# CONSTRUCTING MORTALITY MODELS FOR NATURAL FOREST STATE III IN FOUR PROVINCES OF THE CENTRAL REGION, VIETNAM

**Cao Thi Thu Hien**
*Vietnam National University of Forestry*

In this research, the dead trees were defined as the standing trees that died between the two occasions at which measurements were taken. The data on 300 subplots from 12 permanent sample plots were collected. The response variable was the number of dead trees per subplot. The results suggest that we successfully developed the mortality model by using both generalized linear and generalized linear mixed models for count data to address the problem of overdispersion. Arithmetic mean diameter of the subplot, plot basal area, and provinces as a categorical variable were found to be the most significant explanatory variable. With the generalized linear model, we found that the Negative Binomial GLM was the most appropriate model for predicting the number of recruitment for three groups. From using the provinces as a grouping variable, we realized that the mean numbers of dead trees was different in the four different locations, namely the number of dead trees for group 1 and group 2 in Thua Thien Hue were the highest, while in Ha Tinh these were the lowest. With the generalized linear mixed model, the Negative Binomial GLMM solves overdispersion by treating a plot as a random effect. The GLMM with random intercept was selected as the equation for the direct prediction of dead trees across each of the two species groups, and for *N. melliferum*. The GLMM with a random slope was chosen for *S. wightianum*.

**Keywords: Generalized linear model, generalized linear mixed model, mortality model, Negative Binomial GLM, tropical rainforests.**

## 1. INTRODUCTION

Natural mortality of trees is a crucial process that determines forest dynamics (Rüger *et al.*, 2011). When a tree dies, the reduced competition benefits the trees near the dead tree, positively affecting their growth (Yang *et al*, 2003); in addition, gaps created by dead canopy trees are later filled by new trees (Oliver and Larson, 1996). McCarthy (2001) notes that these gap dynamics are crucial determinants of the structure and composition of a forest stand. For these reasons, the mortality process should be considered in stand simulation models. However, modeling mortality is difficult due to the stochastic nature of mortality events; standing death may be caused by intrinsic senescence (Carey *et al.*, 1994) or extrinsic factors such as disease, insects, fungi, and wind. In previous studies, several statistical methods have been utilized to develop mortality models, including the logistic regression model (Monserud and Sterba, 1999), the two-step approach (Eid and Tuhus, 2001; Álvarez González *et al.*, 2004; Diéguez-Aranda *et al.*, 2005), the three-step approach (Fridman and Stahl, 2001; Meng *et al.*, 2003), and neural networks (Hasenauer *et al.*, 2001).

Moreover, as recruitment data, mortality data consist of many zeros. Fortin and Deblois (2007), for instance, demonstrated that fitting a traditional Poisson distribution to this type of data can underestimate the occurrence of zeros or overestimate the occurrence of larger counts. One way to solve this issue is utilizing a method similar to conditional functions. Fortin and Deblois (2007) predicted tree recruitment with zero-inflated models, and Zhang *et al.* (2012) applied negative binomial mixture models (zero-inflated negative binomial, and Hurdle negative binomial models) to predict tree recruitments of Chinese pine trees (*Pinus tabulaeformis*).

However, in the present study, we did not use zero-inflated models because of two reasons: (1) A zero-inflated model assumes that the zero observations have to come from two different sources, namely “structural” and "sampling" (Hu *et al.*, 2011). The sampling zeros are assumed to occur by chance, while structural zeros are observed due to some specific structure in the data. (2) Zero-inflated models are recommended if the overdispersion parameter is larger than 15 or 20 (Zuur *et al.*, 2009), what was not the case with the data of the 12 sample plots under study.

For this study, due to the low number of plots compared to the large number of species the purpose of this research is not to tackle the mortality of each single species, but rather, to concentrate on predicting tree mortality of particular species groups: across all tree species, all locally (province) important tree species, and important species spread over provinces. Two approaches were used here, generalized linear models (Poisson, Quasi-Poisson and Negative Binomial models), and generalized linear mixed effects models (Negative Binomial mixed model), the latter to take random plot effects into account.

## 2. RESEARCH METHODOLOGY

### 2.1. Study area

Measurements were taken in a tropical rainforest, in four different provinces of Central region of Vietnam: Ha Tinh province, Thua Thien Hue province, Binh Dinh province and Khanh Hoa province. There were three plots in each of the four provinces.

### 2.2. Data collection

In this research, 12 permanent sample plots (PSPs) in four provinces were selected from the network of PSPs which was established by the Forest Inventory and Planning Institute (FIPI) of Vietnam. Data from 2005 inherited, and re-measurement of these plots was done by the author in 2012, 2013.

Each plot has a square shape (100 m x 100 $m^2$) and is divided into twenty five 20 m x 20 m quadrats. It was aligned according to a magnetic-north direction and has four major corner posts made of concrete. All trees equal to or larger than 6 cm diameter at breast height ($DBH \geq 6$ cm) were identified by species and permanently marked using a white metal tag.

