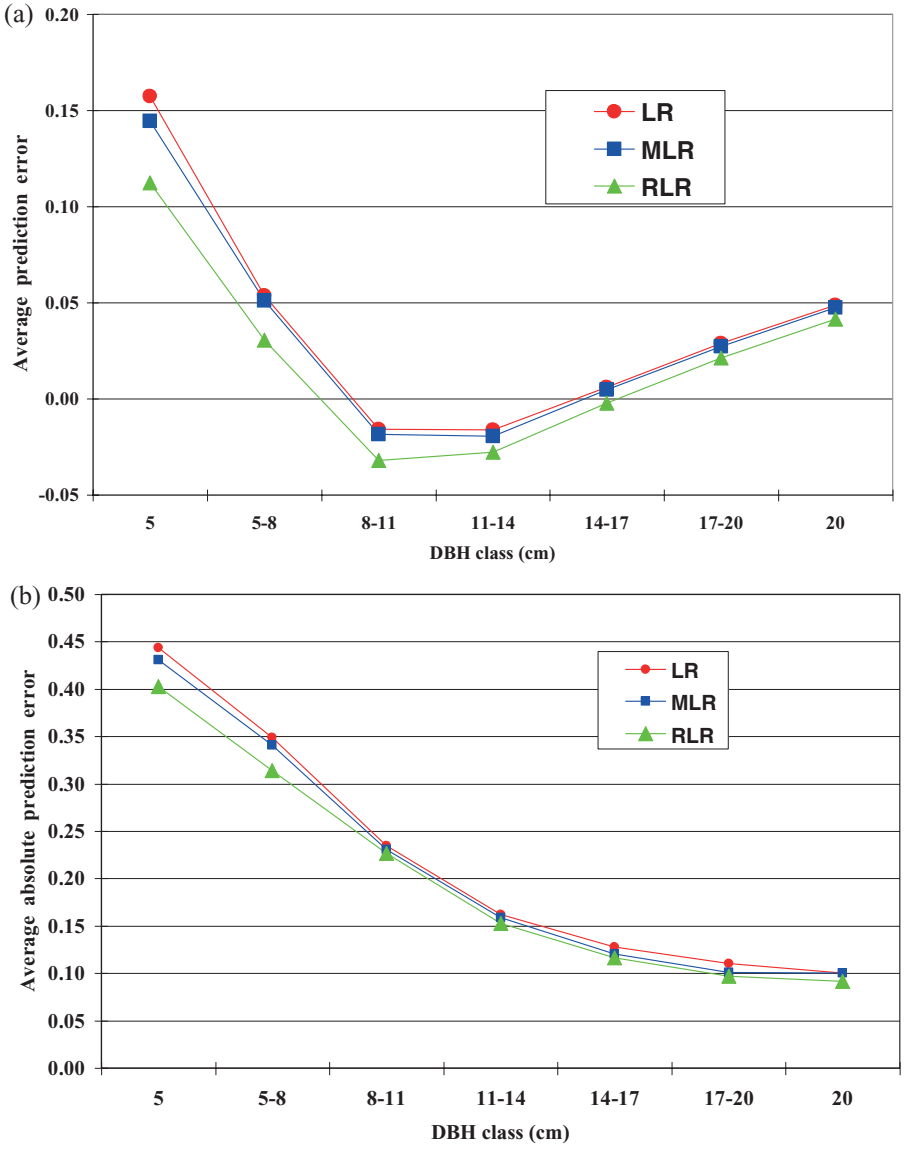


FIGURE 3. Evaluation of model performance using (a) the average prediction error, (b) the average absolute prediction error, and (c) the standard deviation of prediction error by way of the 3 cm diameter classes for the 10-fold cross-validation.

