Modelling and prediction of wind damage in forest ecosystems of the Sudety mountains, SW Poland

Science of the Total Environment 815 (2022) 151972

Contents lists available at ScienceDirect

Science of the Total Environment

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

Modelling and prediction of wind damage in forest ecosystems of the Sudety Mountains, SW Poland

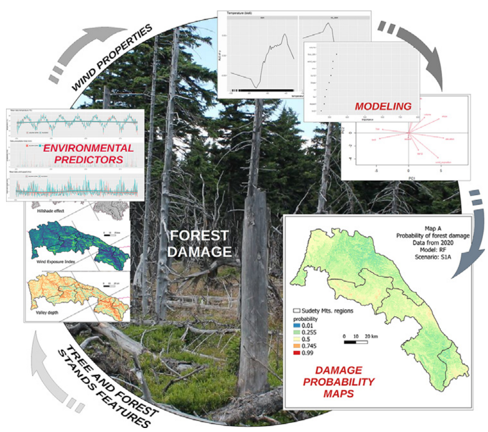
Łukasz Pawlik [a,*], Sandy P. Harrison [b]

[a] University of Silesia, Faculty of Natural Sciences, Institute of Earth Sciences, ul. Będzińska 60, 41-200 Sosnowiec, Poland
[b] University of Reading, School of Archaeology, Geography, and Environmental Sciences, Earley, Reading RG6 6AX, United Kingdom

HIGHLIGHTS

- Five machine learning models give consistent predictions of what controls wind damage.
- Tree volume and age are the most important predictors of forest damage caused by wind.
- Geomorphic and climate predictors are less important.
- Random forest algorithm and gradient boosting modelling offer the best accuracy of prediction.
- Forest stand features might significantly influence probability of forest damage.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 30 May 2021
Received in revised form 9 November 2021
Accepted 22 November 2021
Available online 27 November 2021

Editor: Manuel Esteban Lucas-Borja

Keywords:
Forest damage
Windstorm
Disturbance
Wind climate
Machine learning
Poland

ABSTRACT

Windstorms are one of the most important disturbance factors in European forest ecosystems. An understanding of the major drivers causing observed changes in forests is essential to improve prediction models and as a basis for forest management. In the present study, we use machine learning techniques in combination with data sets on tree properties, bioclimatic and geomorphic conditions, to analyse the level of forest damage by windstorms in the Sudety Mountains over the period 2004–2010. We tested four scenarios under five classification model frameworks: logistic regression, random forest, support vector machines, neural networks, and gradient boosted modelling. Gradient boosted modelling and random forest have the best predictive power. Tree volume and age are the most important predictors of windstorm damage; climate and geomorphic variables are less important. Forest damage maps based on forest data from 2020 show lower probabilities of damage compared to the end of 20th and the beginning of 21st century.

* Corresponding author.
E-mail address: lukasz.pawlik@us.edu.pl (Ł. Pawlik).

http://dx.doi.org/10.1016/j.scitotenv.2021.151972

1. Introduction

Wind damage is one of the most important abiotic disturbances in European forests (Schelhaas et al., 2003; Brázdil et al., 2004; Gardiner et al., 2010; Pawlik, 2013; Kulakowski et al., 2017; Senf and Seidl, 2020; Pettit et al., 2021) having a significant impact on tree mortality, soils, and hillslope dynamics (Šamonil et al., 2009; Pawlik et al., 2016). Wind disturbance as a "*meteorological extreme*" (Brázdil, 1998) is a particularly strong driver of forest change in places where other natural catastrophes, e.g. tsunami, volcanic eruptions or major landslides, are absent or infrequent (Seidl et al., 2017). Wide-scale replacement events (*stand-replacing disturbances*) which damage almost all trees (Evans et al., 2007a, 2007b) happen relatively rarely in temperate forests, but have long lasting consequences on forest ecosystems and society (Gardiner et al., 2010, 2013). Other low magnitude events that create patches and gaps in forest ecosystems are more frequent and increase forest biodiversity and health by eliminating weaker trees and causing the regeneration of trees (*gap-phase dynamics*) growing under the forest canopy (Yamamoto, 2000; Šamonil et al., 2009).

Wind speeds over 29 m s^{-1} (F0 in Enhanced Fujita scale; Edwards et al., 2013) will damage any forest ecosystem regardless of its health and structure. However, in most cases, other factors play a role and impact the spatial patterns and magnitude of the damage (Ruel, 1995; Schindler et al., 2012). The meteorological origin of the wind event, whether associated e.g. with mesoscale low pressure centers of continental scale, orographic winds or tornadoes of regional and local impact, affects wind properties that affect the scale of damage including duration, gustiness, vorticity, speed and direction (Xi and Peet, 2011). Terrain properties such as slope exposure and valley depth, modify the wind flow (Everham and Brokaw, 1996; Xi and Peet, 2011). Soil properties, such as depth, texture, bulk density, moisture content, and organic matter content, affect landscape stability and hence the ability of forests to withstand disturbance (Schaetzl et al., 1989). Finally, forest properties, such as the age and height of the trees, forest composition and fragmentation, and tree health status, can also affect the amount of damage incurred during windstorms (McMaster, 2005; Coates et al., 2018). It is important to emphasize here the fundamental differences between managed and old-growth forests, the latter being more resistant to the damaging force of wind events (Pawlik et al., 2016; Pettit et al., 2021).

Previous studies show that wind has been an important natural disturbance factor in many parts of Europe during the past 300 years (Zielonka and Malcher, 2009; Panayotov et al., 2011; Svoboda et al., 2012, 2013; Janda et al., 2014; Trotsiuk et al., 2014; Pettit et al., 2021), although there is considerable variability in disturbance regimes between various regions and forests (managed vs old-growth forests). It has been suggested that strong wind severity and frequency has increased during the past century and that this has been accompanied by increases in both the area affected and volume of damaged trees (Gregow et al., 2017). During the past 60 years, there have been many strong wind events of disastrous consequences for European forests. This includes the clusters of winter storms in 1990 (windstorm *Daria* – 25–26 January; windstorm *Vivien* and *Wiebke* – 25 February–1 March) and in 1999 (windstorm *Anatol* – 2–4 December; windstorm *Lothar* – 24–27 December; windstorm *Martin* – 25–28 December; Vitalo and Stephenson, 2009; Gardiner et al., 2010). These trends have been attributed to climate change, and it has been suggested that future climate change will further increase wind damage in forests (Lindner and Rummukainen, 2013; Gregow et al., 2017; Seidl et al., 2017). However, it is still unclear whether the increases in wind damage reflect a marked increase in storminess of extra-tropical cyclones as a direct consequence of ongoing climate change (Gregow et al., 2017) or are an indirect effect of the temperature and precipitation changes on soil properties (Peltola et al., 1999; Lindner and Rummukainen, 2013). The effects of warmer soil temperatures in winter appear to have been responsible for greater forest damage during high-speed wind events in Switzerland, for example, when there was no significant change in wind parameters (Usbeck et al., 2010). Most of the damage in European forests was associated with the passage of extra-tropical cyclones and their impact on managed forests (Schelhaas et al., 2003; Brázdil et al., 2004; Gardiner et al., 2010). Although, this impact has increased over the 20th century recent results show it did not affect European primary beech forests (Pettit et al., 2021).

Predictive models of forest damage have a long history (Gardiner, 2021). Two types of models have been tested: 1) hybrid-mechanistic, and 2) statistical models (Hart et al., 2019; Gardiner, 2021). Hybrid-mechanistic models, such as the ForestGALES model (Hale et al., 2015), require two components: 1) understanding and calculation of the physical behaviour of the tree canopy under the wind load, and 2) information on the local and regional wind regime (Gardiner et al., 2008; Gardiner, 2021). The second component needs to be created by interpolation of wind data with topography and surface roughness taken as covariates. Statistical models, including machine learning models, seek to identify the best predictors of forest damage based on large datasets. There are many types of machine learning models (Kuhn and Johnson, 2013; Molnar et al., 2018), including logistic regression, random forest (RF), and artificial neural networks (ANN). These various types of models have different strengths and weaknesses, and it is not always clear which is the most appropriate for the analysis of a given problem. It is common practice, however, to use multiple models to assess the robustness of the findings (see e.g. Hengl et al., 2018).

Along with a growing concern over climate change and the exploitation of natural resources, an assessment of the impact of wind damage on forest ecosystems is important for carbon sequestration, forest management and silviculture, hillslope stability and geohazards, and wildfire risk. This requires a better understanding of the factors that influence the magnitude of damage during any given storm. In this paper, we apply five machine learning techniques, logistic regression (GLM), random forest (RF), radial basis function support vector machines (SVM), neural networks (NN) and gradient boosted modelling (GBM), to analyse what factors influence the scale of forest damage focusing on the Polish part of the Sudety Mountains. Due to its complex relief, forest history, and wind climate, this massif is particularly prone to severe and frequent damage from strong wind-related events (Dmyterko et al., 2015). Our goal is to determine the best predictors of forest damage in order to improve forest management and support analyses in other scientific disciplines, for instance, forest ecology, soil science, and geomorphology, in which information on forest damage can help interpretation and formulation of final conclusions. Given that there is no obvious best choice of machine-learning model: some models perform better than others but this is not a consistent finding across data sets (e.g. Hengl et al., 2018), we test the robustness of our conclusions by training five different models across four data sets. By using several models we can see whether the conclusions are robust or model-dependent.