In 2005: On each plot, all trees in each plot with a diameter at breast height from 6 cm ($DBH \geq 6$ cm) were marked and, identified by species; their diameter was measured at 1.3 m from the ground. The data within the plot were assigned to their 20 m x 20 m quadrat.

In 2012 and 2013: Measurements were repeated on all 12 plots and standing dead trees were also recorded.

### 2.3. Data analysis

#### *2.3.1. Species group*

There are a large number of tree species in natural tropical rainforests. Several species appear more frequently, some occur with only low frequency. For that reason, species might be aggregated into some groups to reduce the number of mortality models and to avoid the need for adding data for species with insufficient number of observations. For our study, simply the importance value index (*IVI*) was used to determine a group of most important species.

Important tree species having $IVI \geq 5\%$ in pooled data from three plots in each province were utilized to construct mortality model.

#### *2.3.2. Mortality model*

In our case, the data on 300 subplots from 12 plots were used to fit the mortality model. The response variable was the dependent variable was the number of dead trees, which are the standing trees that died between the two occasions at which measurements were taken. Explanatory variables were measured at the beginning of the period, including arithmetic mean diameter of the subplot (*DBH*), subplot and plot basal area ($BAL_{subplot}$, $BAL_{stand}$), subplot density (the number of trees on each subplot) (*N*), and provinces as a categorical variable (*provincecode*).

- *Generalized linear model (GLM)*

The GLM consists of two components, the response variable and the link function. The link function defines how the mean of the dependent variable and the linear combination of the explanatory variables are connected (Faraway, 2006).

In this study, a Poisson GLM (log link), Quasi-GLM and Negative Binomial were used, where the Poisson GLM (log link) was used to detect overdispersion.

In the Poisson model, the variance equaled $\phi\mu$, with mean $\mu$ and dispersion parameter $\phi$. The following formula (Zuur *et al.*, 2009) was used in the calculation:

$$\emptyset = \frac{D}{n-p} \tag{1}$$

Where: $D$ is the residual deviance and $n - p$ is degrees of freedom. $n$ is the number of observations, and $p$ is the number of regression parameters (intercept and slopes) in the model. If $\phi$ equals 1, there is no overdispersion and we have the Poisson GLM; if $\phi$ is larger than 1, this is evidence for the suggestion of overdispersion (Zuur *et al.*, 2009), the Quasi-Poisson GLM and Negative Binomial models were used:

$$\log(\mu_i) = \beta_0 + \alpha_k + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \log(time_i) \tag{2}$$

$$or\ \mu_i = \exp(\beta_0 + \alpha_k + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \log(time_i)) \tag{3}$$

Where:

$x_{1i}$ to $x_{4i}$ are independent variables of the $i^{th}$ subplot ($DBH$, $BAL_{subplot}$, $BAL_{stand}$, $N$) and $\alpha_k$ is the effect of province $k$ ($k = 1, 2, 3$), $\alpha_k = 0$ for Ha Tinh;

$\beta_0$ to $\beta_4$ and the $\alpha_k$ are the parameters to be estimated;

$\log(time_i)$ is an offset factor.

To estimate the regression parameters of the GLM, a maximum likelihood estimation was used (Zuur *et al.*, 2013). The procedure for selecting Poisson, Quasi-Poisson, and Negative Binomial models followed Zuur *et al.* (2009).

- *Generalized linear mixed model (GLMM)*

Generalized linear mixed models are an extension of a GLM in which the linear predictor contains random effects in addition to the fixed effects. The random effects can account for the correlation between observations from the same plot in a province. In this study, a random plot effect was added to the intercept (equation 4), the slope (equation 5), or both intercept and slope (equation 6) of each model. The general equation is as follows:

$$\log(\mu_{ij}) = (\beta_0 + b_j) + \alpha_k + \beta 1.x1ij + \beta 2.x2ij + \beta 3.x3ij + \beta 4.x4ij + \log(time_{ij}) \tag{4}$$

$$\log(\mu_{ij}) = \beta_0 + \alpha_k + (\beta 1 + b_{1j}).x1ij + (\beta 2 + b_{2j}).x2ij + (\beta 3 + b_{3j}).x3ij + (\beta 4 + b_{4j}).x4ij + \log(time_{ij}) \tag{5}$$

$$\log(\mu_{ij}) = (\beta_0 + b_j) + \alpha_k + (\beta 1 + b_{1j}).x1ij + (\beta 2 + b_{2j}).x2ij + (\beta 3 + b_{3j}).x3ij + (\beta 4 + b_{4j}).x4ij + \log(time_{ij}) \tag{6}$$

To assess goodness of fit of the data in the GLM and GLMM models, Pearson's $\chi^2$ was used. The parameter estimations for the GLMM in this chapter were fitted with *glmmPQL* in "MASS" package available from the open source statistical software R.

All hypothesis testing was performed at the $\alpha = 0.05$ significance level.