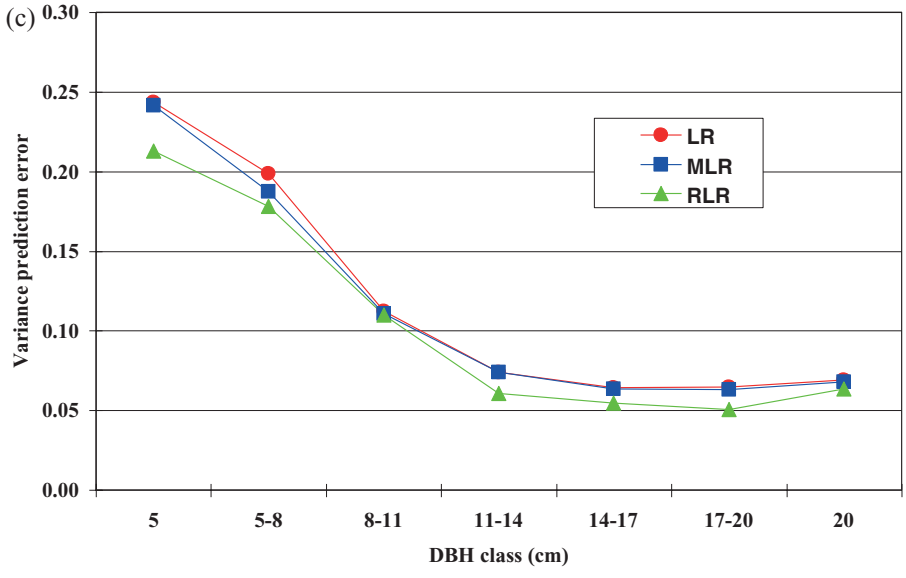


FIGURE 3. Continued.

(1986) only allows for marginal inference. Thus, regression coefficients of the MLR model show the average trend of regression lines. This is why the MLR model is referred to as "population-averaged." In contrast, the RLR model incorporates random effects and thus allows for subject-specific inference. The RLR model assumes that correlation between repeated measurements within individuals under investigation derived from some latent or unobserved characteristic. Random effects are therefore used to reflect these characteristics (Allison, 1999; Fieberg et al., 2009; Koper and Manseau, 2009). In this assessment, plot random effect was included within the RLR model. The random coefficients of the plot effect can reflect unobserved plot effects such as slope, soil characteristics, drainage, and the history of the stand. For this study, plot random effect was statistically significant based on the likelihood ratio test.

For parameter estimates, statistical tests changed as did magnitudes. For example, BAL was significant in the MLR model but insignificant in the RLR model (Tables 3 and 4). The BA variable changed from a nonsignificant variable to a significant variable if we set $\alpha = 0.05$. It seems BA also changed the sign from positive in LR and MLR models to negative sign in RLR model, but the estimate of BA in LR ($p = 0.1860$) and MLR ($p = 0.2224$) models are not significant at $\alpha = 0.05$. In other words, the estimates of BA in LR and MLR models are not different to zero and there is no relationship between tree mortality and BA variable. Thus, BA in LR and MLR has no significant impacts on tree mortality according to results from LR

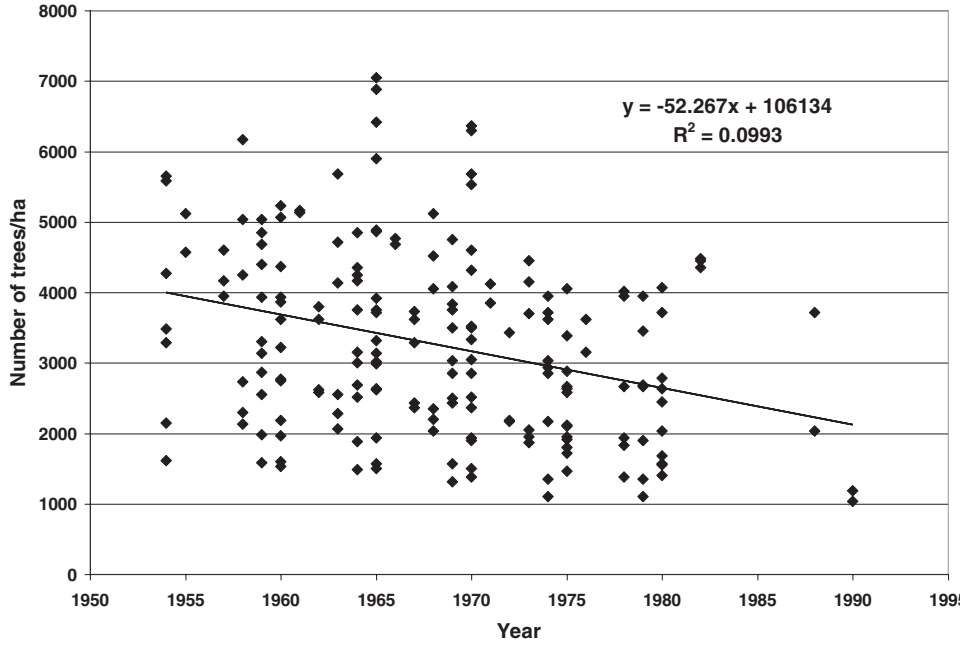


FIGURE 4. Relationship between number of trees and time.