2. Data and methods

2.1. Study area

The Sudety Mountains (Fig. 1) belong to the Bohemian Massif and the region is highly complex in terms of geology and geomorphology. The highest part of the massif, the Giant Mountains reaching 1603 m a.s.l. at Mt. Śnieżka, is in the western part of the Sudety and is formed from different types of granite and metamorphic rocks. The Middle Sudety has a post volcanic and tableland landscape dissected by deep and wide river valleys. In some places, the pattern of hillslope morphology has been shaped by large-scale landslides (Migoń et al., 2010). The eastern part of the Sudety Mountains is built of metamorphic rocks that are easily eroded and weathered. The entire massif is tectonically unstable and dissected into smaller blocks that are uplifted or lowered along normal faults (Różycka and Migoń, 2017), resulting in large differences in elevation and in hillslope steepness that can reach >30°. The climate is influenced by regional circulation and also by topography. During the period analysed in this study, the mean daily temperature at Jelenia Góra and Kłodzko was 7.9 °C, and mean daily precipitation total was 1.8 mm. Mean daily wind speed was 2.9 m s^{-1} (https://danepubliczne.imgw.pl/).

The regional forest cover has been strongly impacted by human activities over the last several hundred years, with many parts of the Sudety deforested for agricultural purposes (Mazurski, 1986). Higher demand for

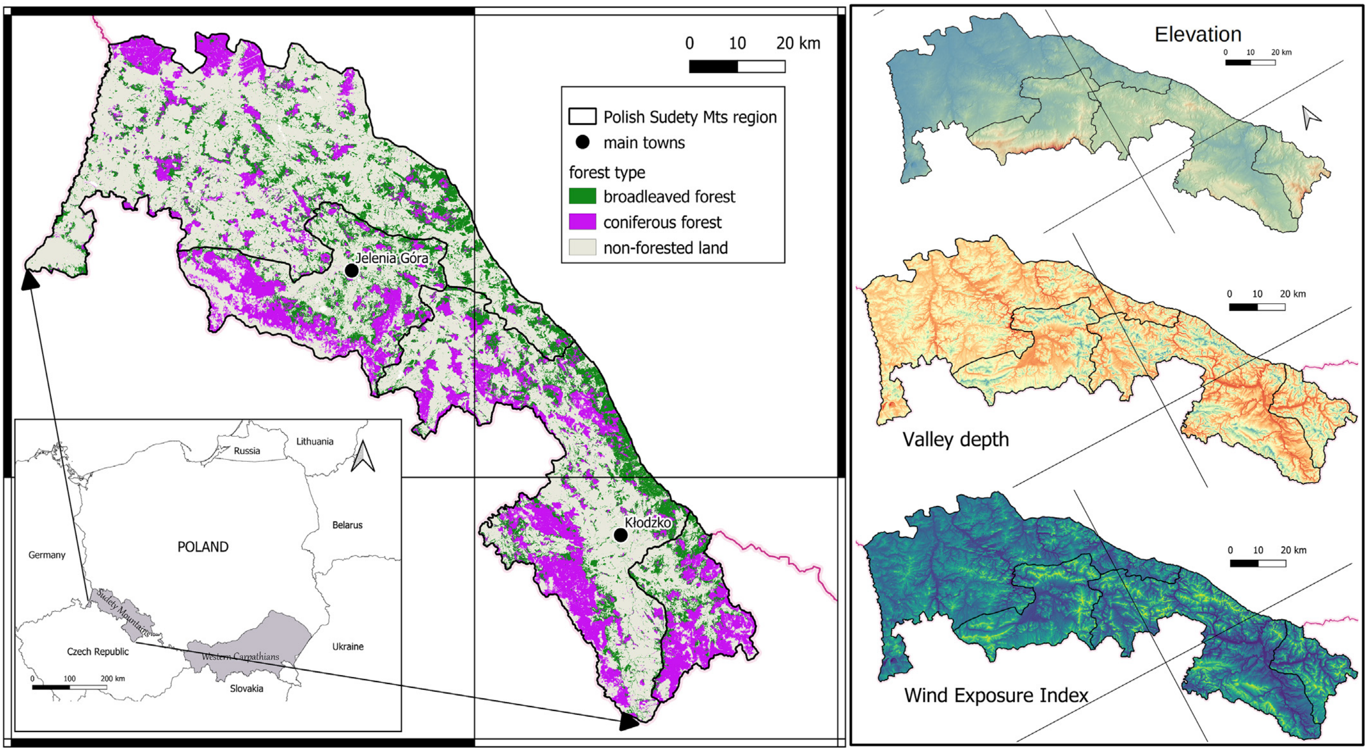


Fig. 1. Study area and its forest cover. The right panel shows selected terrain properties that have been used as potential predictors: elevation (in m a.s.l.), valley depth (in meters), and Wind Exposition Index (nondimentional metric).

wood resulted in the extensive planting of *Picea abies* in the second part of the 19th century and in the first part of the 20th century. This had negative consequences for forest stability and resistance against natural disturbances, e.g. bark beetle outbreaks or high-wind impact (Pawlik et al., 2016). The greatest wind damage in the Polish part of the Sudety Mountains was recorded in 2007, when the *Kyrill* windstorm of 18–19 January damaged over 1 million m^3 of trees in the Lower Silesia Province (Fink et al., 2009a, 2009b; Klaus et al., 2011; Pawlik, 2013).

2.2. Data sources and preparation

We obtained data on the characteristics of individual forest units and wind damage from the Polish Institute of Forest Research. Although the forest characteristics (tree age, tree height, diameter and breast height, stand volume, dominant tree species) are recorded every 10 years for each of the 430 forest districts in Poland, wind damage in each forest district (measured in m^3 of damaged trees per forest unit) is recorded and mapped immediately after a storm event of any magnitude. However, when damage is extremely high, as was the case for the Kyrill windstorm in 2007, recording, mapping, and forest cleaning can take several months (Pawlik, 2013). Measurements of damage do not distinguish tree species only total damage. We used wind damage estimates made between 2004 and 2010 (Table 1). The forest damage data were recorded and mapped based on a vector layer of basic forest units (Fig. 1S, Appendix 1).

The predictor climate variables (Table 1), obtained as rasters at 30′ resolution either from the WorldClim database version 2.1 (www.worldclim.org, Harris et al., 2014; Fick and Hijmans, 2017) or from the CHELSA database (http://chelsa-climate.org/; Karger et al., 2017), were average monthly wind speed from January (wind1) to December (wind12), annual mean temperature (bio1), minimum temperature of coldest month (bio6), mean temperature of driest quarter (bio9), mean temperature of coldest quarter (bio11), annual precipitation (bio12), precipitation of driest month (bio14), precipitation of the driest quarter (bio7), and precipitation of the coldest quarter (bio19). Unfortunately, neither of the gridded data sets provide daily or maximum climate variables. Topographic predictors (as raster layers), including slope, aspect, topographic wetness index (TWI), valley depth, and wind exposure, were calculated from the Shuttle Radar Topography Mission (SRTM) digital elevation model (Rodriguez et al., 2005), which has an original resolution of 30 m. Valley depth is a measure of the difference between the elevation and ridge level (Conrad et al., 2015). TWI is a measure of soil potential water content and flow accumulation (Kopecký and Čížková, 2010; Dyderski and Pawlik, 2020). Wind Exposition Index (WEI) calculates the average '*Wind Effect Index*' for all directions using an angular step. Values below 1 indicate wind shadow areas whereas values above 1 indicate areas exposed to wind (Boehner and Antonic, 2009). Normally this index is closely related to elevation. Slope and aspect were calculated according to Horn (1981). Vegetation properties (Table 1), including tree age, height, diameter at breast height, tree volume per forest unit (77,882 observations of 4 variables), were obtained from tree census data obtained in 1998 (Jawor and Świeradów forest inspectorates), 1999 (Bardo Śląskie, Kamienna Góra, Śnieżka, Szklarska Poręba, Wałbrzych forest inspectorates), 2000 (Bystrzyca Kłodzka, Lądek Zdrój, Międzylesie, Zdroje forest inspectorates), 2001 (Jugów, Lwówek Śląski, Świdnica, Złotoryja forest inspectorate) or 2006 (Pieńsk forest inspectorate).

We used the central point of each forest unit area to extract the relevant information for data that were available as raster layers, including the climate and topographic predictors. The spatial resolution, projection and extent of these raster layers were homogenized to create a raster stack prior to data extraction. For manipulation of the raster layers, we used several R geopackages: *sp*, *sf* (Pebesma, 2018), *rgdal*, and *raster* (Hijmans, 2020).

2.3. Explanatory data analysis

We used Principal Component Analysis (PCA) to identify the relationships between all continuous predictors and to establish the final set of potential explanatory variables using the *vegan* (Oksanen et al., 2016) R package (R Core Team, 2019). When two variables were highly correlated ($|r| > 0.75$), one was removed from the data set to avoid multicollinearity using the *caret* function *findCorrelation*(). This function calculates the

Table 1
Set of variables used for the present study.

