Statistical Analysis of Tree-Forest Damage by Snow and Wind: Logistic Regression Model for Tree damage and Cox Regression for Tree Survival

Ken-ichi Kamo¹, Masashi Konoshima², Atsushi Yoshimoto³*

Abstract: Forest stands and individual trees are often devastated by natural disasters such as typhoons and heavy snowfall in Japan, resulting in significant economic losses to the forestry sector. Our objective is to identify key risk factors that affect the degree of damage. We apply two types of statistical approach: one is, a logistic regression model to snow damage data to investigate if there is any geographical element affecting the degree of damage at the stand level, and the other is, a survival analysis on tree failure data for factors affecting the degree of damage at the individual tree level. A logistic regression analysis revealed that the risk probability of snow damage is higher on older and thin stands. The analysis also indicates taking advantage of certain geographic conditions to reduce wind burden could decrease the degree of damage. A Cox regression analysis showed that tree age, diameter at breast height, and species were key factors that influenced the degree of tree failure. Specifying risk factors throughout statistical modeling helps to provide a comprehensive, systematic, and objective method to assess risk in forest management.

Keywords: snow damage, wind damage, logistic regression, Cox regression

1. Introduction

Forests and trees are often devastated by natural disasters such as typhoons and heavy snowfall in conjunction with strong winds. Sugi (Cryptomeria japonica) and hinoki (Chamaecyparis obtusa), the most commercially produced species in Japan, have been damaged by events of this nature throughout Japan, resulting in significant economic loss and ecological damage. Avoiding or reducing these risks has been an important forest management objective. If we can identify physical and/or geographical characteristics affecting the degree of devastation, it may become possible to reduce the risk of damage through an appropriate management strategy (Nykänen et al., 1997; Jalkanen and Mattila, 2000). However, the process of quantifying and integrating risk into (long-term) forest management strategy is not well developed in forestry (Hanewinkel et al., 2011).

Our objective is to identify key risk factors and quantify the associated risk of damage caused by snowfall at both the stand and individual tree levels. Regression models have been most widely applied when several factors are thought to influence objective observations. Logistic regression has been used to evaluate the risk of natural disturbance and identify risk factors in forest management (Jalkanen and Mattila, 2000; Kamo et al., 2008; Kulla and Marusák, 2011). Jalkanen and Mattila (2000) for example, applied a logistic regression model to predict the susceptibility of forest stands to both wind and snow damage. We apply logistic and Cox regression models to evaluate the risk of damage. In the logistic regression model, we regard snow load damage as a one-time, instantaneous event in some parts of the forest stand.

Snow damage can be viewed as a collection of structurally compromised trees in a forest stand where each tree is subjected to the pressure of accumulated snow until it breaks. Since this phenomenon is similar to survival, we apply survival analysis to describe tree failure events. We use the Cox regression model (Cox, 1984) for survival analysis of tree failure experimental data. Survival analysis has been widely used in the medical and engineering fields. An event (or end point) can represent “death” in the medical fields and “failure” in the engineering arena. For our data, we treat “tree failure” as an event. Survival analysis is unique in the sense that it considers a case where termination occurs before the event (survival time) is observed (censoring of observations) (Collett, 1994; Allison, 1995). The Cox regression model is categorized in a class of statistical methods that investigate the occurrence and timing of events. At the most basic level, this type of analysis estimates a survival function for the probability that an event of interest will occur after some point in time, where survival time is a function of various factors. Unlike logistic regression models, the Cox

Received Jan. 14, 2016; Accepted Apr. 8, 2016
¹Sapporo Medical University, Japan
²University of the Ryukyus, Japan
³Institute of Statistical Mathematics, Japan
*Corresponding Author:yoshimoa@ism.ac.jp
model is not often applied to the risk of natural disturbance in forestry (Cyr et al., 2007; Jansen et al., 2008). There are few studies that have used the model to address factors of tree failure caused by natural disturbances such as wind and snow.

In both analyses, a model selection procedure is used to select the most suitable model. Our procedure uses both Akaike’s information criterion (AIC) (Akaike, 1973) and Baysian information criterion (BIC) (Schwarz, 1978). When both AIC and BIC lead to the selection of two different models, the deviance is used to select the best model and the best combination of explanatory variables (McCullagh and Nelder, 1989).