## 3. RESULTS

### 3.1. Results of the GLM

A total of 1323 dead trees belonging to 189 species were counted at the four locations. The number of dead trees for all species and all important species counted per plot ranged from 21 to 221 and from 10 to 133 respectively. This was in correspondence with the number of species, which was respectively, from 17 to 58 and from 8 to 21 (Table 1).

**Table 1. Descriptive statistics of the mortality data used for the model development**

| Province | Plot | No. trees | $\overline{DBH}$ (cm) | Stand basal area (m$^2$/ha) | All tree species | | All important tree species | |
|---|---|---|---|---|---|---|---|---|
| | | | | | No. dead trees | No. species | No. dead trees | No. species |
| Ha Tinh | 1 | 416 | 20.95 | 18.65 | 21 | 17 | 10 | 8 |
| | 2 | 352 | 19.95 | 13.84 | 69 | 33 | 33 | 9 |
| | 3 | 391 | 19.05 | 15.54 | 66 | 37 | 25 | 9 |
| ***Total*** | | | | | ***156*** | ***66*** | ***68*** | ***14*** |
| Thua Thien Hue | 4 | 932 | 17.35 | 33.04 | 154 | 44 | 109 | 21 |
| | 5 | 855 | 18.20 | 33.50 | 77 | 41 | 45 | 19 |
| | 6 | 1092 | 16.53 | 34.72 | 190 | 54 | 133 | 21 |
| ***Total*** | | | | | ***421*** | ***73*** | ***287*** | ***21*** |
| Binh Dinh | 7 | 1151 | 16.05 | 31.01 | 221 | 57 | 130 | 17 |
| | 8 | 967 | 16.62 | 31.72 | 184 | 58 | 108 | 17 |
| | 9 | 893 | 18.34 | 32.44 | 96 | 43 | 47 | 12 |
| ***Total*** | | | | | ***501*** | ***86*** | ***285*** | ***17*** |
| Khanh Hoa | 10 | 800 | 17.53 | 28.95 | 35 | 18 | 25 | 9 |
| | 11 | 782 | 17.52 | 24.46 | 82 | 27 | 59 | 9 |
| | 12 | 901 | 17.47 | 28.73 | 128 | 27 | 99 | 11 |
| ***Toal*** | | | | | ***245*** | ***42*** | ***183*** | ***11*** |
| **Total (4 prov.)** | | | | | **1323** | **189** | **823** | **49** |

The fitted Poisson GLM model for two species groups (all tree species and all important tree species) supported evidence for overdispersion through the ratio of deviance to degrees of freedom larger than 1 (2.79 and 2.38, respectively). Thus, we refitted the data to correct the standard errors using Quasi-Poisson and Negative Binomial GLMs.

The estimated parameters, standard errors, and the *p*-values of Poisson, Quasi-Poisson, and Negative Binomial models are represented in table 2. The deviance across all species and all important species was the lowest when analyzed with a Negative Binomial GLM, leading us to conclude that the Negative Binomial model was preferrable over the Poisson and Quasi-Poisson models. The Negative Binomial GLM is given as follows:

For all species:

$$\log(\mu_i) = \beta_0 + \alpha_k + \beta_1 BAL_{standi} + \log(time_i) \quad (7)$$

and for all important species:

$$\log(\mu_i) = \beta_0 + \alpha_k + \beta_1 DBH_i + \beta_2 BAL_{standi} + \\ + \log(time_i) \quad (8)$$

In the same way, the Negative Binomial GLM was the selected model for three important species (*Syzygium wightianum*, *Diospyros sylvatica* and *Nephelium melliferum*) spread over three or four provinces:

*S. wightianum:*

$$\log(\mu_i) = \beta_0 + \alpha_k + \beta_1 BAL_{standi} + \log(time_i \quad (9)$$

*D. sylvatica*:

$$\log(\mu_i) = \beta_0 + \alpha_k + \log(time_i) \quad (10)$$

*N. melliferum*:

$$\log(\mu_i) = \beta_0 + \log(time_i) \quad (11)$$

Estimated parameters and p-value values of the Poisson, Quasi-Poisson and Negative Binomial equations are shown in table 2.

**Table 2. GLM (Poisson, Quasi-Poisson, Negative Binomial) results for standing dead trees (0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05)**