Variable	Data format	Unit	Spatial resolution	Time period covered and data source	Number of observations
Response variable					
Damage class	Category	1 — “dam”; 0 — “no_dam”	–	2004–2010	464,328
Potential predictors					
1. Forest unit area	Continuous	ha		PIFR	
2. Aspect	Continuous	rad		From SRTM**	
3. Slope	Continuous	rad		From SRTM	
4. Elevation	Continuous	m a.s.l.	30 m	SRTM	
5. Topographic Wetness Index	Continuous	Nondimensional		From SRTM	
6. Valley depth	Continuous	m		From SRTM	
7. Wind Exposition Index	Continuous	Nondimensional		From SRTM	
8. Monthly mean wind speed	Continuous	m s^{-1}	30 s	1970–2000 (WorldClim)	
9. Annual mean temperature (bio1)	Continuous	°C*10	30 s	1979–2013 (CHELSA)	
10. Min temperature of coldest month (bio6)	Continuous	°C*10			
11. Mean temperature of driest quarter (bio9)	Continuous	°C*10			
12. Mean temperature of coldest quarter (bio11)	Continuous	°C*10			
13. Annual precipitation (bio12)	Continuous	mm			
14. Precipitation of driest month (bio14)	Continuous	mm			
15. Precipitation of driest quarter (bio17)	Continuous	mm			
16. Precipitation of coldest quarter (bio19)	Continuous	mm			
17. Tree age	Integer	years		1998–2006	77,882
18. Tree height	Integer	m			
19. Tree dbh	Integer	cm			
20. Forest stand volume	Integer	m^3			

* PIFR – Polish Institute of Forest Research.
** From SRTM – calculated based on SRTM.

mean absolute correlation of each variable to all other variables and removes the variable with the largest mean absolute correlation in any highly correlated pair (Kuhn, 2020). The initial analyses showed strong and positive correlation between mean monthly wind speeds, and wind speed was also correlated with elevation. Most of the bioclimatic variables were also highly correlated with one another. Removing the highly correlated variables resulted in the selection of 10 or 11 variables, depending on the analysis. Even after removing the highly correlated variables, there are some weak relationships between variables (Fig. 2). Slope and valley depth (vd) are related to tree age and volume, while topographic wetness index (TWI) is related to mean temperature of coldest month (bio6) and precipitation of coldest quarter (bio19).

2.4. Data preprocessing

The general workflow of data preprocessing and modelling is shown on Fig. 3. Data preprocessing and modelling was done using the *caret* R package (Kuhn, 2020).

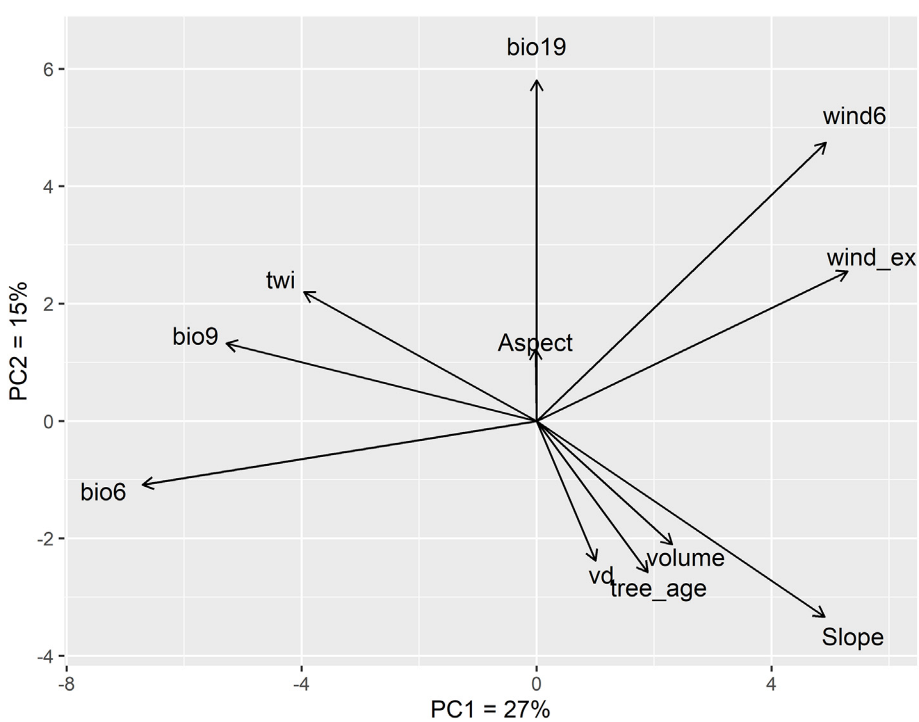


Fig. 2. PCA biplot for continuous variables included in the training dataset for Analysis 1. Only variables with correlation |r| < 0.75 were used. Abbreviations: bio19 – precipitation of coldest quarter, wind_exp – wind exposition index, twi – topographic wetness index, vd – valley depth, bio9 – mean temperature of driest quarter, bio6 – mean temperature of coldest month, wind – mean wind speed in June, volume – tree volume.

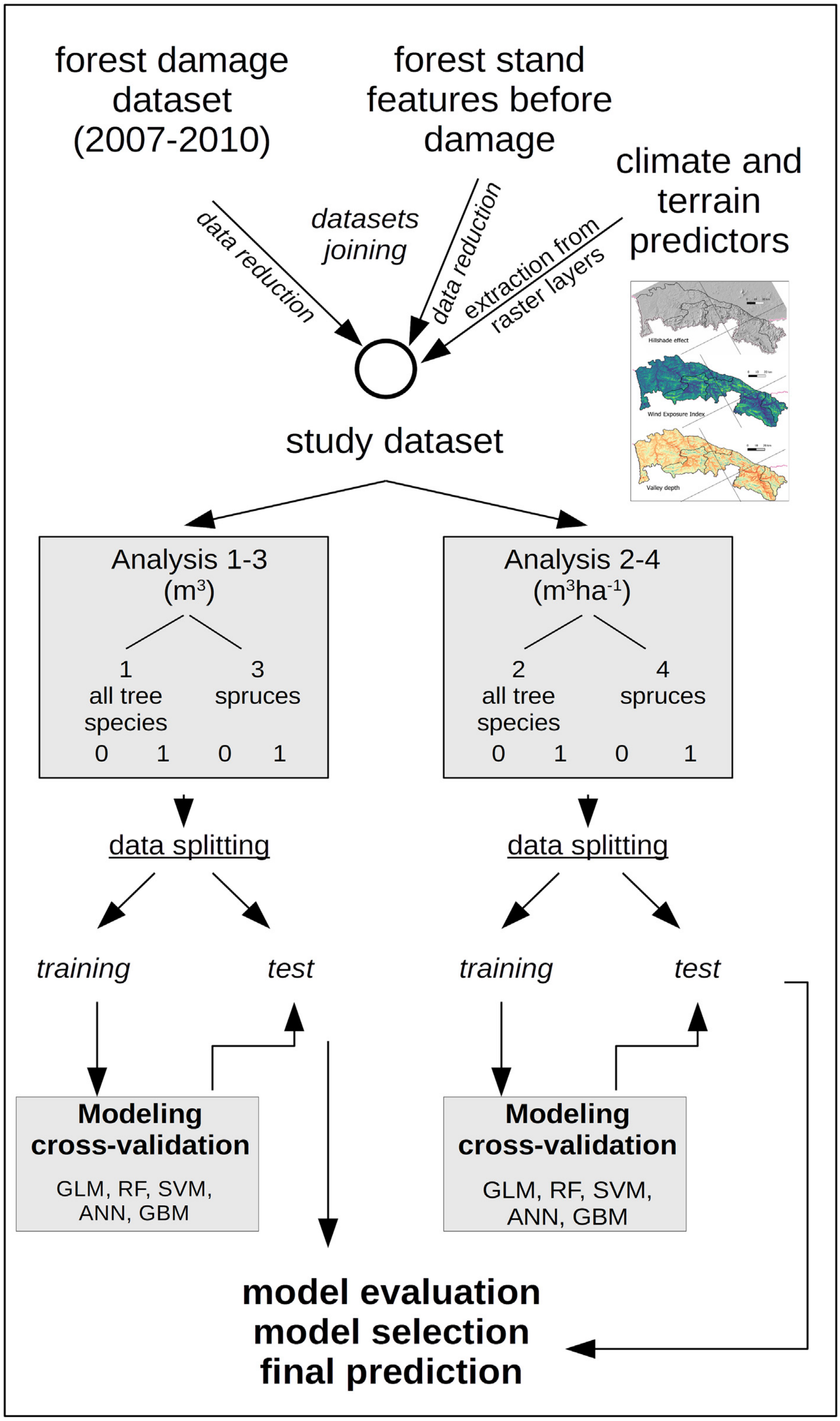


Fig. 3. Workflow adopted for data preparation and modelling in the present study.

