Reliability Engineering and System Safety 177 (2018) 68–79

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Statistical modeling of tree failures during storms

Elnaz Kabir*[a], Seth Guikema[a], Brian Kane[b]

[a] Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI, 48109, USA
[b] Department of Environmental Conservation, University of Massachusetts, 160 Holdsworth Way, Amherst, MA, 01003, USA

ARTICLE INFO

Keywords:
Tree failure predictive modeling
Statistical learning
Tree risk assessment
Random forest
Boosting
Ensemble modeling

ABSTRACT

The failure of trees during storms imposes strong economic and societal costs. Statistical modeling for predicting the probability of a tree failing during storms has the potential to help improve tree risk management. The purpose of this study is to explore the potential predictability of tree failure using advanced predictive modeling approach. These models also have broader applicability for modeling failures of technical systems during adverse weather events. To train and test models, we use a data set from a real case study in Massachusetts, USA. We compare the out-of-sample predictive accuracy of several machine learning models including logistic regression, classification and regression trees, multivariate adaptive regression splines, artificial neural network, naive-Bayes regression, random forest, boosting, and an ensemble model of boosting and random forest. Our results demonstrate that the ensemble model of boosting and random forest achieves the best prediction accuracy in predicting the failure probability of trees for the case study storm. Our results can help tree care professionals make better decisions to reduce the risk of tree failure prior to the storm.

1. Introduction

One main component of tree failure risk assessment is estimating the probability of a tree failing. Trees are more susceptible to fail under high winds, and so thunderstorms, hurricanes, and tornadoes frequently induce tree failure. The estimates of tree failure probability during a storm form an important basis for emergency planning and response decisions by arborists, urban foresters and land managers with responsibility for trees. Between 1995 and 2007, 407 deaths were recorded as a result of wind-related tree failures in the United States [47]. Among 72 deaths from Hurricane Sandy, 20 of them were attributable to falling trees [40]. Considering that over 250 billion trees are spreading over the United States, the rate of U.S. tree fall fatalities are estimated to be more than 100 deaths per year on average [25]. In addition to directly causing deaths and serious injuries, fallen trees and branches may interfere with overhead power lines and cause considerable electricity power outages which, in turn, may result in loss of services from a number of critical infrastructure systems.

Even though falling trees may lead to widespread economic and non-economic losses to society, trees are an invaluable component of the environment making it more pleasant and healthy for humans and community [7]. Improving air and soil quality, plant life and ecological function in urban ecosystem, providing shade and energy conservation, reducing the heat island effect in cities, and increasing property values are some examples of the numerous benefits of trees [31]. It is important that the tree failure risk estimates are as accurate as possible to balance risk with the numerous benefits of trees. Using these estimates, proper actions (such as pruning the branches or removing hazardous trees) should be taken in order to manage the risks that trees present to society.

Effective tree safety management can reduce the risk of severe consequences from tree failure whilst maintaining the benefits conferred by trees. To evaluate the stability of trees during the storms, a range of techniques has been proposed in the literature. Most of these techniques are either explanatory models or mechanistic ones. Explanatory models aim to understand the causal relationships between various factors and tree failure. These models are based on physical parameters related to drag-induced stress and wood properties, and are suitable for situations in which not enough historical data is available. Mechanistic models estimate tree's risk of failure by calculating the horizontal and vertical forces on the trees [35,43] which requires to have detailed information (e.g., stem mass, soil strength, and wind induced bending moment) for each tree. Due to lack of very detailed and complex information on each tree which is required for the mechanistic models, their use is not practical in many cases [7].

Awareness of tree risk assessment and management has risen in the United States over the recent years leading to an increased collection of data about trees. If appropriate methods are used, this data can form a strong basis for tree risk assessment by developing predictive models and understanding the empirical relationships among variables which

---

* Corresponding author.
E-mail address: ekabir@umich.edu (E. Kabir).

https://doi.org/10.1016/j.ress.2018.04.026
Received 13 June 2017; Received in revised form 12 April 2018; Accepted 27 April 2018

enable us to make estimates for future cases. However, existing data-driven risk and reliability assessment methods for trees [7,21,23] and also other applications [2,3,41,42,48] are not appropriate for predictive models. In these methods, instead of validating the models on the test set which has not been used in the training process, some metrics such as *R*-squared are used in order to evaluate the fitness of the model to the training set. This approach may result in over-fitting. Furthermore, most of the employed models in the literature are parametric ones (e.g. generalized linear models), while there are a wide range of non-parametric, and non-linear models (such as boosting trees and random forest) that potentially can model the failure rate of the system accurately.