| Objects | Variables | Parameters | Poisson | | Quasi-Poisson | | Negative Binomial | |
|---|---|---|---|---|---|---|---|---|
| | | | Parameter estimates | Pr(>\|z\|) | Parameter estimates | Pr(>\|z\|) | Parameter estimates | Pr(>\|z\|) |
| All tree species | Intercept | $\beta_0$ | 0.9764 | 0.0152* | 0.3103 | 0.614253 | 0.4251 | 0.4287 |
| | *DBH* | $\beta_1$ | -0.0306 | 0.0019 ** | - | - | - | - |
| | $BAL_{stand}$ | $\beta_2$ | -0.0969 | 1.02e-05 *** | -0.0999 | 0.010019 * | -0.1082 | 0.0012 ** |
| | *N* | $\beta_3$ | -0.0075 | 0.0046 ** | - | - | - | - |
| | *Thua Thien Hue* | $\alpha_k$ | 2.8502 | 5.10e-12 *** | 2.7961 | 0.000130 *** | 2.9610 | 1.73e-06 *** |
| | *Binh Dinh* | | 2.8215 | 2.68e-14 *** | 2.7672 | 2.39e-05 *** | 2.9166 | 1.33e-07 *** |
| | *Khanh Hoa* | | 1.6269 | 3.19e-09 *** | 1.6007 | 0.000901 *** | 1.7165 | 2.99e-05 *** |
| | AIC | | | 1684.3 | | NA | | 1457.9 |
| | Deviance | | | 816.81 | | 831.72 | | 322.56 |
| All important tree species | Intercept | $\beta_0$ | -0.1060 | 0.8359 | -1.2507 | 0.0033 ** | 0.1144 | 0.8782 |
| | *DBH* | $\beta_1$ | -0.0433 | 0.0004 *** | -0.0429 | 0.0280 * | -0.0396 | 0.0287 * |
| | $BAL_{stand}$ | $\beta_2$ | -0.0718 | 0.0095 ** | - | - | -0.0910 | 0.0245 * |
| | *Thua Thien Hue* | $\alpha_k$ | 2.6382 | 3.24e-07 *** | 1.3537 | 2.24e-09 *** | 3.0239 | 5.53e-05 *** |
| | *Binh Dinh* | | 2.4632 | 1.05e-07 *** | 1.3239 | 6.41e-09 *** | 2.7783 | 3.51e-05 *** |
| | *Khanh Hoa* | | 1.7083 | 8.98e-07 *** | 0.8861 | 0.0002 *** | 1.9593 | 9.33e-05 *** |
| | AIC | | | 1389.9 | | NA | | 1233.6 |
| | Deviance | | | 700.32 | | 706.97 | | 319.22 |
| *S. wightianum* | Intercept | $\beta_0$ | -2.3011 | 0.0512. | -2.3011 | 0.0615. | -2.2846 | 0.0566. |
| | $BAL_{stand}$ | $\beta_1$ | -0.1859 | 0.0059** | -0.1859 | 0.0081** | -0.1869 | 0.0063** |
| | *Thua Thien Hue* | $\alpha_k$ | 5.4731 | 6.58e-05*** | 5.4731 | 0.0002*** | 5.4926 | 8.15e-05*** |
| | *Binh Dinh* | | 4.7208 | 0.0002*** | 4.7208 | 0.0004*** | 4.7376 | 0.0002*** |
| | *Khanh Hoa* | | 4.7277 | 8.96e-07*** | 4.7277 | 3.48e-06*** | 4.7409 | 1.21e-06*** |
| | AIC | | | 375.95 | | NA | | 377.85 |
| | Deviance | | | 216.95 | | 216.95 | | 210.84 |
| *D. sylvatica* | Intercept | $\beta_0$ | -6.5974 | 6.72e-11*** | -6.310 | 6.69e-08*** | -6.3090 | 4.12e-10*** |
| | $BAL_{subplot}$ | $\beta_1$ | 0.4233 | 0.0384* | - | - | - | - |
| | *Binh Dinh* | $\alpha_k$ | 2.8643 | 0.0054** | 3.135 | 0.0071** | 3.1370 | 0.0026** |
| | *Khanh Hoa* | | 2.9082 | 0.0047** | 3.135 | 0.0071** | 3.1280 | 0.0027** |
| | AIC | | 235.37 | | | NA | | 231.26 |
| | Deviance | | 152.10 | | | 155.85 | | 106.08 |
| *N. melliferum* | Intercept | $\beta_0$ | -4.1127 | < 2e-16*** | -4.1127 | < 2e-16*** | -4.1101 | < 2e-16*** |
| | AIC | | | 176.37 | | NA | | 176.36 |
| | Deviance | | | 124.77 | | 124.77 | | 98.79 |

From the Negative Binomial model output, the coefficient of $BAL_{stand}$ was always negative (all species, all important species, and *S. wightianum*) denoting that the number of dead trees declined as the stand basal area became larger, or 0 (i.e. nonsignificant) for the other two species (Table 2). Similarly, the number of dead trees of all important species decreased with an increasing *DBH*, indicating a higher number of dead trees among small, as opposed to larger trees. This number should rise in age-related senescence as the tree becomes older; however, the data in our model did not demonstate this relationship. The number of dead trees for the two groups and *S. wightianum* was found to be the highest in Thua Thien Hue in comparison with Ha Tinh, Binh Dinh, and Khanh Hoa. Similarly, *D. sylvatica*, which appeared in three locations, had a much larger number of dead trees in Binh Dinh and Khanh Hoa than in Ha Tinh.