The volume of damage response variable was initially expressed in m^3 and this was converted to volume of damage per unit area expressed in m^3 ha^{-1}. This ensures that differences in forest unit area do not affect the model. Units with no or low damage could be considered as sites that are less susceptible to wind damage, but on the other hand damage on such sites may be due to wind events that only affect single trees. We tested whether there was a difference in the response when we excluded sites with low levels of damage, using a threshold of 1 m^3 ha^{-1} which is typically considered as indicating severe damage (Jewuła, 1974; Vicena et al., 1979; Brázdil, 1998). We therefore performed model training and testing first on the complete data set and then on the data set after removing forest units where damage levels were <1 m^3 ha^{-1}. The majority of the forest units were dominated by Norway spruce (*Picea abies* L. Karst), i.e., 206,163 units (70.8% by volume, Fig. 2S, Appendix 1), although there were also substantial numbers of plots dominated by *Pinus sylvestris* (33,939 obs.), *Quercus* spp. (31,905 obs.), *Betula* (29,205 obs.), and *Fagus sylvatica* (23,463 obs.). Unfortunately, the wind damage data did not differentiate which species were most affected in a forest unit. However, we tested whether the uneven representation of units dominated by different species in the data set had an impact on the model, by re-running the training and testing using only forest units dominated by Norway spruce. The data for the other dominant species was insufficient to build robust models. This resulted in four model approaches: Analysis 1 uses all the data, Analysis 2 uses only units dominated by Norway spruce but makes no distinction with respect to damage level, Analysis 3 uses data where damage was >1 m^3 ha^{-1} but makes no distinction with respect to dominant species, and finally Analysis 4 uses only units dominated by Norway spruce where damage was >1 m^3 ha^{-1}.

Before modelling, we classified each record as a place with ("dam") or without ("no_dam") damage. Data were zero-inflated, since there were many more units that are not affected by wind damage than are affected by wind damage. We overcome the problem of zero-inflation by stratified downsampling using the *caret::downSample()* function. This function randomly samples the data set so that all classes have the same frequency, i.e. so that there are as many samples from the "no dam" class as from the "dam" class. Before training, the data were centered and scaled (Kuhn and Johnson, 2013). The dataset was split into a training set (60%) and a test set (40%). We used data for 2007–2010, when there were more observations. After splitting, we used 10-fold cross-validation resampling repeated 5 times (results from 50 different data sets were averaged and used for the evaluation of model efficacy), specified in the *caret::trainControl()* function, during model training to avoid over-fitting (Kuhn and Johnson, 2013). In order to find the best combination of hyperparameters the tuning procedure was performed through the *base::expand.grid()* function (Section 2.5). In addition to the cross-validation of the 2007–2010 training data set, we used the data from 2004 to 2006 as an additional validation of the models.

2.5. Data modelling

Different approaches may vary in terms of interpretability, computation time and predictive power. We tested five models that have been successfully applied to analyse other environmental dependencies (e.g. Regmi et al., 2014; Bistinas et al., 2014; Hengl et al., 2017; Dyderski and Jagodziński, 2019): logistic regression (GLM), random forest (RF), radial basis function support vector machines (SVM), neural networks (NN) and gradient boosted machine (GBM).

Logistic regression is a special case of generalized linear models (GLM) commonly used for solving classification and regression problems (Musa, 2013; Kuhn and Johnson, 2013; Regmi et al., 2014). A GLM does not require normally distributed predictors and can handle binary response variables (0 = false, 1 = true), but a prerequisite is the lack of correlation (collinearity) between covariates. In the present study, we use two definitions of the type 1 response: 1) damage >0 m^3 ("dam") and 2) "catastrophic damage" interpreted as damage ≥1 $m^3 ha^{-1}$ ("dam"). The method finds the probability of the event (1 – damage, 0 – no damage) through regression between a response (dependent) variable and independent variables (here mostly continuous covariates).

The random forest method (Breiman, 2001; Boulesteix et al., 2012; James et al., 2013; Probst and Boulesteix, 2017; Probst et al., 2019) is a tree-based algorithm supported by multiple decision or classification trees, which increases model stability and accuracy. When building a tree, a random sample of *m* predictors – *mtry* is taken at each split. The typical default is $mtry \approx \sqrt{p}$ for classification (Kuhn and Johnson, 2013). The use of small values of *mtry* helps to overcome collinearity between the predictors because only a fraction of observations is sampled at each split. Hyperparameter tuning automates the selection of the *min_n* and *mtry* parameters (Table 1S, Appendix 1). After hyperparameter tuning, we selected the best parameters using the measure of accuracy and ROC_AUC (*min_n* = 40, *mtry* = 2). Low values of *mtry* lead to more weakly correlated trees and, while this is desirable, the resulting trees could be based on sub-optimal variables if the sample size (i.e. the number of selected trees) is small (Probst et al., 2019). This, and the fact that the stability of the final RF model can also be affected by the choice of other hyperparameters, means it is important that the model is based on a sufficient number of weakly correlated trees (Probst and Boulesteix, 2017). We therefore set the number of trees to 1000, following Kuhn and Johnston (2013).

Support Vector Machine (SVM) is a technique based on a mathematical definition of the *hyperplane* that is a boundary between data point groups representing similar values (Boser et al., 1992; Brett, 2013). One of the steps in SVM is finding the maximum margin hyperplane that is the greatest separation between the binary classes (Boehmke and Greenwell, 2019; Thurnhofer-Hemsi et al., 2020). Separation can follow a linear or non-linear function (polynomial or radial). For non-linear transformation the so-called *kernel trick* is used, i.e. a function that can transform the data into higher dimensional space and allows separation. For the present classification problem, we tested the polynomial and radial basis function kernel (Ben-Hur and Weston, 2010; Kuhn and Johnson, 2013; Kuhn, 2020). Because SVM is a supervised classification method, it requires various arguments to be specified, including cost (C), degree (of polynomial function), scaling factor (for polynomial kernel), and gaussian sigma σ (for radial kernel) (Table 1S, Appendix 1). C is the regularization parameter that controls the trade-off between minimizing the training set error and maximizing the hyperplane margin. The degree of the polynomial kernel controls the flexibility of the resulting classifier. Since we obtained similar results with the polynomial and radial kernel, we adopted the radial kernel because it is less computationally demanding.

Monotone Multi-Layer Perceptron (MONMLP) Neural Network, implemented in the *caret* package through the *monmlp* package (Cannon, 2017), is a form of artificial neural network (ANN). ANN is a 'black box' approach (Lek and Guégan, 1999) able to learn from incomplete, disturbed and noisy datasets (Hanewinkel et al., 2004). MMPNN is a popular and powerful neural network type with architecture consisting of non-linear elements (neurons) arranged in successive layers, and a hidden layer that is used for unidirectional flow of the information from input to output layer (Lek and Guégan, 1999; Kuhn and Johnson, 2013). Thus, there are two tunable hyperparameters: the number of hidden layers and the number of neurons. We trained the ANN models with one hidden layer and up to 3 neurons (Table 1S, Appendix 1). It is desirable to limit the number of neurons because using too many neurons can result in overfitting (Heaton, 2008). It has been shown that one hidden layer is sufficient for most problems (Heaton, 2008; Bochenek et al., 2021).

The gradient boosting machine (GBM) is based on an ensemble of decision tree models, where the initial model is boosted by adding new models ("weak learners") to the ensemble sequentially to improve accuracy (Friedman, 2001; Natekin and Knoll, 2013; Touzani et al., 2018). This allows the reduction of the model variance through averaging several decision trees. New weak learners are based on the residuals of previous models. We used the search grid method to estimate the following hyperparameters: 1) interaction depth, 2) shrinkage, 3) minimum number of observations at node, 4) number of trees (Table 1S, Appendix 1). The shrinkage parameter α, also called the learning rate, can take values between 0 and 1, when lower values imply a higher number of small steps and higher accuracy (Friedman, 2001; Natekin and Knoll, 2013; Touzani et al., 2018). Both RF and GBM use an ensemble of decision trees. However, there are two general differences between RF and GBM: 1) in RF trees are combined at the end of the process of model building, whereas in GBM trees are added at each iteration, i.e., when a new tree is grown, it uses information from a previous model (Boehmke and Greenwell, 2019), 2) in RF many deep independent trees are grown, whereas GBM uses shallow trees that individually are considered as weak predictive models (Boehmke and Greenwell, 2019).

For all the models, we used model prediction strength accuracy and the area under (AUC) the receiver operating characteristic (ROC) curve as metrics. The accuracy is calculated as the total number of correct predictions (for both classes of the response variable) divided by the total number of predictions. The ROC curve computes the sensitivity and specificity over a continuum of different event thresholds (Kuhn and Silge, 2020). The area under the ROC curve (roc_auc) is a measure of probability of correctly identifying a positive signal (here 'dam') above the noise (here 'no_dam') (Hanley and McNeil, 1982). The threshold of probability level used for discrimination between a positive and negative class is typically set to 0.5. We adopt the following interpretation of *roc_auc*: roc_auc < 0.5 indicates worse than random discrimination, roc_auc = 0.5 indicates random discrimination between binary classes, roc_auc of (0.5, 0.7] indicates weak discrimination, roc_auc of (0.7, 0.8] indicates acceptable discrimination, roc_auc of (0.8, 0.9] indicates excellent discrimination, roc_auc of (0.9, 1.0) indicates outstanding discrimination, and roc_auc = 1.0 indicates perfect discrimination between classes (Suvanto et al., 2019; Hosmer et al., 2013).

The impact of individual variables on model performance was evaluated by calculating feature importance and using accumulated local effects (ALE) plots (Molnar et al., 2018; Molnar, 2020). Feature importance is calculated using a permutation method that randomly permutes the values of each predictor variable in the training data set and computes the associated reduction in model performance or increase of the predictive error (Boehmke and Greenwell, 2019). ALE are computed as accumulated differences in predictions over the conditional distribution (Molnar, 2020). Feature importance was calculated using the *varImp()* function and the ALE was assessed with *FeatureEffects()* function from *iml* package (Molnar et al., 2018).