This paper is organized as follows: In Section 2, we describe the data and the methods used in this study (i.e. logistic regression analysis and Cox regression analysis). In Section 3, we show the results obtained by each model, followed by discussion of the results in Section 4.

2. Materials and Methods

2.1. Data for snow damage

Snow damage data was collected for a sugi (<i>Cryptomeria japonica</i>) forest in Toyama Prefecture, Japan. Snow damage occurs when snow loading on leaves and branches cause the trunk to bend, break, or leads to complete uprooting of the tree. Figure 1 shows a trunk-bending scenario in Toyama, Japan. The degree of snow damage depends on several factors, including forest stand and geographic characteristics (Cremer et al., 1983; Kato and Nakatani, 2000; Jalkanen and Mattila, 2000).

![Figure 1. Snow damage–stem compression from snow loading–in a sugi forest stand.](image)

In our analysis candidate variables include geographic elements such as slope aspect, and forest stand attributes such as stand age and average tree height. The data was collected from 47 different sample plots (20m x 20m) in 2004. In total, 1,761 trees were sampled, of which 599 were damaged. Thus, we estimated an overall damage probability for the region as 34%. We observed 16 factors that might have affected the risk probability. These include: 1) Forest stand age (“Age”), 2) Average tree DBH (“DBH”) (diameter at breast height), 3) Average tree height (“Height”), 4) Ratio of height to DBH (“H/D ratio”), 5) Forest stand density (“Density”) (number of trees per ha), 6) Species (“Species”) (dummy variables were used to represent different species), 7) Forest stand volume (“Volume”), 8) Altitude (“Altitude”), 9) Slope gradient (“Slope”), 10) Index of snow water flow (“ISWF”) (calculated by numerical examination), 11) Topographic wetness index (“TWT”), 12) Plain curvature (“Plain curvature”) (curvature along with horizontal direction, see Moore et al., 1993), 13) Profile curvature (“Profile curvature”) (same definition as plain curvature, but along the vertical), 14) Over-ground openness (“Over openness”) (or degree of outlook), 15)
Under-ground openness ("Under openness") (or degree of interruption by the earth), and 16) Slope aspect ("North", "West", "South", "East").

Note that the topographic wetness index is defined by \( \log \frac{a}{\tan b} \), where \( a \) is contributing area and \( b \) is slope gradient (Beven, 1997). In the analysis, the slope aspect variable was constructed by combining four dummy variables to identify eight aspects. The full combination is shown in Table 1. For example, "north" is identified by the set of (east, west, south, north) = (0, 0, 0, 1), while "northwest" is identified by (1, 0, 0, 1). These four dummy variables used for slope aspect were incorporated into the model to represent all eight aspects. Variables related to 1) – 7) are those elements for forest stand characteristics, while those for 8) – 16) are for geographical characteristics. Variables 10) – 15) describe the degree of forest stand surface convexity from several viewpoints. All variables are candidate factors (Kamo et al., 2008).

Table 1. Combination of dummy variables to express slope aspect.

<table>
<thead>
<tr>
<th>Variables</th>
<th>North</th>
<th>Northeast</th>
<th>East</th>
<th>Southeast</th>
<th>South</th>
<th>Southwest</th>
<th>West</th>
<th>Northwest</th>
</tr>
</thead>
<tbody>
<tr>
<td>East</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>West</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>South</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>North</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

2.2. Data for the tree-pulling experiment

In the tree-pulling experiment, a load was placed on standing trees to measure the force necessary to cause tree failure (Figure 2) (See Kato (2001) for a detailed explanation of this tree-pulling experiment). Generally speaking, there are two types of tree failure—stem bending or breakage and uprooting (Tanaka and Yagisawa, 2009).