As mentioned above, for dealing with different time intervals of each plot in each province, adding an offset variable (as log(*time*)) for the model. To avoid plotting two lines in each location, we calculate the annual number of recruits per province. Thus, the annual number of dead trees can be estimated using the Negative Binomial GLM for all species, for all important species per province and for four important species occurring on all plots in three/four locations are:

For all species (with $k = 2.487$)
Ha Tinh: $\log(\mu_i/time_i) = 0.4251 - 0.1082BAL_{standi}$ (12)
Thua Thien Hue: $\log(\mu_i/time_i) = 0.4251 - 0.1082BAL_{standi} + 2.9610$ (13)
Binh Dinh: $\log(\mu_i/time_i) = 0.4251 - 0.1082BAL_{standi} + 2.9166$ (14)
Khanh Hoa: $\log(\mu_i/time_i) = 0.4251 - 0.1082BAL_{standi} + 1.7165$ (15)
For all important species (with $k = 1.993$):
Ha Tinh: $\log(\mu_i/time_i) = 0.1144 - 0.0396DBH_i - 0.0910BAL_{standi}$ (16)
Thua Thien Hue:
$\log(\mu_i/time_i) = 0.1144 - 0.0396DBH_i - 0.0910BAL_{standi} + 3.0239$ (17)
Binh Dinh: $\log(\mu_i/time_i) = 0.1144 - 0.0396DBH_i - 0.0910BAL_{standi} + 2.7783$ (18)
Khanh Hoa: $\log(\mu_i/time_i) = 0.1144 - 0.0396DBH_i - 0.0910BAL_{standi} + 1.9593$ (19)
For *S. wightianum* ($k = 13.786$):
Ha Tinh: $\log(\mu_i/time_i) = -2.2846 - 0.1869BAL_{standi}$ (20)
Thua Thien Hue: $\log(\mu_i/time_i) = -2.2846 - 0.1869BAL_{standi} + 5.4926$ (21)
Binh Dinh: $\log(\mu_i/time_i) = -2.2846 - 0.1869BAL_{standi} + 4.7376$ (22)
Khanh Hoa: $\log(\mu_i/time_i) = -2.2846 - 0.1869BAL_{standi} + 4.7409$ (23)
For *D. sylvatica* ($k = 0.667$):
Ha Tinh: $\log(\mu_i/time_i) = -6.3090$ (24)
Binh Dinh: $\log(\mu_i/time_i) = -6.3090 + 3.1370$ (25)
Khanh Hoa: $\log(\mu_i/time_i) = -6.3090 + 3.1280$ (26)
and for *N. melliferum* ($k = 0.772$): $\log(\mu_i/time_i) = -4.1101$ (27)

### 3.2. Results of the GLMM

The significant negative effects of the *DBH* (for the dead trees of all important species) and $BAL_{stand}$ (for the dead trees of two species groups) as predicted by the Negative Binomial GLM became insignificant under the Negative Binomial GLMM, leading to removal of those variables from the model. Thus, for all species and all important species, the respective fixed effects models (7) and (8), were compared to the following mixed effects models:

$$\log(\mu_{ij}) = (\beta_0 + b_j) + \alpha_k + \log(time_{ij}) \tag{28}$$

$$\log(\mu_{ij}) = (\beta_0 + b_j) + (\alpha_k + b_{kj}) + \log(time_{ij}) \tag{29}$$

$$\log(\mu_{ij}) = \beta_0 + (\alpha_k + b_{kj}) + \log(time_{ij}) \tag{30}$$

For *S. wightianum*, the fixed effects model (3.3) was compared to four mixed effects models:

$$\log(\mu_{ij}) = (\beta_0 + b_j) + \beta_1 BAL_{standij} + \alpha_k + \log(time_{ij}) \tag{31}$$

$$\log(\mu_{ij}) = \beta_0 + (\beta_1 + b_{1j}) BAL_{standij} + \alpha_k + \log(time_{ij}) \tag{32}$$

$$\log(\mu_{ij}) = (\beta_0 + b_j) + \beta_1 BAL_{standij} + (\alpha_k + b_{kj}) + \log(time_{ij}) \tag{33}$$

$$\log(\mu_{ij}) = \beta_0 + \beta_1 BAL_{standij} + (\alpha_k + b_{kj}) + \log(time_{ij}) \tag{34}$$

While for *D. sylvatica*, the fixed effects model (10), along with candidate mixed effects models (28), (29), and (30) were assessed, and for *N. melliferum*, the fixed effects model (11) was compared to only one mixed effects model (35):

$$\log(\mu_{ij}) = (\beta_0 + b_j) + \log(time_{ij}) \tag{35}$$

Pearson's $\chi^2$ values of fixed and mixed models for predicting the number of dead trees across all species, all important species, one important species (*S. wightianum*) occurring in four, and two others (*D. sylvatica*, *N. melliferum*) found in three locations are presented in table 3.

