Evaluating and Predicting Southern Pine Beetle Occurrence in Relation to Forest Stand Conditions and Silvicultural Treatments

C. Kuetsinya
Department of Experimental Statistics, Louisiana State University

1.0 Introduction

1.1 A Brief Overview

The southern pine beetle (SPB), *Dendroctonus frontalis Zimmermann*, is a major concern in pine forests as its outbreaks can lead to severe impacts on forest productivity and health, resulting in billions of dollars in economic losses [1]. Despite the extensive research on SPB ecology, there remains a lack of complete understanding of its population dynamics and outbreak patterns. The USDA Forest Service and the Southern Group of State Foresters have been encouraging and providing cost-share assistance to silvicultural treatments such as prescribed burns and thinning to reduce stand susceptibility to SPB in the southeastern United States through the Southern Pine Beetle Prevention Program (SPBPP) since 2003 [2].

Recent research and observations suggest that rapid detection and removal of infestations as well as silvicultural practices provide some stands or forests with resistance against SPB infestation [2]. This study provides an understanding of the factors that influence the incidence of SPB occurrence in pine forests, as well as insights into the effectiveness of silvicultural practices in preventing and managing SPB outbreaks.

1.2 Purpose of the Project

The study predicts SPB occurrence in Bienville and Homochitto National Forests in Mississippi using 2012 and 2016 data. This study used the diameter class of trees, basal area of pine trees, Palmer drought severity index, burn information, stand age, soil type and site index to predict SPB infestation in loblolly pine forests. The findings of this study provided valuable information to forest management agencies for better management and control of SPB occurrence in the future.

1.3 Research Questions

1. How well can environmental factors of interest and silvicultural activities predict the population dynamics of southern pine beetles?
2. Which variables are important to help reduce SPB occurrence and improve the management of pine forests?

2.0 Literature Review

Southern pine beetle (Dendroctonus frontalis Zimmermann) is a major pest of pine forests in the southeastern United States, causing extensive damage and economic losses. To manage these infestations effectively, it is critical to evaluate and predict their incidence and severity in relation to forest stand conditions and silvicultural treatments.

Thinning and prescribed burning are commonly recommended management strategies to reduce the risk of beetle infestations. The effectiveness of these treatments may depend on several factors, including the specific tree species, site conditions, and timing and frequency of treatments. Further research is necessary to better understand the underlying causes of the decline in beetle outbreaks and to develop effective management strategies for preventing future infestations.

After southern pine beetle infestations, Xi et al. (2012) discussed a goal-oriented approach to forest landscape restoration. The authors emphasized the importance of considering the ecological goals of restoration efforts, such as restoring forest structure, composition, and function, as well as controlling beetle infestations. According to the authors, a combination of silvicultural treatments such as thinning, prescribed burning, and planting resistant tree species could be effective in restoring forest landscapes after beetle infestations.

