A Bayesian network model to disentangle the effects of stand and climate factors on tree mortality of Chinese fir plantation

Yihang Jiang[1,2], Zhen Wang[1], Hanyue Chen[1], Yuxin Hu[1], Yancheng Qu[1], Sophan Chhin[3], Jianguo Zhang[1], Xiongqing Zhang[1,2,*]
1 State Key Laboratory of Efficient Production of Forest Resources, Key Laboratory of Tree Breeding and Cultivation of the National Forestry and Grassland Administration, Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China.
2 Collaborative Innovation Center of Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing, 210037, P. R. China
3 Division of Forestry and Natural Resources, West Virginia University, 322 Percival Hall, PO Box 6125, Morgantown, West Virginia, 26506, USA

Introduction

Tree mortality refers to the process in which the vitality of trees gradually weakens and eventually dies under the combined effects of environmental interference and stand characteristics, affecting tree regeneration, stand structure, stand productivity, and carbon storage. Stand and climate factors are the main driving factors determining the spatiotemporal patterns of tree mortality. Tree mortality is a complex process that not only needs to consider the impact of various factors, but also consider the uncertainty and link the various long-term effects of the factors on each other. Previous studies mainly focused on the factors of mortality or simple linear relationship between the factors, while ignoring the uncertain relationship between factors.

Machine learning techniques have been widely used and proved to be powerful tools in the field of forest management. Bayesian network (BN) is a useful technique which utilizes the probability calculus together with an underlying graphical structure to provide a theoretical framework for modeling uncertainty. The random forest (RF) algorithm is another nonparametric machine learning algorithm that performs well in classification and regression.

Chinese fir (*Cunninghamia lanceolata* (Lamb.) Hook.), a native species with a capacity for fast growth, is widely planted in subtropical China. Effective management of this widely distributed tree species needs accurate assessment of tree mortality and understanding the complex mechanisms of tree mortality. This study was to model tree mortality in relation to stand and climate factors using Bayesian network, random forest and logistic regression methods and find out which method is better, disentangle the effects of stand and climate factors on tree mortality and analyze the multilateral relations among the stand and climate factors affecting tree mortality.

Materials and Methods

A long-term spacing trial including 60 permanent plots located in Fujian, Jiangxi, Guangxi, and Sichuan provinces in southern China were used to explore tree mortality in relation to stand and climate variables using three models. These plots were established in 1981 or 1982 and measured every two or three years until 2012. Stand variables including relative density (RD), density (N), age, and stand structure (Gini), climate variables including annual precipitation (AP), winter mean minimum temperature (WMMT), and mean annual temperature (MAT).

Logistic equation

$$P_S = \{1 + \exp[-(\alpha_1 + \beta X)]\}^{-q}$$

Where P_s is the probability of tree survival, and X is a series of indicator variables selected for this study, including stand age, stand dominant height, competition intensity, stand structure, and climate factors. α_1 is the intercept of the model, β is the parameter vector containing the intercept, and q is the year between two measurements.

Bayesian network (BN)

The Bayesian network model can be implemented through the "bnlearn" package in R, the probabilistic relationship between nodes is determined by the following equation.

$$\theta_{X_i|\pi_i} = P_\alpha（X_i|\pi_i）$$

X represents each variable, and parameter α quantitatively describes this dependency. Assuming that the parent node set of variable *Xi* in a directed acyclic graph is π_i , then α contains the conditional probability of each variable. θ represents the specific interdependency values of the target attribute and other variables.

By selecting the target node in Netica, the degree of influence of other nodes can be sequentially exported. Bayesian Network model can be interpreted to The Most Probable Explanation (MPE). It can find the combination of multiple causes (node states) that may lead to a certain result from the combination of multiple causes (node states), and the combination with the greatest possibility is the most possible explanation. The most likely combination of causes can be found through the MPE function of Netica.