**Table 3. A comparison of Pearson's $\chi^2$ values between the fixed effects model and the mixed effects models. Selected models are bolded**

| Objects | *n* | Equation | Model specification | Pearson's $\chi^2$ Negative Binomial GLMM |
|---|---|---|---|---|
| All tree species | 300 | 3.1 | Fixed effects model (GLM) | 330.35 |
| | | 3.22 | FM + plot intercept | **289.50** |
| | | 3.23 | FM + plot intercept + plot slope (*provincecode*) | 289.55 |
| | | 3.24 | FM + plot slope (*provincecode*) | 289.55 |
| All important tree species | 300 | 3.2 | Fixed effects model (GLM) | 324.64 |
| | | 3.22 | FM + plot intercept | **289.59** |
| | | 3.23 | FM + plot intercept + plot slope (*provincecode*) | 289.68 |
| | | 3.24 | FM + plot slope (*provincecode*) | 289.68 |
| *S. wightianum* | 300 | 3.3 | Fixed effects model (GLM) | 312.14 |
| | | 3.25 | FM + plot intercept | 300.00 |
| | | 3.26 | FM + plot slope ($BAL_{stand}$) | 300.00 |
| | | 3.27 | FM + plot intercept + plot slope (*provincecode*) | 298.13 |
| | | 3.28 | FM + plot slope (*provincecode*) | **298.13** |
| *D. sylvatica* | 225 | 3.4 | Fixed effects model (GLM) | **216.80** |
| | | 3.22 | FM + plot intercept | 225.00 |
| | | 3.23 | FM + plot intercept + plot slope (*provincecode*) | 224.99 |
| | | 3.24 | FM + plot slope (*provincecode*) | 224.99 |
| *N. melliferum* | 225 | 3.5 | Fixed effects model (GLM) | 235.47 |
| | | 3.29 | FM + plot intercept | **223.50** |

Table 3 presents a comparison of selected models based on Pearson's $\chi^2$ values. Here, we see that the Pearson's $\chi^2$ statistic for the fixed model is significantly larger than that of mixed effects model, with the exception of one important species appearing in three places (*D. sylvatica*). The Negative Binomial GLMM performed better than the Negative Binomial GLM. In similar fashion to the recruitment model, the mixed models with random intercept/random slope (*provincecode*) was not different from the mixed models with random slope (*provincecode*) effects. Therefore, the mixed model with a random slope was chosen for *S. wightianum* for it was simpler, while the mixed model with random intercept was selected as the equation for the direct prediction of dead trees across each of the two species groups, and for *N. melliferum*, because it had the smallest Pearson's $\chi^2$ value (Table 3). For *D. sylvatica*, the fixed model using only the *provincecode* as a predictor (Equation 3.4) was used.

The summary statistics for the parameter estimations, standard deviation errors, and the *p*-values for the Negative Binomial GLMM are reported in table 4. In general, for the two species groups, the number of dead trees in Thua Thien Hue and Binh Dinh was much greater than that in Ha Tinh, with the single exception of Khanh Hoa, where there was no significant difference when compared with Ha Tinh ($p > 0.05$). For *S. wightianum*, in Thua Thien Hue, Binh Dinh, and Khanh Hoa were significantly higher numbers of mortality in comparison with Ha Tinh.

For *N. melliferum*, there was no difference in the number of both recruits and dead trees in three provinces Ha Tinh, Binh Dinh, and Khanh Hoa.

The variance component of the random plot effects in Table 4 was rather small (from 7.22% to 25.77%); however, the random effect demonstrated evidence of unexplained variation at the plot level and provided a suitable adjustment for dispersion (the overdispersion parameter was more or less 1).

## 4. DISCUSSION

The Negative Binomial regression for the mortality model expressed in this paper used variables $BAL_{stand}$ and *provinceocde* for all species, and *DBH*, $BAL_{stand}$ and *provincecode* for all important species as predictors in predicting mortality. The *DBH* had a negative sign, resulting in the high mortality of small diameter trees and suggesting that suppressed trees are more likely to be eliminated from stand level competition (Adame *et al.*, 2010); the negative *DBH* coefficient also indicated that stand mortality is more likely in forest stands with many small trees as compared to those with larger trees (Juknys *et al*., 2006). This result was supported by Zhang *et al.* (2014) who likewise found that stand mortality was negatively associated with a stand arithmetic mean diameter among Chinese pines (*Pinus tabulaeformis*).

The stand basal area was suggested as a measure of two-sided competition that can take into account both the vertical competition for light and the horizontal competition for rooting space, water, and nutrients (Yang *et al.*, 2003). This indicator is a good measure of stand crowding because it accounts for both tree size and density. Trees in a stand with a larger basal area will experience more competition than those in another stand with a smaller stand basal area (Yang *et al.*, 2003). The number of dead trees should grow along with the increase in the stand basal area as a result of competition pressure. In this study, however, the negative coefficient of the stand basal area demonstrated that with an increasing stand basal area, the number of dead trees decreased. This may imply that inter-specific competition does not cause tree mortality for these stands. Another study from Bravo *et al.* (2001) found that stand basal area was an insignificant predictor of Douglas-fir mortality across a range of stands.