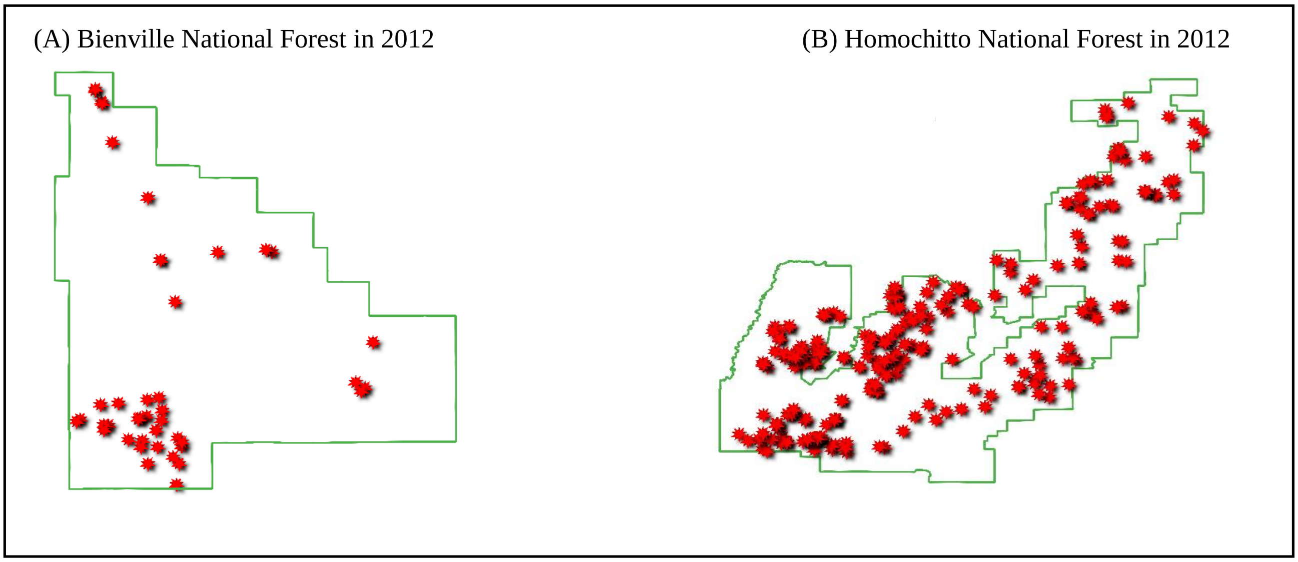
Nowak et al. (2015) discovered that forest stands that had been treated with prescribed burning or thinning were less likely to have beetle infestations than untreated stands. However, they found conflicting results about the impact of prescribed burns on SPB infestation. Data collected in the Homochitto National Forest revealed that stands that received a prescribed burn had lower SPB activity than those that did not, whereas data collected in the Bienville National Forest revealed that prescribed fire had no effect, or had a positive effect, on SPB activity. The disparity between these two forests in close proximity creates uncertainty, and a closer examination of the data to determine potential causes is warranted.

3.0 Methodology

3.1 Data Source

The data was sourced from publicly available GIS files provided by the USDA Forest Service. It contains information on SPB spots in stands of loblolly pine forest that are less than 45 years old. The variables in the data include the longitude and latitude of each SPB spot, the year of infestation, the type of National Forest, the diameter class of the trees, basal area of the pine trees (BA_PINE), Palmer drought severity index (PDSI), burn information, site index, soil type, the number of infested trees and the number of treated acres.

Figure 1. SPB spots in Bienville National Forest and Homochitto National Forest



3.2 Data Analysis

The data in this study has two potential response variables: the number of infested trees, which was obtained during the preliminary stages of SPB spot identification, and the number of treated acres, which was obtained during the treatment stage. Therefore, this study explored both response variables allowing for a more nuanced analysis of the data.

The study employed the random forest algorithm and linear regression techniques, to develop different models. By using a combination of these approaches, the study aimed to provide a more comprehensive understanding of the factors influencing SPB occurrences and to develop more accurate predictive models for future SPB outbreaks.

Random Forest is a highly effective ensemble learning method that is used for classification and regression problems. It operates by constructing decision trees from random samples of the training data, creating a "forest" of trees that can be used for making predictions. Each tree in the forest is generated using a random subset of the available features as the split criteria, ensuring that each tree makes predictions based on different features, thereby minimizing the risk of overfitting to any particular feature. In regression problems, the final prediction is made by averaging the predictions of all trees in the forest, while for classification problems, the final prediction is determined by a majority vote.

Linear regression, on the other hand, is a statistical method used to model the relationship between one or more independent variables and a dependent variable. It works by identifying the line of best fit that describes the relationship between the variables. The line of best fit is a straight line that minimizes the difference between the actual values of the dependent variable and the predicted values based on the independent variable(s). This method is useful for predicting the dependent variable based on the values of the independent variable(s).

The steps of the analysis included;