Random Forest (RF)

Random forest model is a highly flexible machine learning algorithm based on Bagging integrated learning model. In terms of model tuning, RF has two main parameters: mtry (the number of randomly selected split prediction variables at each split) and ntree (the number of decision trees in the random forest). The default value of mtry is usually one-third of all predicted variables, we set ntree equals 1000 as a replacement, and set other parameters to default values. The RF model was implemented with the "randomForest" package in the R software.

Model evaluation

Operating Characteristic (ROC) Curve is used to evaluate the predictive performance of binary models. If a model has a larger AUC value, it indicates that the model is more suitable for data. The widely accepted evaluation rule for AUC value is: 0.9-1 is excellent, indicating that there can be a good distinction between forest survival and withering; 0.8-0.9 is good; 0.7-0.8 is average; 0.6-0.7 is poor; 0.5-0.6 is very poor.

Results

The AUC value of Bayesian network model was 0.9997, and that of random forest model and Logistic regression was 0.9273 and 0.8844, respectively, indicating that the three models could effectively fit tree mortality, and Bayesian network model performed the best.

Results showed that under the initial condition of any variable (not adjusting), the survival probability was 78% (Fig. 1). Setting the mortality state as 100%, the relationship between different levels of each variable and tree mortality can be analyzed according to the changes of each level of the variables (Fig. 2).

In this study, there were 2187 (3×3×3×3×3×3×3) combinations of causes that affect tree mortality in the Bayesian network model. It can be observed that the combination of the most likely cause for tree mortality was: low level RD, middle level AP, WMMT, age, and Gini, and high level N and MAT (Fig. 3).

As shown in Tab. 1, it can be seen that the mutual information (0.1147) of node “RD” was the largest, which indicated that it had the strongest impact on “Alive”, followed by “N” which had mutual info of 0.0688 and then followed by MAT, WMMT, AP, age and Gini.

The physiological and growth responses of trees to environmental forcing were not linear. Stand conditions could regulate the response of tree mortality to climate change (Tab. 2). Under the influence of different stand variables, the effect of climate factors on tree mortality was different. At the same time, under different climate variable conditions, the impact of stand factors on mortality was also different.

Discussion and Conclusions

Bayesian network could automatically find the dependency relationship between data and provide a theoretical framework for modeling uncertainty by using probabilistic calculus and underlying graph structure. In addition, Bayesian network has higher estimation accuracy and fitting stability than logistic regression and random forest.

RD was the dominated stand factor affecting tree mortality, and temperature (including MAT and WMMT) was the climate factor with the highest posterior probability affecting tree mortality, whereas stand structure (Gini coefficient) was less important.

Results showed that tree mortality was negatively correlated with moderate AP, WMMT, MAT and stand structure (Gini) as well as low age, but positively correlated with low RD, high N and age. According to the Most

Probable Explanation, the most likely cause of tree mortality could be determined as low RD, middle AP, WMMT, Age, and Gini, and high N and MAT.

Stand conditions modulated the response of tree mortality to climate and regulated the resilience of forests to climate stress. Moderate level of competition condition and stand structure heterogeneity weakened the negative impact of climate variables on tree mortality. The impact of age on tree mortality was greater for old trees than for young trees, which indicated that old trees were more sensitive to extreme weather conditions than young trees.

In future forest management, we can take different measures according to different stand conditions, such as managing density or adjusting stand structure, while also taking into account stand age and site conditions, so as to mitigate the adverse effects of climate change on tree mortality.