Plot level random effects on mortality models can address some of the unexplained variation in these processes due to unobserved plot level variables, which included topography, soil, microclimate, nutrients, and moisture (Ma *et al.*, 2013).

**Table 4. Parameter estimates for the Negative Binomial GLMM across all species, all important species, and important species occurring in four or three provinces**

| Objects | Equation | Fixed effects | | | | | Variance components | | | % variation explained by the plot | Overdispersion parameter |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Variables | Parameters | Parameter estimates | Std-error | Pr(>\|z\|) | $\sigma^2_{\text{ran-in}}$ | $\sigma^2_{\text{ran-slop}}$ | $\sigma^2_{\text{res}}$ | | |
| All species | 5.56 | Intercept | $\beta_0$ | -1.3576 | 0.2586 | 0.0000 | 0.1661 | - | 0.8138 | 16.95 | 0.9814 |
| | | *Thua Thien Hue* | $\alpha_k$ | 1.0295 | 0.3600 | 0.0212 | | | | | |
| | | *Binh Dinh* | | 1.2125 | 0.3595 | 0.0097 | | | | | |
| | | *Khanh Hoa* | | 0.4374 | 0.3627 | 0.2623 | | | | | |
| All important species | 5.56 | Intercept | $\beta_0$ | -2.1680 | 0.3048 | 0.0000 | 0.2200 | - | 0.7779 | 22.04 | 0.9817 |
| | | *Thua Thien Hue* | $\alpha_k$ | 1.4325 | 0.4192 | 0.0091 | | | | | |
| | | *Binh Dinh* | | 1.4350 | 0.4191 | 0.0090 | | | | | |
| | | *Khanh Hoa* | | 0.9419 | 0.4218 | 0.0560 | | | | | |
| *S. wightianum* | 5.62 | Intercept | $\beta_0$ | -2.5911 | 1.2212 | 0.0347 | - | 0.3476 | 1.0013 | 25.77 | 1.0141 |
| | | *BALstand* | $\beta_1$ | -0.1670 | 0.0696 | 0.0475 | | | | | |
| | | *Thua Thien Hue* | $\alpha_k$ | 5.1282 | 1.4038 | 0.0084 | | | | | |
| | | *Binh Dinh* | | 4.2865 | 1.3349 | 0.0152 | | | | | |
| | | *Khanh Hoa* | | 4.5158 | 0.9926 | 0.0026 | | | | | |
| *N. melliferum* | 5.63 | Intercept | $\beta_0$ | -4.1223 | 0.2257 | 0.0000 | 0.0762 | - | 0.9793 | 7.22 | 1.0022 |

## 5. CONCLUSION

For this study, overdispersion becomes an issue as a result of the huge number of zero counts. Because it can affect the regression parameters, overdispersion is dealt with here by using a generalized linear mixed model, treating a plot factor as a random effect and integrating the evoked overdispersion by this factor into the model. The Negative Binomial GLMM therefore appeared to be a suitable model due to its ability to capture overdispersion and within-plot correlation. This analysis illustrates that appropriate statistical models are effective in tackling the challenge of modeling mortality and the association of dead trees with data that has a high frequency of zero captures.