Figure 2. Photos from the tree-pulling experiment in a sugi forest demonstrating uprooting under weight burden.
Figure 3. Relationship between “Moment” (point of failure) and other explanatory variables.
The tree-pulling experiments were conducted in 1982, 1984, and 2007 in Toyama Prefecture. Throughout the experiments, six types of data were collected from twenty-eight sample trees. These included: 1) Moment of tree failure, 2) Tree age, 3) DBH (diameter at breast height), 4) Tree height, 5) Tree volume, and 6) Species. Two variants of sugi were used in these experiments—Bokasugi and Tateyama-sugi, which are species commonly planted in this region. The “Moment” of tree failure directly influences the degree of tree failure, so this was regarded as a chief candidate for a response variable. Figure 3 shows the relationship between the “Moment” variable and the other four variables—“Tree age”, “DBH”, “Tree height”, and “Tree volume”—with “Moment” set on the vertical axis. The marks “●” and “□” denote the “Species” variable for the Bokasugi and Tateyama-sugi, respectively. Because these experiments were conducted in plantation forests, the value of “Tree age” was either 14, 21, 23, or 28. As Figure 3 depicts, “DBH”, “Tree height”, and “Tree volume” appear to have a positive linear relation to “Moment” at the time of tree failure.

2.3. Logistic regression model

The logistic regression model assumes the frequency of damage occurrence is distributed according to the binomial distribution. That is,

\[ y \sim \text{Bin}(n, p), \]

where \( n \) is the number of trees, \( p \) is the probability of damage occurrence and \( y \) is the number of damaged trees. In the logistic regression model, \( \logit p = \log \frac{p}{1-p} \) is assumed to be a linear function of explanatory variables. Let \( y_i, n_i, \) and \( p_i \) be those values for the \( i^{th} \) sample tree \((i = 1, 2, \ldots, m)\), and \( x_i \) be the explanatory variable vector in the same sample. Then, \( p_i \) is expressed as follows:

\[ \logit p_i = x_i' \beta \]

Maximum likelihood method of \( \beta \) is obtained by

\[ \hat{\beta} = \arg \max_{\beta} L(\beta) = \arg \max_{\beta} \sum_{i=1}^{m} \{ y_i x_i' \beta - n_i \log(1 + \exp(x_i' \beta)) \} \]

so that we have

\[ \hat{p}_i = \frac{\exp(x_i' \hat{\beta})}{1 + \exp(x_i' \hat{\beta})} \]

The model selection procedure is designed to find the optimal subset of explanatory variables. The most commonly used criterion for selecting variables is Akaike’s information criterion (AIC) (Akaike, 1973). AIC estimates a risk function based on predicted Kullback-Leibler values (Kullback and Leibler, 1951). AIC can be interpreted as a measure of the distance between the true model and the candidate model. Thus, the model with minimum AIC is considered the best fit. In logistic regression, AIC can be defined as

\[ \text{AIC} = -2L(\beta) + 2k \]

where \( k \) is the number of unknown parameters. AIC is often interpreted as the sum of a model’s “goodness of fit” \((-2L(\hat{\beta}))\) and a “penalty” for its complexity \((2k)\). Schwarz (1978) later proposed Bayesian information criterion (BIC) as

\[ \text{BIC} = -2L(\beta) + k \log n \]

where \( n = \sum_{i=1}^{m} n_i \). Although AIC estimates Kullback-Leibler risk (Kullback and Leibler, 1951), BIC is based on a Bayesian framework. Like AIC, the model with minimum BIC is considered the best model.

When different models are selected by AIC and BIC, the deviance, \( D \), (McCullagh and Nelder, 1989) is used to select the single best model, defined by

\[ D = 2 \sum_{i=1}^{m} \left\{ y_i \log \left( \frac{y_i}{n_i \hat{p}_i} \right) + (n_i - y_i) \log \left( \frac{n_i - y_i}{n_i - n_i \hat{p}_i} \right) \right\} \]
The deviance is defined as a likelihood ratio between the candidate and the full model, which follows a chi-squared distribution with the degree of freedom of \((\text{sample size}) - (\text{number of explanatory variables}) - 1\) under the hypothesis that the candidate is correct. In our model the number of trees damaged by snow is assumed to follow a binomial distribution, with the total number of trees and the probability, \(p\), of snow damage, where the expected value of the probability transformed by the logit function, \(\log\frac{p}{1-p}\) for \(p\), is expressed by a linear function of explanatory variables. Because the resultant \(D\) follows the chi-squared distribution with \((m - k - 1)\) degree of freedom, it is used to determine the best model. A detailed discussion of logistic regression models can be found in Cameron and Trivedi (1998).