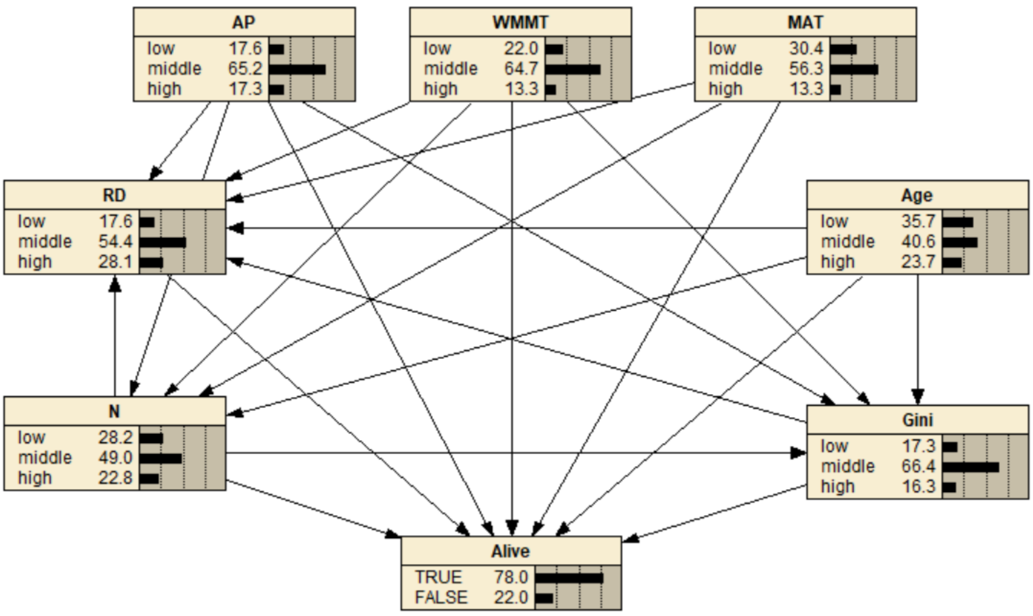

Fig. 1 Bayesian network of tree mortality in the initial case and its prior probability.

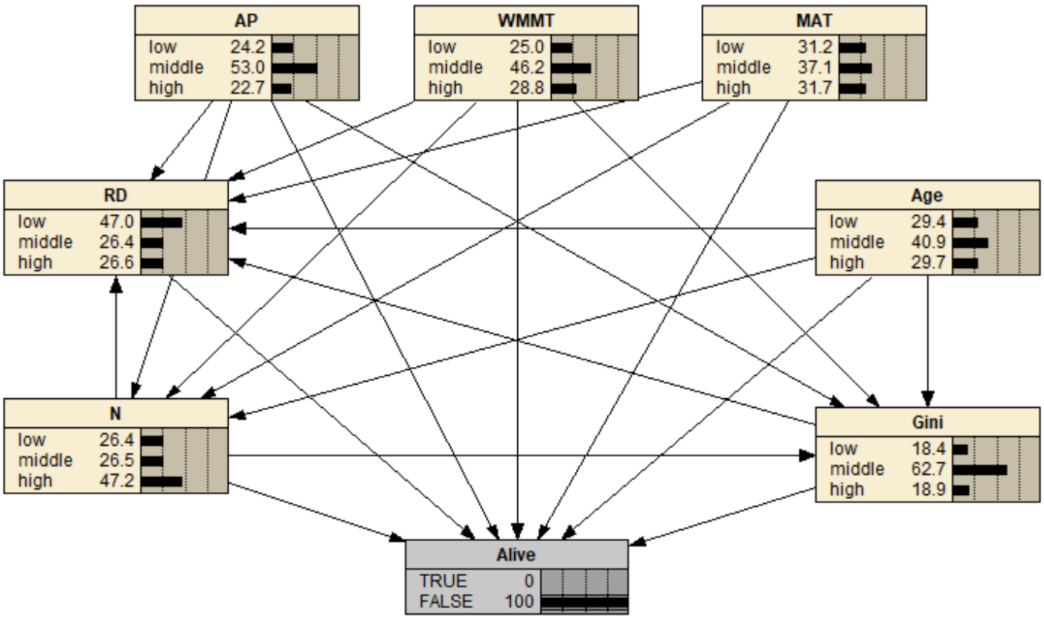

Fig. 2 Bayesian network of tree mortality after adjusting survival rate.