and LMR models. In recent years, ecological studies have shown that the presence of autocorrelation has a strong influence on estimating the variability of regression coefficients. It may change the relative importance (e.g., magnitude changes in our study) of explanatory variables in the model (e.g., Bini et al., 2009). Because the rationale for selecting a suitable modeling technique is commonly lacking, one may need to try a number of modeling methods in order to choose a suitable one for the given data set (Dormann et al., 2007). In other words, one may not know which method is best for the data set with autocorrelation until the methods are applied and compared. From those previous studies, it is therefore not surprise to find the significance of BAL changed.

An interesting discovery was that the signs of Num Trees variable were all negative. This indicates that mortality rates may increase as trees decline in number. To investigate this relationship, the trends of the number of trees were plotted with year (Figure 4). It was revealed that a significant declining trend in tree number occurred as time increased. Stand level mortality rates were also calculated. They were computed as the number of dead trees divided by the number of live trees for each measurement of each plot. Stand level mortality and time were then plotted (Figure 5). A significant increasing trend was found in relation to stand level

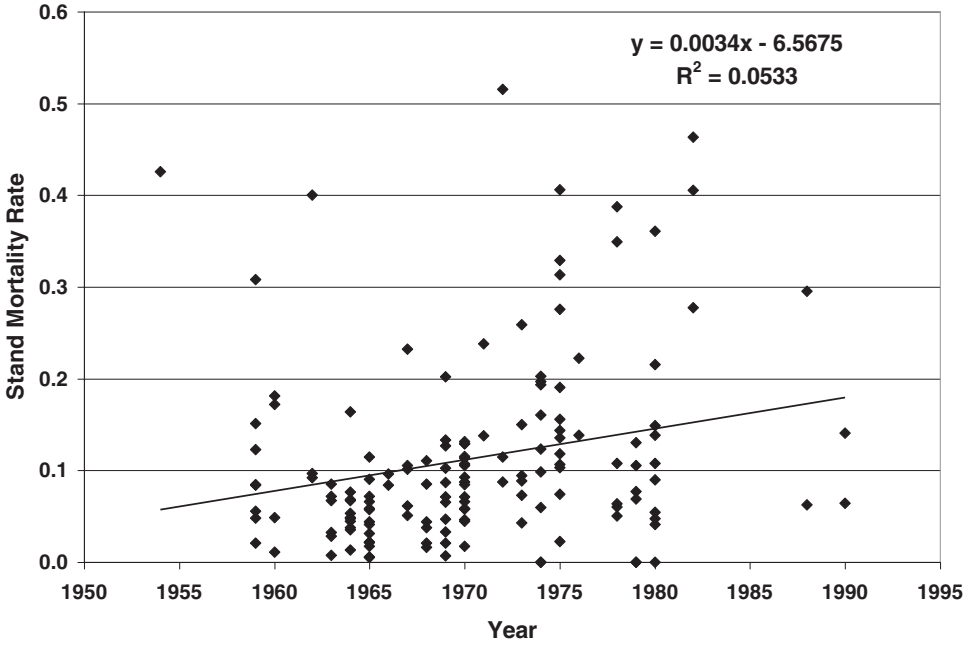


FIGURE 5. Relationship between stand level mortality rate and time.

mortality rates. This may indicate that the probability of individual tree morality rates also increased. The inference of this finding suggests that morality rates may have increased in recent years, and forest stands suffered more stress from other factors such as climate change (e.g., an increase in temperature may promote an increase in insects and disease that attack trees) (van Mantgem et al., 2009).