Finally, we used data from two periods to examine spatial patterns of the probability of forest damage: 1998–2006, and 2020 (see Section 2.6 for details). Raster layers with probability were calculated with the *raster::predict()* function from the *raster* package. All analyses were conducted in R 4.0.2 (R Core Team, 2019). The final maps were constructed using QGIS 3.14.

2.6. Prediction prerequisites

Models with the highest predictive scores were used for prediction of potential forest damage probability using the most recent forest inventory data. The data were downloaded from the Forest Data Bank (FDB, https://www.bdl.lasy.gov.pl/portal/). Data in FDB is updated annually based on information received from forest inspectorates; the last update of data on forest features for all inspectorates was 2020. The most important information was tree volume and age. It was combined with a vector layer and converted to raster layers. Other predictors were assumed to be constant. From these layers a raster stack was built and used for prediction using the most optimal models.

3. Results

All of the models have similar predictive power for any given scenario (Table 2), with AUC values close to 0.7 and accuracy >0.65 (Fig. 3S, Appendix 1). Although the GBM model produces marginally better results and the GLM model marginally worse results, the similarity across the models indicates that the results are not dependent on the choice of machine-learning approach used.

The models using the dataset with all tree species (Analysis 1–2) are better than those obtained for data with only spruce (Analysis 3–4). There is little degradation of the scores when using the test data from 2004 to 2006 (Table 2S, Fig. 3S, Appendix 1). However, the scores for the tests based on individual years are less good overall (Table 2S). The consistency across models and across data sets shows that it is possible to predict the factors influencing forest damage with some confidence.

Generally, tree characteristics were the most important variables in predicting the level of damage. Tree volume is the most important predictor in all models and in all scenarios (Figs. 4 and 5). Tree age is the second most important variable for all analyses, except in the GBM model in Analyses 1, 2 and 3. Except in the GLM case, there is a large drop in predictive importance between these vegetation-related variables and the bioclimatic or terrain variables. Furthermore, the relative importance of the bioclimatic and terrain variables changes between models and scenarios. Mean wind speed in June (wind6), precipitation of coldest quarter (bio19) and mean temperature of driest quarter (bio9) are the most important bioclimatic predictors, with either wind6 or bio19 being the most important bioclimatic predictor for Analysis 1 and 2 across all the models. However, mean wind speed in June is not the most important predictor in Analysis 2 and, although precipitation of coldest quarter (bio19) remains important, the mean temperature of driest quarter (bio9) and minimum temperature of coldest month (bio6) are the most important bioclimatic variable in some of the models under these scenarios. Wind exposure is the most important topographic variable in half of the models under Analyses 1 and 2, although slope and valley depth are the most important variable in the other models (4 and 2 cases respectively). However, wind exposure is never the most important topographic variable in Analyses 3 and 4. Instead, TWI is chosen equally often as slope (4 cases). This suggests that the topographic variables convey similar information about susceptibility such that there is no clear value in using one predictor over another.

The ALE plots show the nature of the relationship between individual variables and wind damage. Tree volume has a positive effect on damage up to a threshold value of 5000 m^3 per forest unit, after which there is no further change (Fig. 6, see Fig. 4S–6S in Appendix 1 for other models). The probability of forest damage decreases with tree age up until ca 25 years, then increases up to 100 years. The apparent decrease in damage with trees older than 100 years may reflect under-sampling (Fig. 6, RF model). Wind exposure and valley depth have a generally positive effect on damage, whereas slope has a generally negative effect on damage. TWI has a positive effect on damage up to 7.0 and a negative impact above this value. The ALE plots show that west-facing slopes have the highest probability of damage. Precipitation of the coldest quarter (bio19) has a strong negative effect on the probability of forest damage up to ca 150 mm but has no impact above this level. The impact of mean

Table 2
Metrics of accuracy (acc) and ROC AUC (auc) for each trained model and their evaluation based on test sets. The highest values are indicated by bold type.

Analysis	Model data input		Binomial GLM		Random forest		SVM radial kernel		Neural network		Gradient boosting machine	
			acc	auc	acc	auc	acc	auc	acc	auc	acc	auc
1	m^3	Training	0.637	0.687	0.649	0.709	0.654	0.709	0.651	0.708	0.656	**0.714**
	All tree species	Test	0.639	0.691	0.654	0.715	0.654	0.711	0.652	0.710	0.657	**0.717**
2	$m^3\ ha^{-1}$	Training	0.625	0.672	0.642	0.697	0.643	0.696	0.638	0.690	0.644	**0.699**
	obs. > 1 $m^3\ ha^{-1}$ All tree species	Test	0.619	0.667	0.642	**0.701**	0.641	0.694	0.637	0.689	0.644	**0.701**
3	m^3, only spruces	Training	0.631	0.677	0.641	0.698	0.645	0.698	0.646	0.700	0.648	**0.704**
		Test	0.634	0.681	0.644	0.701	0.648	0.701	0.648	0.703	0.651	**0.706**
4	$m^3\ ha^{-1}$	Training	0.618	0.664	0.637	0.692	0.636	0.689	0.632	0.687	0.639	**0.694**
	Obs. > 1 $m^3\ ha^{-1}$ Only spruces	Test	0.620	0.662	0.637	0.688	0.633	0.686	0.634	0.686	0.641	**0.689**

Analysis 1 – all tree species, damage in m^3, damage >0 m^3.
Analysis 2 – all tree species, damage in $m^3\ ha^{-1}$, damage >1 $m^3\ ha^{-1}$.
Analysis 3 – only Norway spruce, damage in m^3, damage >0 m^3.
Analysis 4 – only Norway spruce, damage in $m^3\ ha^{-1}$, damage >1 $m^3\ ha^{-1}$.

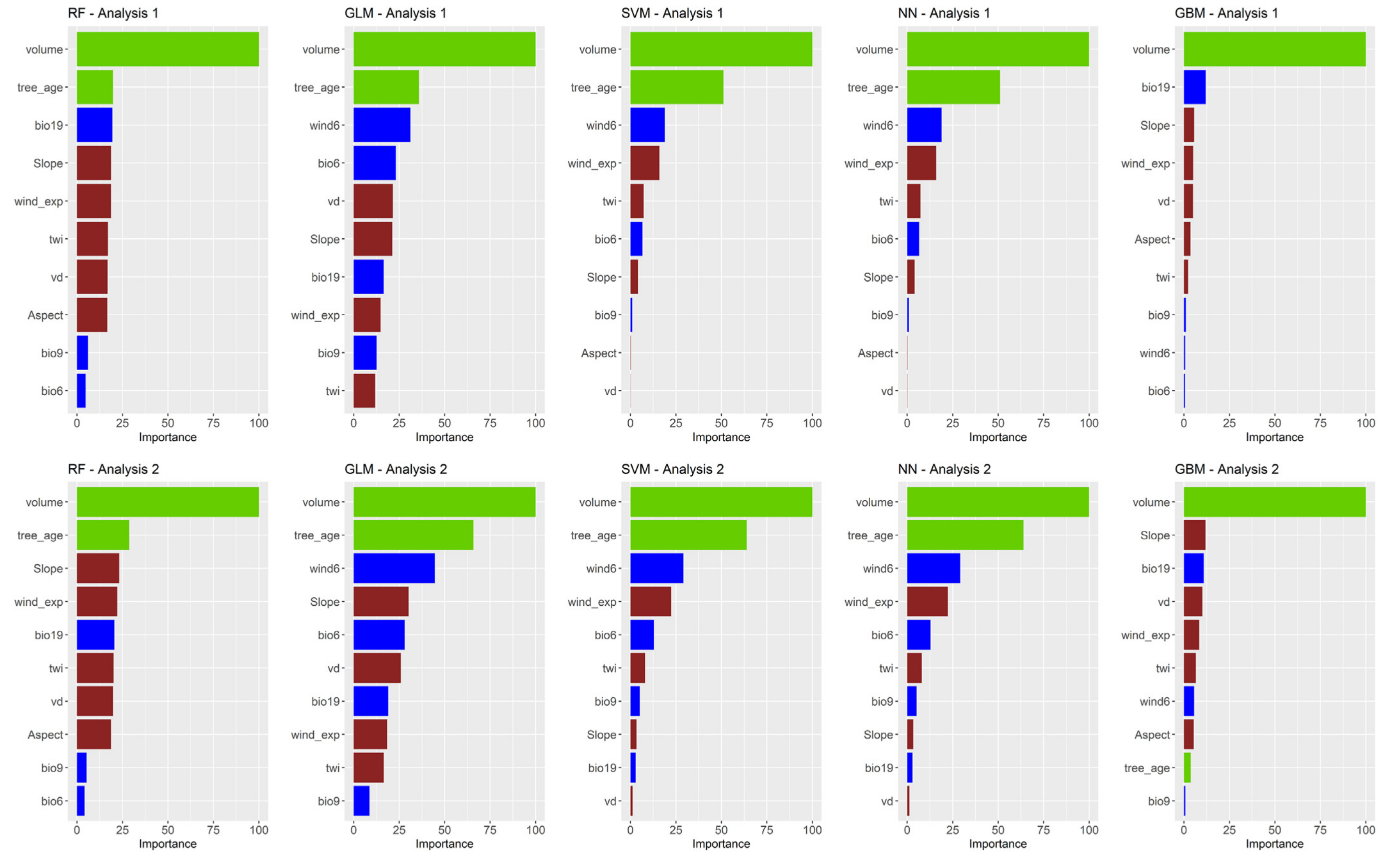


Fig. 4. Importance of variables for different models applied under Analyses 1 and 2 where all tree species are taken into account and damage volume is expressed in m^3 (A1) or $m^3\ ha^{-1}$ (A2). Colours indicate general type of variable: tree features – green; climate features – blue; terrain features – brown. Abbreviations: bio19 – precipitation of coldest quarter, wind_exp – wind exposition index, twi – topographic wetness index, vd – valley depth, bio9 – mean temperature of driest quarter, bio6 – mean temperature of coldest month.

temperature of the driest quarter (bio9) shows an initial decrease in probability until ca -2.5 °C and then an increase. As expected, wind speed has a positive effect on damage probability. These patterns are similar in the ALE plots for other models and scenarios (Appendix 1, Figs. 4S–6S).