2.4. Cox regression model

When applying the Cox regression model in survival analysis, we often use the hazard rate, which is defined as the probability of occurrence per unit time, given that the event has “survived” up to some point in time. This is often called the Cox proportional hazard model. In this hazard model, the hazard rate becomes a response variable of a survival function. In our study, we used the Cox proportional hazard model to analyze experimental data on tree failure. Let \(S(t)\) be the survival function for the probability that the case has survived to time \(t\), i.e. a random variable \(T\) exceeds time \(t\), resulting in the following:

\[
S(t) = \Pr(T \geq t)
\]

Here \(T\) is a random variable denoting the time of event occurrence, where \(t\) ranges from 0 to \(\infty\). The survival function monotonically decreases over time with \(S(0) = 1\) and \(S(\infty) = 0\).

The survival function is estimated by the Kaplan-Meier method, where the Kaplan-Meier curve is treated as one estimator of the survival curve. Let \(t_i\) be the time event occurrence at the \(i\)th observation. At this time, let \(n_i\) be the number of “at risk” (or the risk set) and \(d_i\) be the number of “tree failures” at the \(i\)th observation \((i = 1, 2, ..., m)\). Here, “at risk” implies the collection of individual trees that have survived (not suffered failure) to time \(t_i\). Thus, the survival function is estimated by

\[
S(t) = \prod_{i=1}^{m} \frac{n_i - d_i}{n_i}
\]

Throughout this method, we assume the survival function’s value only changes when an event occurs.

In the Cox regression model, the response variable is set to be the hazard function. Here “hazard” implies the immediate potential per unit time for the event to occur, defined by

\[
\lambda(t) = \frac{Pr(t \leq T \leq t + dt \mid T \geq t)}{dt} = \lim_{\Delta t \to 0} \frac{S(t) - S(t + \Delta t)}{\Delta t}
\]

where \(\lambda(t)\) on the left side is the hazard at time \(t\). In contrast to the survival function for “not failing” events, the hazard function focuses on “failing” events. In Cox regression analysis, the following assumption applies for the conditional hazard function:

\[
\lambda(t \mid x) = \lambda_0(t) \exp(\beta'x)
\]

where \(\lambda_0(t)\) is the baseline hazard function, \(\beta\) is an unknown parameter vector, and \(x\) is an explanatory variable vector. Eq. [11] holds the proportional hazard ratio property,

\[
\frac{\lambda(t \mid x_1)}{\lambda(t \mid x_2)} = \exp(\beta'(x_1 - x_2))
\]

This indicates that if there are two types of hazard with different explanatory variables \(x_1\) and \(x_2\), then the hazard ratio is independent from \(t\).

Estimation of unknown parameters is achieved by maximizing the partial likelihood, defined by

\[
L(\beta) = \sum_{i=1}^{m} \beta'x_i - \sum_{i=1}^{m} \log \left( \sum_{j \in R_i} \exp(\beta'x_j) \right)
\]
where $R_i$ is the index group for “at risk”. This likelihood is constructed only by the order of event occurrence, so that its estimator is obtained by

$$\hat{\beta} = \arg \max_{\beta} L(\beta)$$

As in the case of the logistic regression model, AIC is defined by,

$$AIC = -2L(\hat{\beta}) + 2k$$

where $k$ is the number of unknown parameters. Note that $L(\beta)$ is the partial log-likelihood, not log-likelihood. Because of this difference, we cannot define BIC; therefore, we only used AIC in this case.

Table 2. Model selection results.

<table>
<thead>
<tr>
<th>Candidate Variables</th>
<th>The number of Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13 14 15</td>
</tr>
<tr>
<td>DBH</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>Height</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>H/D ratio</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>Density</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>Species</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>Volume</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>Altitude</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>Slope gradient</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>Contributing area</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>Topographic wetness</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>Plain curvature</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>Profile curvature</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>Over openness</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>Under openness</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>Slope aspect*</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>Best model</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
</tbody>
</table>

The notation “0” denotes the variable selected by AIC and BIC. For all candidate models with regard to the number of variables, the same combination of variables is selected as the best by AIC and BIC. “: Slope aspect includes any of North, West, East, South dummy variables.