MLR model estimation was not based on maximum likelihood. Due to this, certain maximum likelihood-based criteria for model selection such as the Akaike information criterion (Burnham and Anderson, 1998) and the traditional $\chi^2$goodness of fit tests are unavailable and cannot be applied to the MLR model. $K$-fold cross-validation was used to evaluate model predictions of the LR, MLR, and RLR models as a result. Koper and Manseau (2009) applied $K$-fold cross-validation to show the predictive capacity of the GEE and the generalized linear mixed effects models. Their study, however, did not directly compare model predictions between the two models. Here, we set $K = 10$ according to the study of McLachlan et al. (2004). For this study, 10% of the total number of trees was used as validation data. The remaining 90% of trees served as calibration data for each cross-validation. We did not select a subset data to serve as an independent validation data set because we can not obtain the plot-specific random effects for the independent validation plots

and we can only get the estimated random effects for plots involved the model calibration. Thus, it may be very difficult for us to compare the model performance between three modeling techniques. Irregular measurement time periods were treated as independent dummy variables. This made the prediction of individual tree mortality rates over specific time periods possible. However, by treating the census interval as dummy variables, we can only use the fitted models to predict mortality for some measurement periods that are used in the model. Finally, four evaluation statistics based on model residuals were calculated and plotted. The RLR model provided better overall model predictive capacity compared to the MLR model. This resulted from the RLR model capturing variation between plots that derived from certain unobserved variables (e.g., soil types and drainage) through the inclusion of random effect. The MLR method based on GEE was not able to correct itself due to the omission of these variables (Allison, 1999). On the other hand, the MLR model, which considered correlations among repeated measurements, did not produce noticeably better model prediction than the LR model, which assumed independent observations. A possible reason for this result can be attributed to the fact that there are a very large number of trees from different plots. The correlation structure used by the MLR model only reflect the average correlation for within trees from different plots, therefore, the average correlation structure may not capture the variation of correlation for trees from different plots. Moreover, the influence of the average correlation structure used by the MLR model was so small that it had little impact on regression coefficients. For the RLR model, we obtained random coefficient for each plot. Thus, the random coefficients reflect the plot random variations and may better capture the correlation within in trees in the plots. The estimates that accommodate correlations between repeated measurements typically are different from the model-based standard errors that assume independent observations. The robust standard errors of the MLR model will, therefore, provide correct inferences (e.g., the standard errors seen in Table 3 are different to those in Table 2).

The choice between MLR and RLR models depends primarily on the objectives of the study. If the response variable is normally distributed and "identity" is the link function, the MLR and RLR models will perform identically and, thus, provide identical interpretations. However, when using such models with nonlinear link functions (e.g., logit) for a non-normal response variable, a selection between the different approaches must be made since the regression coefficients of the two approaches possess different interpretations (Carriere and Bouyer, 2002; Fitzmaurice et al., 2004; Fieberg et al., 2009). If the aim of a researcher is to provide an inference of the mortality rate of the entire population, the MLR model may be preferred to provide marginal population estimates. More specifically, estimated regression coefficients in the MLR model describe the effects of independent variables on population average response (Fitzmaurice et al., 2004). For example, the MLR model can be used to answer the question: how management actions (such as thinning) will influence the mortality rate of entire populations. In contrast, if the aim is

to study the mortality rate of trees in specific sites that are developed over time, the RLR model would be preferred to answer such questions. That is to say the estimated regression coefficients may vary across different sites, and random coefficients can reflect the correlation between trees clustered within the same plot and the correlation between repeated measurements taken for given trees (Fitzmaurice et al., 2004). For example, the RLR model may be applied to answer the question: how management actions (such as irrigation and fertilizer usage) will impact the mortality rates of trees within specific sites.

**8. Conclusion.** Recently, MLR and RLR model statistical methods were used to model tree mortality rates by considering correlations among observations. However, little information exists in scientific literature regarding their relative performance. For this assessment, three models (LR, MLR, and RLR) were fitted using the same set of independent variables. Their performance was compared by means of the $K$-fold cross-validation approach. The RLR model incorporating both plot and time random effects gave the best performance based on the evaluation statistics of the mean of the prediction error, the average absolute prediction error, the variance prediction error, and the mean square error. The MLR model did not provide an evident improvement of model performance. This was due to the usage of a large number of trees for which few measurements were taken per individual tree. As a consequence, the working correlation structure had little impact on regression coefficients. This study has shown that the RLR model incorporating both plot and time random effects was able to remove serial correlation from repeated measurements as well as variations derived from plots while the MLR model was capable in dealing with serial correlation from repeated measurements but unable to reflect the variation between plots due to the omission of certain unobserved variables.