In constructing prediction maps, we used data on tree volume and age based on forest census data for 1998–2006 (map A) and compared this with 2020 (map B), assuming that climate- and terrain-based variables do not change in the short-term. All the models predict that a larger proportion of Sudety forests had a higher probability of damage in the period 1998–2006 compared to 2020 (Fig. 7, and Figs. 7S–9S in Appendix 1). The GBM, GLM, and SVM models show the biggest difference between the two periods. Although, the probability of damage decreased for all models, the regions with a high probability of damage in the two periods are the same.

4. Discussion and conclusions

Gradients of wind speed along storm paths result in gradients of damage to trees and forest stands (Martin, 2013). However, spatially and temporally complete information on wind speed and direction allowing precise prediction of forest damage is often difficult to obtain and assessments of forest vulnerability to strong winds must therefore rely on less direct measures. It is likely that the use of average monthly wind speed, rather than daily maximum wind speeds, leads to an underestimate of the importance of wind speed and duration on forest damage in our analyses. This is supported, indirectly, by the fact that wind exposure is an important predictor. However, although wind speed and duration are likely to be important influences, our results show that forest properties (tree volume, age) have the major influence on the amount of damage sustained.

A similar conclusion was reached in analyses of forest damage by wind in Finland (Suvanto et al., 2016, 2019), Sweden (Fridman and Valinger, 1998; Valinger and Fridman, 2011), and Pennsylvania, USA (Evans et al., 2007a, 2007b). Hanewinkel et al. (2013) also concluded that tree and stand characteristics are more important than site characteristics. Several studies have related the importance of tree and forest characteristics to mechanistic features that allow trees to withstand the impact of strong wind (Peltola and Kellomäki, 1993; Gardiner et al., 1997). The study by Gardiner et al. (1997) on Sitka spruce, for example, indicated increasing resistance to breakage with increasing tree diameter driven by wider spacing between trees at the stand level, however these trees may be more vulnerable to uprooting. Stand openness and thinning have been seen as factors increasing forest damage (Pellikka and Järvenpää, 2003), but this is not apparent in our analyses which show that high volume forests (>5000 m^3) sustain more damage. Dobbertin (2002) analysed the damage caused by windstorms *Vivian* (in 1990) and *Lothar* (in 1999) in Switzerland and found forest stand features such as stand height, development stage, and percentage of conifers, influencing higher probability of damage. Hart et al. (2019) suggested that stand characteristics such as density and canopy roughness might be more important than tree characteristics such as age or height during severe storms and that tree features do not modify the effective forest vulnerability to the wind.

Our analyses indicate that higher age-class forests are more susceptible than either young forests (trees <20 years old) or forests with large numbers of old trees (>100 years). Evans et al. (2007a, 2007b) found the highest damage in the tree age class of 80–100 years, similar to our results. Stand age was found to increase damage probability in southern Sweden as a result of the *Gudrun* winter storm (Valinger and Fridman, 2011), with 90 year old Norway spruce stands being twice as susceptible to damage as 50 year old stands of the same species. Stand height was an important

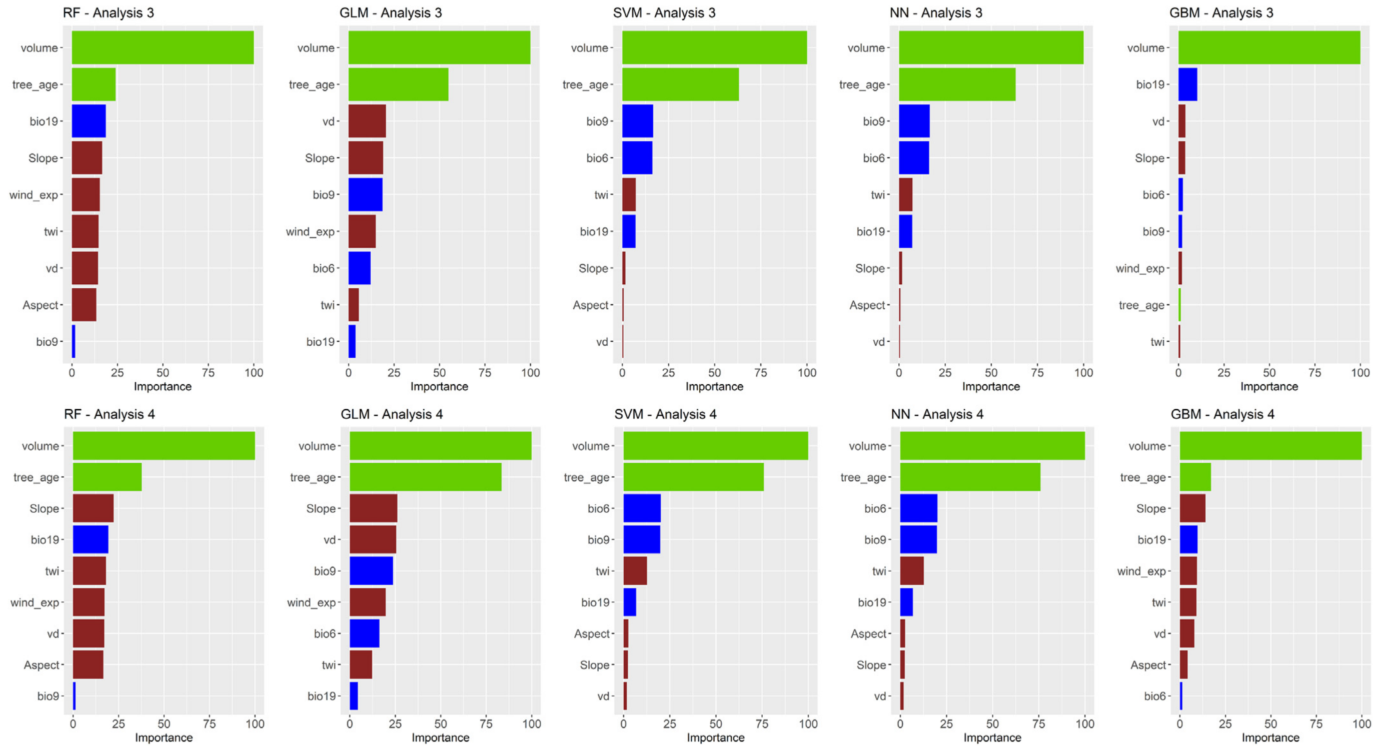


Fig. 5. Importance of variables for different models applied under Analysis 3 and 4 where only Norway spruce is taken into account and volume of damage is expressed in m^3 (A3) or m^3ha^{-1} (A4). Colours indicate general type of variable: tree features – green; climate features – blue; terrain features – brown. Abbreviations: bio19 – precipitation of coldest quarter, wind_exp – wind exposition index, twi – topographic wetness index, vd – valley depth, bio9 – mean temperature of driest quarter, bio6 – mean temperature of coldest month.

predictor of the damage after *Gudrun*, but since stand height is closely related to tree age and DBH we excluded this as a predictor in our study. There may be two reasons for the apparent reduction in damage with trees more than 100 years old shown in our study. Firstly, there is a reduction in sample size because most of the analysed forests are managed and older trees have been harvested. In Poland, *P. sylvestris* is generally harvested at ca 80 years and *P. abies* at ca 110 years. This is shown in Fig. 10S as a fast drop in *P. abies* age density distribution in 2000 (years 1998–2006). Secondly, any remaining older trees (in strictly and partly protected forests) may have survived because they are adapted to extreme climate conditions such as strong wind. A similar conclusion was drawn by Valinger and Fridman (2011).

Climate and terrain variables appear to play a less important role than forest characteristics in explaining susceptibility to damage. Although wind speed (represented by wind6) is clearly important, winter precipitation appears to have a role in reducing forest damage under relatively dry conditions. On the other hand, the impact of temperature during the driest period (bio9) generally has a positive impact on damage when temperatures are above zero. Other studies have suggested that wind damage is more common on saturated soils (Hanewinkel et al., 2013, 2014), but this does not seem to be supported by our results. We observe that only a tiny fraction of the TWI increase (between 6 and 7 units) impacts more significant forest damage. However, it has been shown that increasing TWI (interpreted as higher soil moisture and potential water flow accumulation) has a negative impact on tree biomass of forest stands in the national mountain parks in Poland (Dyderski and Pawlik, 2020). However, when TWI reaches 6–7 units the effect is positive for instance in the forests of the Tatra Mountains National Park (Dyderski and Pawlik, 2021). This suggests a close link between variables driving volume of forest damage and tree biomass but requires further testing to make any solid conclusion.