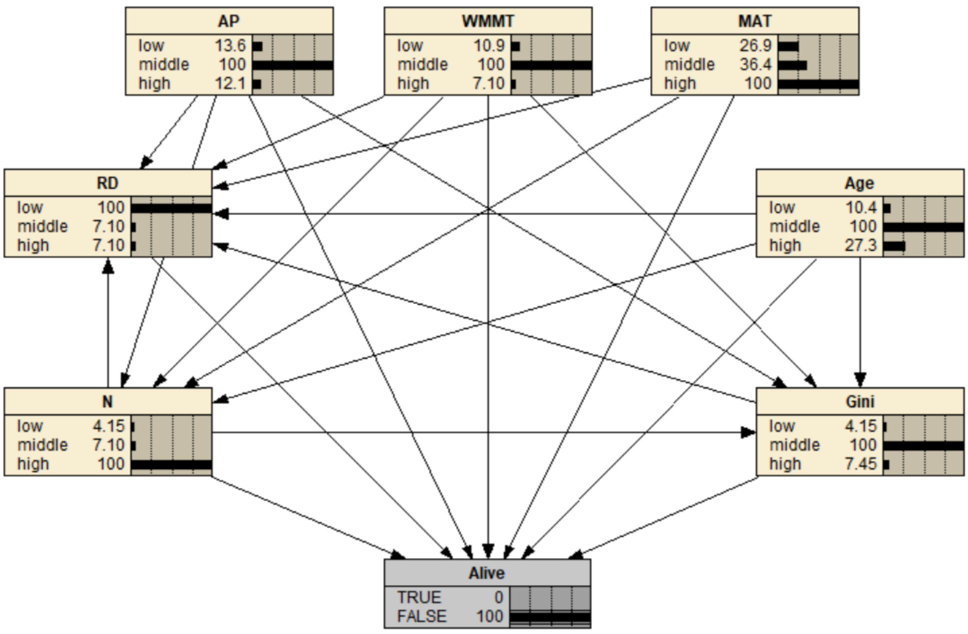

Fig. 3 Most Probable Explanation.

Tab. 1 Sensitivity analysis results ranked in decreasing order of influence on Alive occurrence based on mutual information or entropy reduction (also expresses as a percentage in brackets)

Node	Mutual Info/Entropy reduction (%)
Alive	0.76005 (100)
RD	0.11473 (15.1)
N	0.06883 (9.06)
MAT	0.05666 (7.45)
WMMT	0.0426 (5.61)
AP	0.01298 (1.71)
Age	0.00526 (0.693)
Gini	0.00133 (0.175)

Tab. 2 Changes in tree mortality under the combined influence of climate and stand conditions compared to climate conditions only.

Stand conditions modulate the response of tree mortality to climate change

Change in mortality		AP			WMMT			MAT		
		low	middle	high	low	middle	high	low	middle	high
RD	low	20.3%	45.4%	26.6%	21.4%	53.3%	3.6%	36.6%	41.1%	11.0%
	middle	-11.6%	-10.7%	-11.7%	-10.7%	-10.3%	-3.8%	-8.6%	-9.0%	-9.4%
	high	4.5%	-3.4%	1.7%	5.1%	-4.3%	0.6%	-4.9%	1.0%	-4.6%
N	low	2.2%	-3.6%	2.2%	-0.9%	-3.7%	0.9%	-6.0%	1.7%	-4.3%
	middle	-9.6%	-10.0%	-9.2%	-8.1%	-9.9%	-3.0%	-7.7%	-8.2%	-8.7%
	high	14.4%	29.1%	12.6%	17.7%	29.3%	2.5%	36.6%	18.6%	9.6%
Age	low	-8.2%	-3.7%	-0.4%	-5.9%	-4.5%	2.6%	-4.5%	-3.9%	-2.3%
	middle	3.6%	0.5%	-4.6%	0.2%	0.6%	-2.0%	-4.0%	1.3%	5.0%
	high	6.2%	4.6%	8.4%	8.3%	5.8%	-0.5%	13.8%	3.6%	-5.4%
Gini	low	9.1%	-1.5%	1.4%	9.7%	-4.2%	1.4%	1.3%	2.8%	-4.2%
	middle	-3.3%	0.2%	-3.2%	-3.5%	0.9%	-1.4%	-0.5%	-2.2%	1.6%
	high	1.0%	0.5%	12.6%	1.0%	0.7%	2.2%	0.6%	6.4%	-2.7%