**Acknowledgments.** This study was supported by the National Science and Engineering Research Council of Canada (NSERC) Strategic Network (ForValueNet), an NSERC discovery grant and the China QianRen programme. We thank V. LeMay for her comments and discussions on the earlier draft of the paper and Brian Doonan for his editorial help.

## REFERENCES

P. Allison [1999], *Logistic Regression using the SAS System: Theory and Application*, SAS Institute Inc., Cary, NC, p. 288.

L.M. Bini, J.A.F. Diniz-Filho, T.F.L.V.B. Rangel, T. S.B. Akre, R.G. Albaladejo, F.S. Albuquerque, A. Aparicio, M.B. Araújo, A. Baselga, J. Beck, M.I. Bellocq, K. Böhning-Gaese, P.A.V. Borges, I. Castro-Parga, V.K. Chey, S.L. Chown, P.D. Marco, D.S. Dobkin, D. Ferrer-Castán, R. Field, J. Filloy, E. Fleishman, J.F. Gómez, J. Hortal, J.B. Iverson, J.T. Kerr, W.D. Kissling, I.J. Kitching, J.L. León-Cortés, J.M. Lobo, D. Montoya, I. Morales-Castilla, J.C. Moreno, T. Oberdorff, M.Á. Olalla-Tárraga, J.G. Pausas, H. Qian, C. Rahbek, M.Á. Rodríguez, M. Rueda, A. Ruggiero, P. Sackmann, N.J. Sanders, L.C. Terribile, O.R. Vetaas, and B.A. Hawkins [2009],

*Coefficient Shifts in Geographical Ecology: An Empirical Evaluation of Spatial and Non-Spatial Regression*, Ecography **32**, 193–204.

B.M. Bolker, M.E. Brooks, C.J. Clark, S.W. Geange, J.P. Poulsen, M.H.H. Stevens, and J.S. White [2009], *Generalized Linear Mixed Models: A Practical Guide for Ecology and Evolution*, Trends Ecol. Evol. **24**, 127–135.

K.P. Burnham and D.P. Anderson [1998], *Model Selection and Inference: A Practical Information-Theoretic Approach*, Springer-Verlag, New York.

R. Calama and G. Montero [2005], *Multilevel Linear Mixed Model for Tree Diameter Increment in Stone Pine (Pinus pinea): A Calibrating Approach*, Silva Fenn. **39**, 37–54.

I. Carrière and J. Bouyer [2002], *Choosing Marginal or Random-Effects Models for Longitudinal Binary Responses: Application to Self-Reported Disability Among Older Persons*, BMC Med. Res. Technol. **2**, 15.

C.S. Davis [2002], *Statistical Methods for the Analysis of Repeated Measurements*, Springer-Verlag, New York.

C.F. Dormann, J.M. McPherson, M.B. Araújo, R. Bivand, J. Bolliger, G. Carl, R.G. Davies, A. Hirzel, W. Jetz, W.D. Kissling, I. Kühn, R. Ohlemüller, P.R. Peres-Neto, B. Reineking, B. Schröder, F.M. Schurr, and R. Wilson [2007], *Methods to Account for Spatial Autocorrelation in the Analysis of Species Distributional Data: A Review*. Ecography **30**, 609–628.

B. Efron and R. Tibshirani [1993], *An Introduction to the Bootstrap*, Chapman and Hall, New York, London.

T. Eid and E. Tuhus [2001], *Models for Individual Tree Mortality in Norway*, Forest Ecol. Manag. **154**, 69–84.

J. Fieberg, R.H. Rieger, M.C. Zicus, and J.S. Schildcrout [2009], *Regression Modelling of Correlated Data in Ecology: Subject-Specific and Population Averaged Response Patterns*, J. Appl. Ecol. **46**, 1018–1025.

G.M. Fitzmaurice, N.M. Laird, and J.H. Ware [2004], *Applied Longitudinal Analysis*, John Wiley & Sons, Hoboken, New Jersey.

M. Fortin, S. Bédard, J. DeBlois, and S. Meunier [2008], *Predicting Individual Tree Mortality in Northern Hardwood Stands Under Uneven-Aged Management in Southern Québec, Canada*, Ann. For. Sci. **65**, 205.

J.C. Fox, H.Q. Bi, and P.K. Ades [2007], *Spatial Dependence and Individual Tree Growth Models I. Characterising Spatial Dependence*, Forest Ecol. Manag. **245**, 10–19.