The general increase in damage with exposure, and the fact that west-facing slopes sustain most damage, is not surprising (Hanewinkel et al., 2014). Regional studies show that the prevailing wind direction in the Sudety Mountains is from the W and SW sector (Schmuck, 1969; Kwiatkowski and Hołdys, 1985; Sobik et al., 2013). However, the decrease in damage with increasing slope seems unusual (Fig. 6) compared to other studies. For instance, Kramer et al. (2001) found a positive relationship between damage area and slope in the temperate rain forest in southeast Alaska. However, Evans et al. (2007) found that half of the damage was on the flatter terrain which seems to indicate that slope has only a weak effect on damage. Klaus et al. (2011) found a decrease of damage probability with slope when applying a logistic regression model to wind damage caused by the *Kyrill* windstorm in January 2007. Our data includes damage caused by the *Kyrill* windstorm but is an ensemble of records caused by wind of different intensity and origin, not only winter storms. This may explain why the impact of slope is less than shown by other studies. This also applies to other topographic features, e.g., TWI. Hanewinkel et al. (2014), examining the impact of windstorm *Lothar* (in December 1999) in Switzerland, found a decreasing probability of damage with increasing slope for eastern exposed stands but more severe storm damage on steep slopes for western exposed stands. There is no clear explanation of our results related to the slope. However, in all slope classes plots with damage occur more frequently than plots without damage (Fig. 11S, Appendix 1). The PCA biplot (Fig. 2) shows that the slope variable is weakly correlated with tree volume, age, and valley depth (vd). The geomorphology of the Sudety Mountains is highly complex, and this is reflected in the wide range of the slope parameter (Placek, 2011). Additionally, Hanewinkel et al. (2014) reported that a combination of aspect and slope in a multivariate statistical model reveals lower damage on steeper slopes during winter storms. As pointed by the authors, this effect was frequently observed (e.g., Klaus et al., 2011).

Wind Exposition Index (WEI) and valley depth impact on forest damage in the Sudety Mountains follow our initial hypotheses, but other authors rarely used these parameters. Because WEI is closely correlated with elevation and wind speed, it can capture the effect of these two variables. Valley depth gives information on terrain complexity and can reach high values in

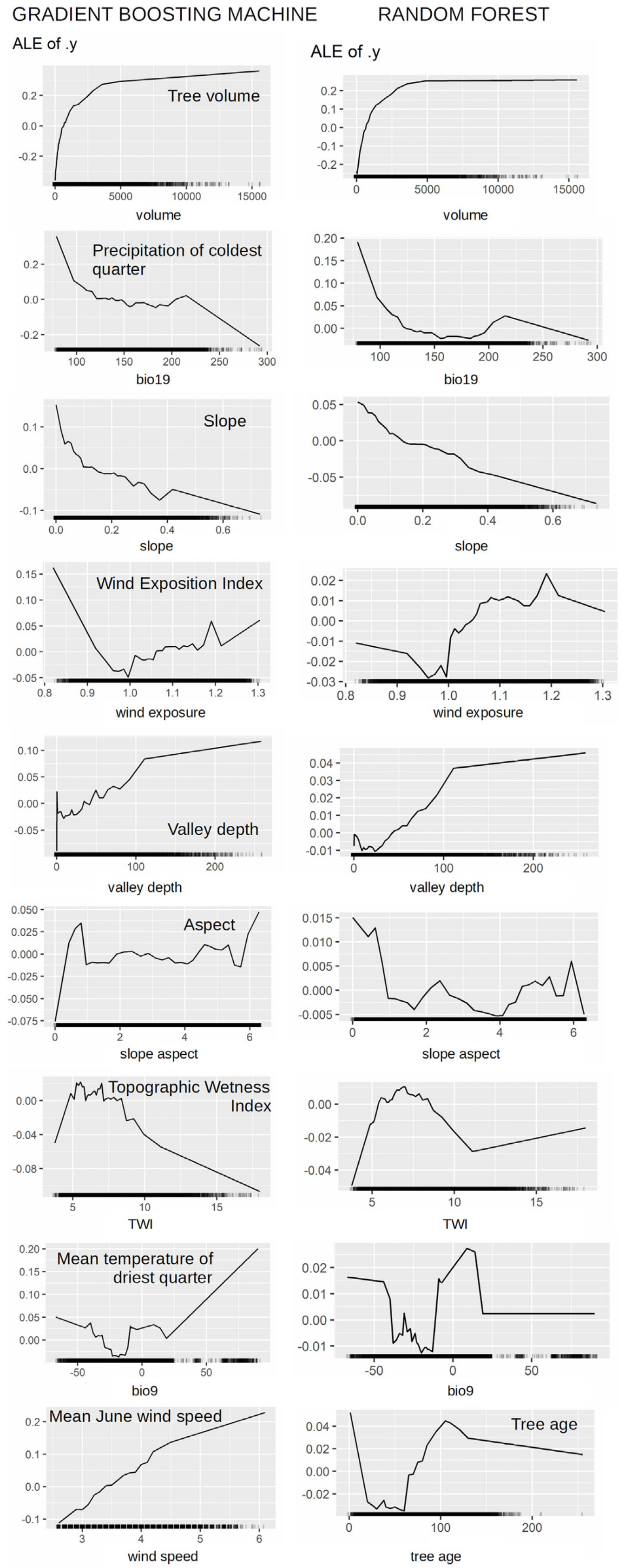


Fig. 6. Accumulated local effects (ALE) plots of selected features having impact on predictors when GBM and RF model is applied under Analysis 1. The density of feature distribution is shown on the x-axis. Regions with low density should be interpreted with caution. The temperature is multiplied by 10 (see Karger et al., 2017).

landscapes dissected by faults and incised by deep river valleys and tectonic troughs. Long and deep valleys are important for bora wind formation, one of the orographic wind types (Grisogono and Belušić, 2008). It can reach high speed and be responsible for catastrophic damage in forest stands as, for instance, in the Tatra Mountains, Slovakia, in 2004 (Kopecka, 2011). Our data do not include information on wind type, but we assume some fraction of the damage may have been caused by bora wind, for which valley depth can be a key driving factor. Our results show that increasing valley depth, over 50 m deep, positively affects damage probability. A similar effect is seen with the nondimensional WEI (over 1.0), but this index expresses different terrain properties, i.e., elevated ridges and isolated massifs.

All of the models used here have a comparable level of predictive power, although the GBM model produces slightly better results and the GLM model marginally worse results. Other studies have shown that logistic regression models perform better than other types of models and had higher predictive power than obtained for the Sudety region (Fridman and Valinger, 1998; Schindler et al., 2009; Klaus et al., 2011; Suvanto et al., 2019), although this is not always the case (see e.g. Hanewinkel et al., 2004; Hart et al., 2019). So far logistic regression has been the most frequently applied algorithm (Klaus et al., 2011; Hanewinkel et al., 2014; Suvanto et al., 2019). The similarity of the results across models and scenarios in our study, and the fact that the relative importance of predictors of wind damage is in agreement with findings from other studies, suggests that the choice of model is not important. However, when applied to unseen data from 2004 to 2006 there was a slight better performance of RF in Analysis 1 (Table 2S, Fig. 3S, Appendix 1). Nonetheless, in most cases AUC values are close to 0.7 for A1.

Despite the fact that ongoing climate change is expected to increase the incidence of wind damage in European forests (Lindner and Rummukainen, 2013; Gregow et al., 2017; Seidl et al., 2017), our analyses suggest that the probability of damage in the Sudety was lower in 2020 than during the interval between 1998 and 2006 as a result of change in forest structure. This does not rule out the possibility that changes in the incidence or strength of windstorms in a warming climate could offset this. However, it does suggest that forest management could play an important role in shaping the response of European forests to climate change.

Other studies have shown that changes in both forest characteristics and storminess have contributed to increased wind damage in recent decades across Europe (e.g. Schelhaas et al., 2003; Seidl et al., 2011; Gregow et al., 2017). Schelhaas et al. (2003) and Seidl et al. (2011) have shown that increasing standing timber volume and the promotion of conifers as dominant tree species has made forests more vulnerable to damage. Thus, the preferential planting of *P. abies.,* a fast-growing tree that is highly susceptible to wind damage (Everham and Brokaw, 1996; Gardiner et al., 2010; Mitchell, 2013), at all elevations in the Sudety would be consistent with high levels of damage. Indeed, our analyses indicate lower probability of forest damage in 2020 than during the interval 1998–2006. This reduction of probability in 2020 compared to the earlier period corresponds to a 9% decrease in the volume of *P. abies* (between 2006 and 2020, Fig. 2S). This clearly points to a role for forest management in mitigating any climate change impacts. It would, however, be useful to use the predictive modelling approach developed here to examine the degree to which the recent reduction in damage will be sustained under future climate change scenarios and also to explore how changes in management could offset the deleterious impact of such changes.