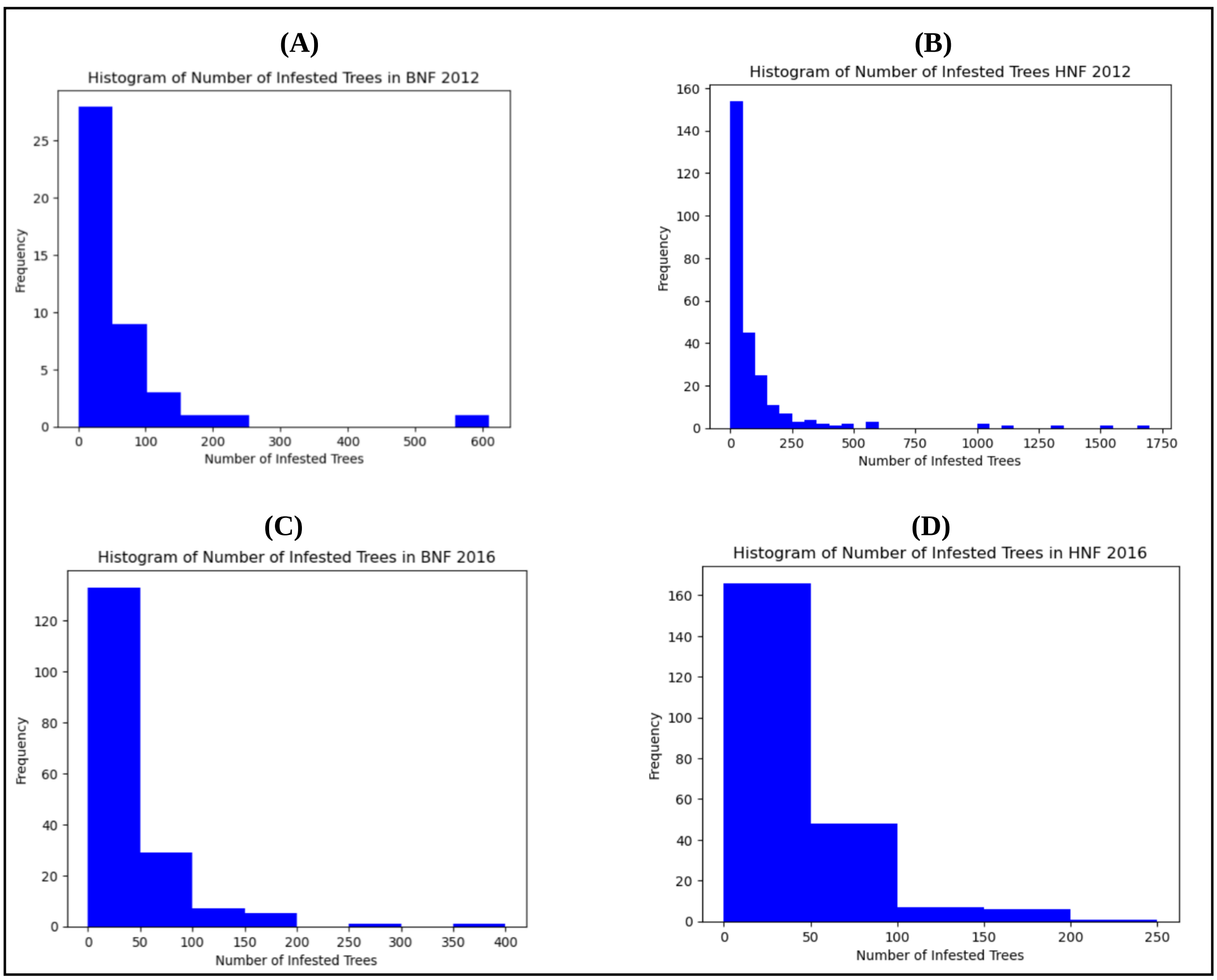
1. Splitting the data into the year by forest (BNF 2012, HNF 2012, BNF 2016 & HNF 2016)
2. Building the models using the BNF 2012 and HNF 2012 as training data
3. Test the models on BNF 2016 and HNF 2016 respectively
4. Compare the models using performance metrics