## REFERENCES

1. Adame, P., Del Río, M., Canellas, I. (2010). Ingrowth model for Pyrenean oak stands in northwestern Spain using continuous forest inventory data. *European Journal of Forest Research* 129, 669-678.
2. Affleck, D.L.R. (2006). Poisson mixture models for regression analysis of stand-level mortality. *Canadian Journal of Forest Research* 36, 2994-3006.
3. Álvarez González, J.G., Castedo Dorado, F., Ruíz González, A.D., López Sánchez, C.A., Gadow, K. (2004). A two-step mortality model for even-aged stands of *Pinus radiate* D. Don in Galicia (Northwestern Spain). *Annals of Forest Science* 61, 439-448.
4. Bravo, F., Hann, D.W., Maguire, D.A. (2001). Impact of competitor species composition on predicting diameter growth and survival rates of Douglas-fir trees in southwestern Oregon. *Canadian Journal of Forest Research* 31, 2237-2247.
5. Carey, E.V., Brown, S., Gillespie, A.J.R., Lugo, A.E., 1994. Tree mortality in mature lowland tropical moist and tropical lower montane moist forests of Venezuela. *Biotropica* 26, 255-265.
6. Diéguez-Aranda, U., Castedo Dorado, F., Álvarez González, J.G., Rodríguez-Soalleiro. R. (2005). Modelling mortality of Scots pine (Pinus sylvestris L.) plantations in the northwest of Spain. *European Journal of Forest Research* 124, 143-153.
7. Eid, T., Tuhus, E. (2001). Models for individual tree mortality in Norway. *Forest Ecology and Management* 154, 69-84.
8. Fridman, J., Stahl, G. (2004). A three-step approach for modelling tree mortality in Swedish Forests. *Scandinavian Journal of Forest Research* 16, 455-466.
9. Hasenauer, H., Merkl, D., Weingartner, M. (2001). Estimating tree mortality of Norway spruce stands with neutral networks. *Advances in Environmental Research* 5, 405-414.
10. Hu, M., Pavlicova, M., Nunes, E.V. (2011). Zero-inflated and Hurdle models of count data with extra zeros: Examples from an HIV-risk reduction intervention trial. *Am J Drug Alcohol Abuse* 37, 367-375.
11. Juknys, R., Vencloviene, J., Jurkonist, N., Bartkevicius, E., Sepetiene, J. (2006). Relation between individual tree mortality and tree characteristics in a polluted and non-polluted environment. *Environment Monitoring and Assessment* 121, 519-542.
12. Ma, Z., Peng, C., Li, W., Zhu, Q., Wang, W., Song, X., Liu, J. (2013). Modeling individual tree mortality rates using marginal and random effects regression models. *Natural Resource Modeling* 26, 131-153.
13. McCarthy, J. (2001). Gap dynamics of forest trees: a review with particular attention to boreal forests. *Environmental Reviews* 9, 1-59.
14. Meng, Q., Cieszewski, C.J., Lowe, R.C., Zasada, M. (2003). A three-step approach to model tree mortality in the State of Georgia. *Fifth Annual Forest Inventory and Analysis Symposium*, United States Department of Agriculture. Forest Service, New Orleans, Louisiana, pp. 91-96.
15. Monserud, R.A., Sterba, H. (1999). Modeling individual tree mortality for Austrian forest species. *Forest Ecology and Management* 113, 109-123.
16. Oliver, C.D., Larson, B.C. (1996). Forest stand dynamics: Update edition. John Wiley & Sons, Inc., New York, 540 pp.
17. Rüger, N., Huth, A., Hubbell, S.P., Condit, R. (2011). Determinants of mortality across a tropical lowland rainforest community. Oikos 120, 1047-1056.
18. Vanclay, J.K. (1992). Modelling regeneration and recruitment in a tropical rain forest. *Canadian Journal of Forest Research* 22, 1235-1248.
19. Yang, Y., Titus, S.J., Huang, S., 2003. Modeling individual tree mortality for white spruce in Alberta. *Ecological Modelling* 163, 209-222.
20. Zhang, X., Lei, Y., Cai, D., Liu, X. (2014). Modeling stand mortality using Poisson mixture models with mixed-effects. iForest-Biogeosciences and Forestry.

# XÂY DỰNG MÔ HÌNH CÂY CHẾT CHO RỪNG TỰ NHIÊN TRẠNG THÁI III Ở 4 TỈNH MIỀN TRUNG VIỆT NAM

**Cao Thị Thu Hiền**
*Trường Đại học Lâm nghiệp*

## SUMMARY

Cây chết trong nghiên cứu này là những cây chết đứng giữa 2 chu kỳ đo. Số liệu từ 300 phân ô của 12 ô đo đếm (ODD) được dùng để xây dựng mô hình cây chết. Biến phụ thuộc là số cây chết/phân ô. Kết quả cho thấy có thể dùng mô hình tuyến tính tổng quát và mô hình tuyến tính tổng quát hỗn hợp để mô phỏng vì hai mô hình này giải quyết được vấn đề phân tán của số liệu. Với mô hình tuyến tính tổng quát (GLM), phương trình Negative Binomial GLM là phù hợp nhất để dựa đoán số cây chết cho 3 nhóm. Đường kính, tiết diện ngang lâm phần và biến phân nhóm "*tỉnh*" là các biến có ảnh hưởng rõ ràng nhất tới mô hình dự đoán cây chết. Từ biến phân nhóm cho thấy số cây chết ở các tỉnh là khác nhau, số cây chết của nhóm 1 và nhóm 2 nhiều nhất ở Thừa Thiên Huế và ít nhất là ở Hà Tĩnh. Với mô hình tuyến tính tổng quát hỗn hợp (GLMM), hàm Negative Binomial GLMM đã loại bỏ được những giá trị phân tán. GLMM với hiệu ứng ngẫu nhiên là ODD cho hệ số tự do được lựa chọn để dự đoán số cây chết cho nhóm 1, nhóm 2 và loài Trường vải. GLMM với hiệu ứng ngẫu nhiên là ODD cho hệ số hồi quy được lựa chọn để dự đoán số cây chết cho Trâm trắng.

**Từ khóa: Mô hình cây chết, Negative Binomial GLM, phương trình tuyến tính tổng quát, phương trình tuyến tính tổng quát hỗn hợp, rừng mưa nhiệt đới.**

**Received : 14/5/2018**
**Revised : 10/9/2018**
**Accepted : 18/9/2018**