3. Results

3.1. Results for the logistic regression model

The best model selected by AIC included thirteen variables, while the best model selected by BIC included six—BIC selected a simpler model. Because these “best models” were different, we finalized the selection by analyzing the deviance. The AIC model resulted in a deviance of 180.9, while the BIC model’s deviance was 202—a difference of 21.1 with a corresponding p-value (degrees of freedom = 7) of 0.004. Thus, the additional 7 variables led to significant improvement at a 0.05 level of significance, resulting in selection of the model derived by AIC. Table 2 shows results from the model selection procedure, while Table 3 presents the estimated parameters from the “best” or final model, which follows:

$$\logit p = -1.145 + 0.0293 \times Age - 0.2625 \times DBH + 0.1326 \times Height - 0.0019 \times Density$$
$$- 2.38 \times Species + 0.0021 \times Volume + 0.0102 \times Altitude$$
$$- 0.0003 \times Contributing area + 0.2804 \times Wetness - 18.18 \times Plain curvature$$
$$- 0.0844 \times Over openness + 1.007 \times Under openness - 0.2099 \times North$$
$$+ 1.412 \times East + 0.0348 \times South + 0.162 \times West$$

[16]
Statistical Analysis of Tree-Forest Damage by Snow and Wind

Table 3. AIC-estimated parameters from the “best” model.

<table>
<thead>
<tr>
<th>Selected Variables</th>
<th>Estimated value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-1.1450</td>
<td>0.7438</td>
</tr>
<tr>
<td>Age</td>
<td>0.0293</td>
<td>0.0264</td>
</tr>
<tr>
<td>DBH</td>
<td>-0.2625</td>
<td>0.0002</td>
</tr>
<tr>
<td>Height</td>
<td>0.1326</td>
<td>0.0981</td>
</tr>
<tr>
<td>Density</td>
<td>-0.0019</td>
<td>0.0068</td>
</tr>
<tr>
<td>Species</td>
<td>-2.3800</td>
<td>0.0000</td>
</tr>
<tr>
<td>Volume</td>
<td>0.0021</td>
<td>0.0216</td>
</tr>
<tr>
<td>Altitude</td>
<td>0.0102</td>
<td>0.0000</td>
</tr>
<tr>
<td>Contributing area</td>
<td>-0.0003</td>
<td>0.0010</td>
</tr>
<tr>
<td>Topographic wetness</td>
<td>0.3804</td>
<td>0.0004</td>
</tr>
<tr>
<td>Plain curvature</td>
<td>-18.1800</td>
<td>0.1435</td>
</tr>
<tr>
<td>Over openness</td>
<td>-0.0844</td>
<td>0.0010</td>
</tr>
<tr>
<td>Under openness</td>
<td>1.0070</td>
<td>0.0000</td>
</tr>
<tr>
<td>North</td>
<td>-0.2099</td>
<td>0.3824</td>
</tr>
<tr>
<td>East</td>
<td>1.4120</td>
<td>0.0000</td>
</tr>
<tr>
<td>South</td>
<td>0.0348</td>
<td>0.8353</td>
</tr>
<tr>
<td>West</td>
<td>0.1620</td>
<td>0.4941</td>
</tr>
</tbody>
</table>

3.2. Results for the Cox regression model

In the analysis of the tree-pulling experiment data, we regarded the moment of tree failure as the time of event occurrence to avoid changing the survival function. Figure 4 shows the survival curve estimated by the Kaplan-Meier method. Here the horizontal axis denotes the moment of tree failure. As mentioned earlier, the value changed in the estimated curve only at the point of event occurrence, resulting in a step-shaped curvature rather than a smooth curvature.