4.0 Results & Discussion

4.1 Exploratory Data Analysis

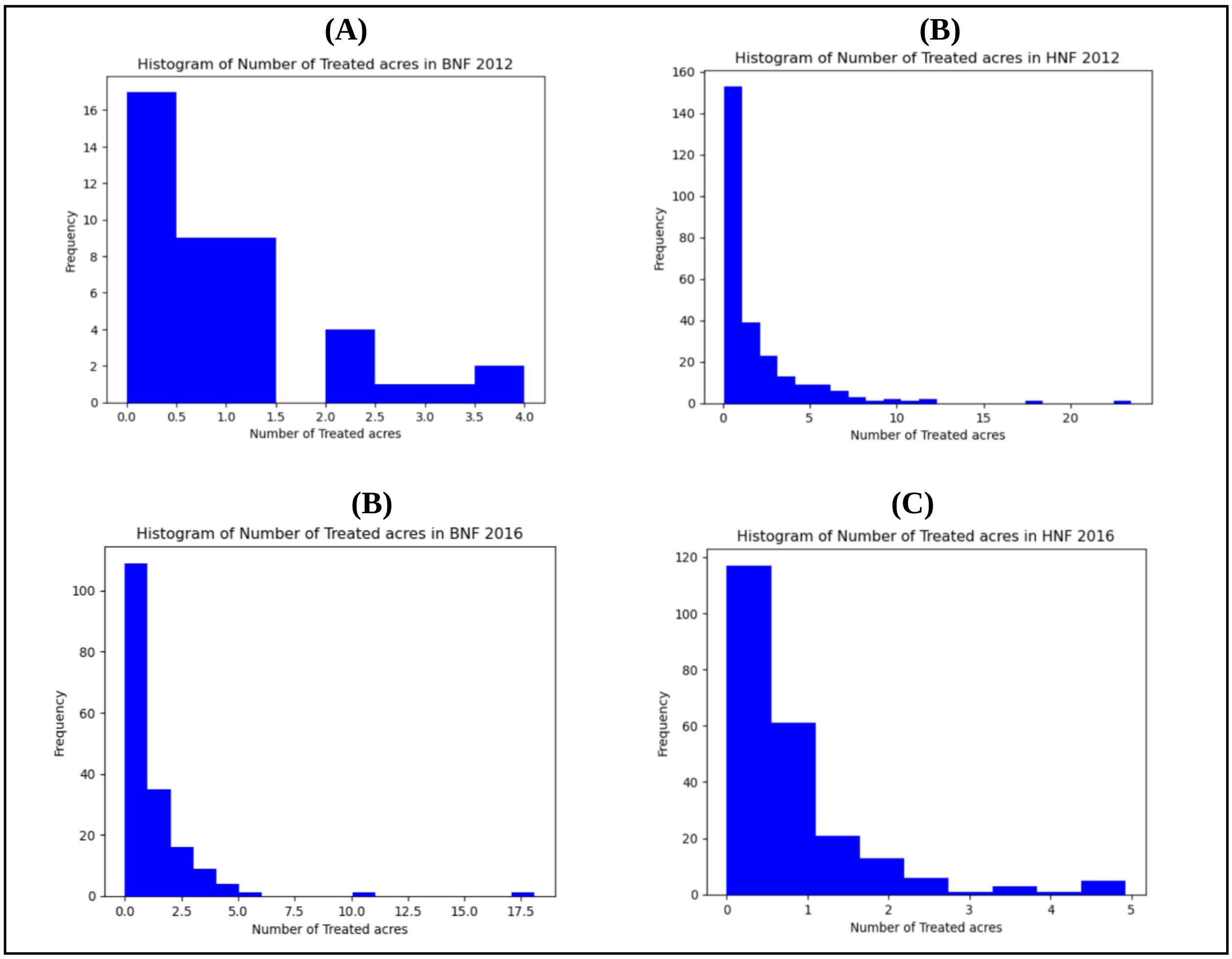
The analysis of the distribution of the number of infested trees across the two national forests and years provides insight into the severity of SPB infestation. The distribution reveals that the majority of SPB spots have infested trees with counts less than 250, indicating a relatively low level of infestation in most cases. However, there are notable outliers, in both forests and across years that suggest a more severe infestation in some cases. Specifically, in 2012, both BNF and HNF experienced a significantly higher number of infestations. These outliers may indicate an unusually severe outbreak of SPB in these years or may be due to specific environmental or other factors that are not present in other years or locations. Overall, the distribution of the number of infested trees suggests that SPB infestation is generally moderate to low in severity in most areas, but there are instances where the infestation is more severe and requires closer attention and targeted management strategies.

Figure 2. Distribution of the number of infested trees



The analysis of the number of treated acres across the SPB spots provides insight into the extent and effectiveness of management strategies used to control SPB infestations. Our analysis reveals that the majority of the SPB spots have treatment areas ranging from 0 to 5 acres. However, there are a few notable exceptions where a higher number of acres were treated, indicating a more extensive or severe infestation in those areas. These areas may have received more resources or attention due to the severity of the infestation or other factors such as proximity to high-value or high-priority areas. The effective use of resources and the identification of high-priority areas for treatment can play an important role in the management of SPB infestations.