The purpose of this paper is to study advanced statistical learning methods to introduce a framework defining a general approach for modeling failures of technical systems. We use this approach to explore the potential predictability of tree failure during storms which is by itself an important problem in risk and reliability assessment. This is done by comparing the predictive accuracy of multiple approaches for estimating tree failure probability during one storm event in Massachusetts. We use various covariates including location, height, diameter at breast height (DBH), the existence of severe defects, whether a tree had been pruned, and whether or not any tree was removed within the immediate vicinity of each tree. More specifically, in order to evaluate the predictive power of different models on our data set, we compare the out-of-sample predictive accuracy of a number of statistical learning models including boosting, random forest (RF), logistic regression (Lg), classification and regression trees (CART), bagging, multivariate adaptive regression splines (MARS), Naive Bayes (NBayes), artificial neural networks (ANN), and an ensemble model of boosting and random forest. Our results show that tree failure can be accurately estimated, and the ensemble model of boosting and random forest yields the best prediction accuracy for estimating the failure probability of trees for our data set. The major contributions and novel aspects of this research are as follows:

- Indicating the potential predictability of tree failure using advanced statistical models. Note that there are few papers that use predictive modeling for storm-related tree failures and most proposed models are either mechanistic or explanatory models, or they try to find some relationships between tree failure and other covariates like height and DBH.
- Making interpretations for importance and effects of predictors on tree failure probability to identify the model's strengths in data inference.
- Demonstrating a range of statistical models that can be used in developing predictive models for tree failure probability. In this study, we go beyond parametric regression models (which have been investigated by [20–22,36]), and present a wide range of non-parametric, and non-linear regression models that potentially can model the failure probability of trees accurately.

The remainder of this paper is organized as follows. Section 2 provides a survey of existing models for assessing the failure rate of trees. In Section 3.1, we briefly review the source of data set and its analysis. The implemented statistical learning methods are briefly discussed in Section 3.2. In Section 4, we provide a summary of our main findings, insights and results. Section 5 explains the practicality of statistical methods for tree risk assessment. We explain the limitation of our data set in Section 6, and conclude the paper in Section 7.

2. Related works

A range of approaches have been proposed in the literature to evaluate the resistance of trees in the face of heavy winds and storms and to manage the associated damage risk of trees. These approaches include (1) explanatory models, (2) mechanistic models, and (3) statistical models. The first group is based on observing key factors such as tree species, tree dimensions, soil quality, uprooting, stem breakage, and branch breakage. These factors are explored to understand the causal relationships between the variables and tree failure [13,46]. Gardiner and Quine [13] assert that the assessment of trees' damage caused by wind should be entirely objective even though it is often subjective in practice due to the lack of complete information. Therefore, they suggest performing statistical learning analysis so as to investigate the effects of wind on trees.

Second, mechanistic models describe the behavior of trees under wind by calculating the horizontal and vertical forces on the trees as well as the maximum wind speed that a tree can withstand [23,35,43]. HWIND [35] and GALES [13] are mechanistic models to assess the risk of wind damage on trees, and they have been extended by other researchers [1]. Gardiner et al. [12] review mechanistic models to predict failure of forest- or plantation-grown trees. Due to complexity and limitations of measuring tree characteristics (e.g., stem mass and tree mass), soil strength, and wind induced bending moment on each tree, which are required information for mechanistic models, often it is imposible to obtain accurate estimations of the failure probability for trees by using mechanistic models [7]. The mechanistic models are also limited by not completely addressing the dynamic interaction of wind and tree [15].

Fewer studies have considered the likelihood of failure of open-grown trees. Open-grown trees typically grow with a more complex crown architecture, and a greater proportion of tree mass is in branches rather than the trunk. Applying simplifying mechanical assumptions (e.g., that the trunk is a simple cantilevered beam of predictable taper and with a uniformly circular cross-section; or that wood is homogeneous and isotropic, yielding constant strength and stiffness throughout the tree's structure) to open-grown trees is particularly problematic. Recent reviews of the static [8] and dynamic [19] load-bearing capacity of open-grown trees discuss these issues in detail.

Finally, statistical models are potentially an attractive approach for predicting the probability of wind-related tree failure. Even though there exists a wide range of statistical models, only a few have been employed in the literature (i.e., [7,20–23,36]). Cifti et al. [7] present a Monte Carlo based method for quantifying the failure probability of trees subject to wind storms. While most previous models in the literature operate on the scale of the forest stand or plantation, Cifti et al. [7] assess the failure probability of individual trees. They apply their method on two maple trees in Massachusetts, USA. One limitation of their proposed method is that it is computationally expensive and not an efficient method for assessing the failure risk of large number of trees.

Others have studied failed and standing trees following severe wind storms and developed logistic regression models to predict failure from parameters such as species, trunk diameter, the presence of severe structural defects, and whether trees had been recently pruned [22,36]. Kamimura et al. [20] also develop a logistic regression model from one storm event and test it on a second storm event. In a study that simulates wind loads by pulling trees with a winch, Kane [23] develops a logistic regression model to predict failure as a function of the loss in load-bearing capacity of a trunk. Despite their utility, however, all of these studies use a single statistical technique: logistic regression. To our knowledge, there are no studies that have compared a variety of statistical methods to determine which best predicts tree failures. This is important because no single statistical technique consistently provides the most accurate predictions for all storm events. Multiple statistical techniques must be compared to choose the most accurate method [27].

3. Data description and methods

In this section, first we introduce and summarize the data set. Then we recapitulate the statistical models.