There were two variants of sugi species (Boka-sugi and Tateyama-sugi) tested in this experiment and the degree of endurance under load differed between them. Figure 5 shows the estimated survival curve for each variant; Boka-sugi was stronger than Tateyama-sugi. We used a Log-rank test to validate the difference between survival functions. This test was based on the null-hypothesis that there is no difference between survival functions. Although we obtained a p-value of 0.0791, implying we should not reject the null-hypothesis at the 0.05 level, its value was very close to the significance level. Thus, species may very well be an important component of a tree’s vulnerability to failure. Details for the Log-rank test can be found in Kleinbaum and Klein (2005).

We selected the best combination of explanatory variables using a stepwise method that employed the “stepAIC” command in the R statistical package (see Venables and Ripley (2002)). With this method, a new candidate model is constructed by removing one variable from the current candidate model. Then, the AIC values of both the current and new candidate models are compared. The model with smallest AIC becomes the candidate model for the next comparison. This process continues, stepwise, until there is no additional improvement derived.

We began by removing the “Volume” variable from the full model, given an initial candidate model with the explanatory variables “Age”, “DBH”, “Height” and “Species”. The “Height” variable was then removed from this model, resulting in a second candidate model with “Age”, “DBH” and “Species”. We were unable to improve this model by removing additional variables and thus concluded it was the final best model. The estimated parameters from each step and p-value for the final model are shown in Table 4.

The final model is expressed as follows:

$$
\log\frac{\lambda(t \mid x)}{\lambda_0(t)} = -0.299 \times \text{Age} - 0.47 \times \text{DBH} + 1.277 \times \text{Tateyama}
$$

where “Tateyama” is a dummy variable for species with a value of 1 corresponding to Tateyama-sugi. The sign of unknown parameters reflects the effect each has on risk. “Age” and “DBH” reduce the risk at a higher value, while Tateyama-sugi increases risk more than Boka-sugi.
Figure 4. Estimated survival curve for all samples.

Figure 5. Estimated survival curve by species.
Table 4. Estimated parameters from the step-wise model selection procedure.

<table>
<thead>
<tr>
<th>Model Selection</th>
<th>Age</th>
<th>DBH</th>
<th>Height</th>
<th>Volume</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full model selected</td>
<td>-0.318</td>
<td>-0.698</td>
<td>-0.434</td>
<td>6.873</td>
<td>1.136</td>
</tr>
<tr>
<td>1st step model selected</td>
<td>-0.277</td>
<td>-0.384</td>
<td>-0.210</td>
<td>–</td>
<td>1.198</td>
</tr>
<tr>
<td>2nd step model selected</td>
<td>-0.299</td>
<td>-0.47</td>
<td>–</td>
<td>–</td>
<td>1.277</td>
</tr>
<tr>
<td>(p)-value</td>
<td>0.0000022</td>
<td>0.000032</td>
<td>–</td>
<td>–</td>
<td>0.0180</td>
</tr>
</tbody>
</table>

4. Discussion

In this paper we considered two types of risk analysis for natural disturbances on both the forest stand and individual tree level. We applied logistic regression because the number of trees damaged by snow was expected to follow a binomial distribution. Using a stepwise AIC process, we selected an optimal model that included thirteen explanatory variables. The sign of the estimated parameter for each variable denotes whether that element increases or decreases the risk probability. Estimated variables with positive coefficients included forest stand age, tree height, tree volume, altitude, wetness index, under-ground openness, east, south, and west. Forest stand age increases risk because older trees are more susceptible to snow loading. As one would expect, taller trees are less stable under snow loading and thus more likely to be damaged. Altitude increases the risk of snow damage toward the ridge where trees are subject to more snow loading under the same weather conditions. The wetness index measured soil moisture and water availability, a key indicator for the fragility of standing trees. Forest stand volumes with tall trees and small DBH imply a dense stand of thin trees with a high height-diameter ratio. Thus, both stand volume and wetness index contributed to an increase in the risk of snow damage. As for slope aspects, the opposite aspects east and west were shown to increase risk. By examining the significance \(p\)-value of magnitude of estimated coefficients, we found that the east-facing sites had greater risk. Based on the weather record, winds blew from the Southwest continuously when snow damage occurred, indicating east-facing sites received the least amount of wind. Wind affects the risk of snow damage in three different ways: 1) Strong wind can remove large volumes of snow from the canopy, which mitigates the risk of snow damage, 2) strong wind forces snow to accumulate on the canopy, increasing the risk of snow damage, and 3) strong winds following snowfall apply an additional load to trees that are already stressed by heavy snow loading on the canopy and branches (Kato and Zushi, 2006). Because of these mixed effects of wind, slope aspect contributes to the risk level in a very complex way, resulting in an optimal model that includes two opposing variables.