Figure 3. Distribution of the number of treated acres



4.2 Predictive Models

The analysis of the results of the predictive models for BNF and HNF reveals insights into the relationship between the response variables, i.e., the number of infested trees and the number of treated acres, and the predictor variables used in the models.

For BNF, the random forest model for the number of infested trees performed better than the linear regression model, with a lower mean absolute error (MAE) of 41.12, a mean squared error (MSE) of 3821.99, and a root mean squared error (RMSE) of 61.82. The linear regression model had a higher MAE of 140.62, a higher MSE of 23425.50, and a higher RMSE of 153.05. For the number of treated acres in BNF, the random forest model had a lower MAE of 1.00, a lower MSE of 3.81, and a lower RMSE of 1.95 compared to the linear regression model, which had an MAE of 1.20, an MSE of 4.42, and an RMSE of 2.10.

For HNF, the random forest model for the number of infested trees had a higher MAE of 178.05, a higher MSE of 52894.92, and a higher RMSE of 229.99 compared to the linear regression model, which had a lower MAE of 116.32, a lower MSE of 16135.73, and a lower RMSE of 127.02. For the number of treated acres in HNF, the random forest model had a higher MAE of 1.77, a higher MSE of 4.57, and a higher RMSE of 2.13 compared to the linear regression model, which had a lower MAE of 1.53, a lower MSE of 2.96, and a lower RMSE of 1.72.

Overall, the results suggest that the random forest models performed better than the linear regression model for the number of infested trees in BNF and the number of treated acres in both BNF and HNF. However, for the number of infested trees in HNF, the linear regression model performed better than the random forest model. These results provide insights into the effectiveness of different predictive models for understanding and predicting the spread and impact of SPB infestations in different forest regions.

Table 1. Results

	Response Variable	**Model**	**MAE**	**MSE**	**RMSE**
BNF results	Number of infested trees	Random Forest	41.12	3821.99	61.82
		Linear regression	140.62	23425.50	153.05
	Number of treated acres	Random Forest	1.00	3.81	1.95
		Linear regression	1.20	4.42	2.10
HNF results	Number of infested trees	Random Forest	178.05	52894.92	229.99
		Linear regression	116.32	16135.73	127.02
	Number of treated acres	Random Forest	1.77	4.57	2.13
		Linear regression	1.53	2.96	1.72

5.0 Conclusion

The results of the predictive models show that the predictors are effective in predicting the number of treated acres in the SPB spots. Therefore, the number of treated acres can be considered the ideal response variable based on the predictive model results.

Looking at the Bienville National Forest (BNF), the random forest model with the number of treated acres as the response variable performs better. The mean absolute error (MAE) for the random forest model is only 1.00, indicating that the predicted values are close to the actual values. The mean squared error (MSE) and root mean squared error (RMSE) for the random forest model are also lower than those of the linear regression model.

For the Homochitto National Forest (HNF), the linear regression model with the number of treated acres as the response variable is the best predictive model. The MAE for this model is 1.53. The MSE and RMSE for the linear regression model with the number of treated acres as the response variable are also lower than those of the random forest model with the number of infested trees as the response variable. However, it is worth noting that the MAE, MSE, and RMSE for the random forest model with the number of infested trees as the response variable are relatively high compared to the results for the other models.

Overall, the predictive models demonstrate that the number of treated acres is a better response variable compared to the number of infested trees. The models also indicate that different types of models perform better for different national forests. While the random forest model is the best for BNF, the linear regression model performs better for HNF.

6.0 References

Meeker, J. R., Dixon, W. N., Foltz, J. L., & Fasulo, T. R. (2004). Southern Pine Beetle, Dendroctonus frontalis Zimmermann (Insecta: Coleoptera: Scolytidae): EENY-176/IN333, rev. 3/2004. EDIS, 2004(4).

Nowak, J. T., Meeker, J. R., Coyle, D. R., Steiner, C. A., & Brownie, C. (2015). Southern pine beetle infestations in relation to forest stand conditions, previous thinning, and prescribed burning: Evaluation of the southern pine beetle prevention program. Journal of Forestry, 113(5), 454-462.

Xi, W., Waldron, J. D., Cairns, D. M., Lafon, C. W., Birt, A. G., Tchakerian, M. D., ... & Coulson, R. N. (2012). Restoration of southern pine forests after the southern pine beetle. A Goal-Oriented Approach to Forest Landscape Restoration, 321-354.