S. Jutras, H. Hökkä, V. Alenius, and H. Salminen [2003], *Modeling Mortality of Individual Trees in Drained Peatland Sites in Finland*, Silva Fenn. **37**, 235–251.

D. Kiernan, E. Bevilacque, R. Nyland, and L. Zhang [2009], *Modeling Tree Mortality in Low- to Medium-Density Uneven-Aged Hardwood Stands Under a Selection System using Generalized Estimating Equations*, For. Sci. **55**(4), 343–351.

R.K. Kobe and K.D. Coates [1997], *Models of Sapling Mortality as a Function of Growth to Characterize Interspecific Variation in Shade Tolerance of Eight Tree Species of Northwestern British Columbia,* Can. J. For. Res. **27**, 227–236.

N. Koper and M. Manseau [2009], *Generalized Estimating Equations and Generalized Linear Mixed-Effects Models for Modelling Resource Selection*, J. Appl. Ecol. **46**, 590–599.

Y. Lee [1971], *Predicting Mortality for Even-Aged Stands of Lodgepole Pine*, Forest. Chron. **47**, 29–32.

K.Y. Liang and S.L. Zeger [1986], *Longitudinal Data Analysis Using Generalized Linear Models*, Bometrika **73**, 13–22.

D. Mailly, M. Gaudreault, G. Picher, I. Auger, and D. Pothier [2009], *A Comparison of Mortality Rates Between Top Height Trees and Average Site Trees*, Ann. For. Sci. **66**(2), 202.

G.J. McLachlan, K.A. Do and C. Ambroise [2004], *Analyzing Microarray Gene Expression Data*, Wiley Series in Probability and Statistics.

W. Pan [2001], *Akaike's Information Criterion in Generalized Estimating Equations*, Biometrics **57**, 120–125.

J.F. Pendergast, S.J. Gange, M.A. Newton, M.J. Lindstrom, M. Palta, and M.R. Fisher [1996], *A Survey of Methods for Analyzing Clustered Binary Response Data*. Int. Stat. Rev. **64**, 89–118.

J.C. Pinheiro and E.C. Chao [2006], *Efficient Laplacian and Adaptive Gaussian Quadrature Algorithms for Multilevel Generalized Linear Mixed Models*, J. Comput. Graph. Stat. **15**, 58–81.

C.E. Rose, D.B. Hall, D.B. Shiver, M.L. Clutter, and B. Border [2006], *A Multilevel Approach to Individual Tree Survival Prediction*, For. Sci. **52**, 31–43.

SAS Institute Inc. [2008]. *The GENMOD Procedure*, Cary, NC, USA.

G.L. Somers, R.C. Oderwald, W.R. Harris, and O.G. Langdon [1980], *Predicting Mortality with a Weibull Function*, For. Sci. **26**, 291–300.

J.W. Twisk [2004], *Longitudinal Data Analysis. A Comparison Between Generalized Estimating Equations and Random Coefficient Analysis*, Eur. J. Epidemiol. **19**(8), 769–776

P.J. van Mantgem, N.L. Stephenson, J.C. Byrne, L.D. Daniels, J.F. Franklin, Z. Fule Peter, M.E. Harmon, A.J. Larson, J.M. Smith, A.H. Taylor, and T.T. Veblen [2009], *Widespread Increase of Tree Mortality Rates in the Western United States*, Science **323**(5913), 521–524.

X. Yao, S.J. Titus, and A. MacDonald [2001], *A Generalized Logistic Model of Individual Tree Mortality for Aspen, White Spruce, and Lodgepole Pine in Alberta Mixedwood Forests*, Can. J. For. Res. **31**, 283–291.

L. Zhang [1997], *Cross-Validation of Non-Linear Growth Functions for Modeling Tree Height-Diameter Relationships*, Ann. Bot. **79**, 251–257.

L. Zhang, C. Peng, and Q. Dang [2004], *Individual-Tree Basal Area Growth Models for Jack Pine and Black Spruce in Northern Ontario*, Forest. Chron. **80**, 366–374.

D. Zhao, B. Borders, and M. Wang [2006], *Survival Model for Fusiform Rust Infected Loblolly Pine Plantations with and without Mid-Rotation Understorey Vegetation Control*, Forest Ecol. Manag. **235**, 232–239.