Variables with negative coefficients included “DBH”, “Density”, “Species”, “Contributing area”, “Plain curvature”, “Over openness”, and the slope aspect of “North”. Trees with large DBH can tolerate heavier snow loads. In higher density areas the total risk is shared among many trees, reducing the average risk per tree. It is known that Boka- and Kawaidani-sugi do not tolerate heavy snow loads; thus, the dummy variable was set to 0 for these species. Three geographic variables – contributing area, plain curvature, and over-ground openness – represent concavity of the forest stand surface. There was little wind effect in these areas. Based on our analysis of all these factors, we conclude that stand age, height-diameter ratio (tall, thin trees), and geographic conditions that minimize wind effects all contribute to an increase in the risk probability.

The Cox regression model often leaves the baseline hazard function unspecified, meaning a specific survival function cannot be described. Thus, it can be used to quantify the magnitude of impact from explanatory variables, but the relative hazard of an event cannot be applied for the projection. If the specific form of survival function is available a priori, it can be used to develop a generalized linear regression model for the projection. Another useful characteristic of this regression model, especially in our study, is that it can accept the censor case. The censor case occurs when we do not know the exact survival time. Suppose we would like to analyze the risk of uprooting, but a tree may break before uprooting occurs. In such a case, this particular experiment does not provide useful data unless we consider a censor case. With the Cox regression model we can avoid such information loss and maintain the number of samples.

We specified three risk factors—tree age, DBH, and species—when running the Cox regression analysis. The coefficients’ signs (positive or negative) were consistent with our intuition and ob-
servations. Tree age reduces the risk of failure because the older a tree becomes, the stronger its root system, up to the point when it becomes more susceptible to decay and other defects. Diameter at breast height (DBH) also reduces the risk of tree failure. All things being equal, the larger the DBH, the stronger and more resistant is it to an applied load and therefore the tree remains. Strength increases monotonically with DBH. Lastly, we found that species has a significant impact on the risk of failure. Specifically, we found that Boka-sugi is stronger than Tateyama-sugi. After adjusting for tree age and DBH using the Cox regression model, we were able to identify the effect of species on the risk of failure. Because the regression model estimates the adjusted effect along with other variables, if, for example, Boka-sugi is genetically predisposed to grow a thicker stem than Tateyama, it is difficult to know if increased strength can be attributed to DBH or species. In our study, we recognized the effect of species even after adjusting for DBH.

Two variables—height and volume—were not used in the model. When we performed the step-wise variable selection procedure, volume was removed from the candidate model. Because volume is a measure that represents an entire tree, it may be an insignificant factor compared to other variables (which describe discrete characteristics of a tree). Following volume, the height variable was also removed from the model. In the full model, the sign of height’s coefficient was negative, indicating height reduces risk. One might expect that taller trees are more resistant to uprooting because of their size; however, a tall tree with small DBH is generally less resistant to uprooting. Therefore, the effect of height is not as significant as DBH. If tree size must be introduced, the variable reflecting size in the tree-pulling experiments should have been DBH rather than height. We showed the Cox regression model could be applied to tree failure (due to natural disturbance) by interpreting “the weight-burden required to break a tree” as “survival time”.

Identifying the risk factors, namely physical and/or geographical characteristics affecting the degree of devastation, is generally the first step in risk management (Hanewinkel et al., 2011). Carefully identifying these factors should be a complete objective exercise (Gardiner and Quine, 2000). Risk analysis using statistical models, which we demonstrate here, provides a comprehensive, systematic, and objective method to assess risk in forest management.

Acknowledgement

This work was supported by Open Partnership Joint Projects of JSPS Bilateral Joint Research Projects (2014–2016)

References


