Mapping Oak decline susceptibility with data mining

Lavado Contador, J. F.1, Herguido Sevillano, E.1; Gómez Gutiérrez, Á.1; Schnabel, S.1; Cardillo Amo, E.2

1: University of Extremadura, Spain; 2: CICYTEX (IPROCOR), Extremadura, Spain

ABSTRACT

Oak decline and sudden death, also known as b在同一 phenomenon that causes widespread occurrence in the region of the SUD phenomenon, is a significant threat to oak forests in Spain. To address this issue, we propose a method that combines data mining techniques with geographic information systems (GIS) to map the susceptibility of oak forests to oak decline. The method is based on the analysis of various topographical and environmental variables, including mean temperature, elevation, and soil fertility. The results show that the method is effective in mapping the susceptibility of oak forests to oak decline, and it can be used to inform management decisions to mitigate the impact of this phenomenon.

INTRODUCTION

Oak decline and sudden death is a complex phenomenon that can be influenced by various environmental factors. The impact of this phenomenon on oak forests in Spain is significant, and there is a need for effective methods to map its susceptibility.

METHODS

The method proposed in this study combines data mining techniques with Geographic Information Systems (GIS) to map the susceptibility of oak forests to oak decline. The method involves the analysis of various topographical and environmental variables, including mean temperature, elevation, and soil fertility.

RESULTS

The results show that the method is effective in mapping the susceptibility of oak forests to oak decline, and it can be used to inform management decisions to mitigate the impact of this phenomenon.

CONCLUSIONS

The proposed method is effective in mapping the susceptibility of oak forests to oak decline. It provides valuable information for decision-makers to manage oak forests more effectively.

REFERENCES


MODEL DESIGN

DATA MODEL FOR TRAINING AND VALIDATION

A survey of decline affected areas provided geo-localized locations of places to be used in the predictive models. A total number of 1742 data records were used for modeling that included 52 AOS as a plot feature layer. 896 records out of the 1742 were determined as undergoing oak decline, the rest were randomly assigned to decline area. The set of data points was split into two groups:

- 75% (879 of the 1174) were used for model training.
- 25% (277 of the 1174) were used as test data for model validation.

A data matrix was then related to a set of 115 variables considered as explaining factors (see model variables). Each of the explaining variables was available or constructed as a map where data points were consulted.

STATISTICAL METHODS

RESULTS OF THE INDIVIDUAL MODELS

CONSTRUCTION OF THE SUSCEPTIBILITY MAPS

- The prominence of different areas to undergo oak decline was expressed as maps in two ways:
  - As constant values reflecting the modeled probability of seca (Figs. 1.2, 3).
  - By thresholding binary maps of predicted seca vs. no seca areas adopting to the baseline model response threshold for the different data mining algorithms (Figs. 3, 4, 5).

MODEL ENSEMBLING

In order to explore whether the combination of the responses obtained with the different data mining algorithms could be improved by ensembling, model ensembles were constructed in two ways:

- A binary map was constructed by majority vote based on the values of the thresholded maps (Fig. 7).
- A constant probability ensemble was constructed with the weighted sum of the model probabilities, using as weighting factor the AUC value of the respective models (Fig. 8).

An assessment of the reliability of the new ensembles was also performed with independent test data. To this, the AUC was estimated:

- Areas of Weighted Probability ≥ 0.75
- Areas of Weighted Probability ≥ 0.95
- Areas with a number of seca voxels ≥ 3
- Areas with a number of seca voxels ≥ 2

CONCLUSIONS

Data mining has demonstrated to be reliable when mapping the potential regional susceptibility to undergo oak decline (seca). Patterns were similar to different mining algorithms. Results are promising considering the complexity of predictors’ nature, dealing well with algorithms with multicollinearity predictors’ recovery and diversity of scales.

ACKNOWLEDGEMENTS

This work was financed by the Spanish Ministry of Economy and Competitiveness through the JERICO (COE0101-22531) project. This support is gratefully acknowledged.

CONTACT

J.F. Lavado Contador
Geo-Environmental Research Group
University of Extremadura, Spain.
filgado@unex.es, filgado@unex.es

REFERENCES


ROC AUC of the ensemble models. Weighted probability and number of votes

ROC

Weighted probability

Number of votes

Fig. 6

Fig. 7