CRediT authorship contribution statement

Łukasz Pawlik: Conceptualization, Methodology, Data curation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Sandy P. Harrison:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Visualization, Supervision.

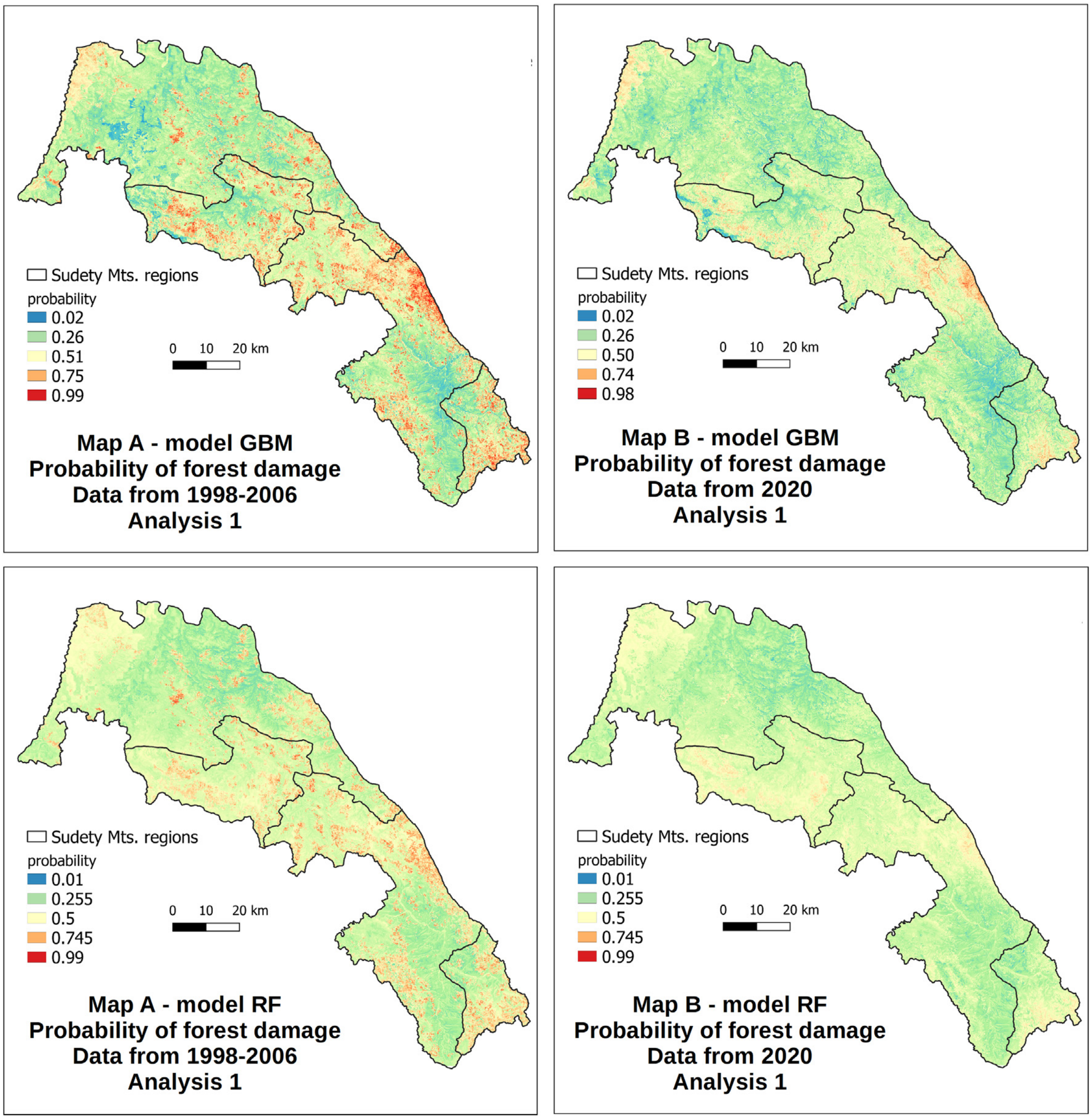


Fig. 7. Probability maps of forest damage based GBM and RF for data from two periods: 1998–2006 (Map A), and 2020 (Map B), under Analysis 1.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Łukasz Pawlik was supported by the National Science Centre project under the UWERTURA 2 funding scheme (project no. 2018/28/U/ST10/00075). Sandy P. Harrison acknowledges support from the ERC-funded project GC 2.0 (Global Change 2.0: Unlocking the past for a clearer future; grant number 694481). We would like to thank Marcin K. Dyderski (Institute of Dendrology, Polish Academy of Sciences) for suggestions and discussion. We thanked providers and administrators of PLGrid. A large part of modelling for this project was supported by PLGrid Infrastructure (http://www.plgrid.pl/). Maciej Kryza (Wrocław University, Poland) is thanked for providing access to the computer cluster of the Department of Climatology and Atmosphere Protection. We also acknowledge his technical support during the project. Lastly, we want to thank three anonymous reviewers whose many comments and suggestions allow for a significant improvement of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2021.151972.

References

Ben-Hur, A., Weston, J., 2010. A user's guide to support vector machines. In: Carugo, O., Eisenhaber, F. (Eds.), Data Mining Techniques for the Life Sciences, Methods in Molecular Biology. 609. https://doi.org/10.1007/978-1-60327-241-4_13.

Bistinas, I., Harrison, S.P., Prentice, I.C., Pereira, J.M.C., 2014.. Causal relationships versus emergent patterns in the global controls of fire frequency. Biogeosciences 11, 5087–5101. https://doi.org/10.5194/bg-11-5087-2014.

Bochenek, B., Jurasz, J., Jaczewski, A., Stachura, G., Sekuła, P., Strzyżewski, T., Wdowikowski, M., Figurski, M., 2021. Day-ahead wind power forecasting in Poland based on numerical weather prediction. Energies 14, 2164. https://doi.org/10.3390/en14082164.

Boehmke, B., Greenwell, B.M., 2019. Hands-On Machine Learning with R. The R series. 488 ppCRC Press, A Chapman & Hall Book. https://bradleyboehmke.github.io/HOML/.

Boehner, J., Antonic, O., 2009. Land-surface parameters specific to topo-climatology. In: Hengl, T., Reuter, H. (Eds.), Geomorphometry - Concepts, Software, Applications. Developments in Soil Science. 33. Elsevier, pp. 195–226.

Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers. In: Haussler, D. (Ed.), 5th Annual ACM Workshop on COLT. Pittsburgh, PA, ACM Press, pp. 144–152.

Boulesteix, A.-L., Janitza, S., Kruppa, J., König, I.R., 2012. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. WIRES 2 (6), 493–507.

Brázdil, R., 1998. Meteorological extremes and their impacts on forests in the Czech Republic. In: Beniston, M., Innes, J.L. (Eds.), The impacts of climate variability on forests. Springer, Berlin-Heidelberg-New-York, pp. 19–47.

Brázdil, R., Dobrovolný, P., Štekl, J., Kotysa, O., Valášek, H., Jež, J., 2004. History of weather and climate in the Czech Lands VI: Strong winds. Masaryk University, Brno.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Brett, L., 2013. Machine learning with R. Packt Pub.

Grisogono, B., Belušić, D., 2008. A review of recent advances in understanding the meso- and microscale properties of the severe bora wind. Tellus Ser. A Dyn. Meteorol. Oceanogr. 61A, 1–16.

Cannon, A.J., 2017. monmlp: Multi-Layer Perceptron Neural Network with Optional Monotonicity Constraints. R package version 1.1.5. https://CRAN.R-project.org/package=monmlp.

Coates, K.D., Hall, E.C., Canham, C.D., 2018. Susceptibility of trees to windthrow storm damage in partially harvested complex-structured multi-species forests. Forests 9 (199). https://doi.org/10.3390/f9040199.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for automated geoscientific analyses (SAGA) v. 2.1.4. Geosci. Model Dev. 8, 1991–2007. https://doi.org/10.5194/gmd-8-1991-2015.

Dmyterko, E., Mionskowski, M., Bruchwald, A., 2015. Risk of the wind damage to the forests in Poland on the basis of a stand damage risk model (in Polish, with English abstract). Sylwan 159, 361–371.

Dobbertin, M., 2002. Influence of stand structure and site factors on wind damage comparing the storms Vivian and Lothar. For. Snow Landsc. Res. 77, 187–205.

Dyderski, K., Jagodziński, A.M., 2019. Seedling survival of Prunus serotina ehrh., Quercus rubra L. and Robinia pseudoacacia L. in temperate forests of Western Poland. For. Ecol. Manag. 450. https://doi.org/10.1016/j.foreco.2019.117498.

Dyderski, M.K., Pawlik, Ł., 2020. Spatial distribution of tree species in mountain national parks depend on geomorphology and climate. For. Ecol. Manag. 474, 118366. https://doi.org/10.1016/j.foreco.2020.118366.

Dyderski, M.K., Pawlik, Ł., 2021. Drivers of forest aboveground biomass and its increments in the Tatra Mountains after 15 years. Catena 205. https://doi.org/10.1016/j.catena.2021.105468.

Edwards, R., LaDue, J.G., Ferree, J.T., Scharfenberg, K., Maier, C., Coulbourne, W.L., 2013. Tornado intensity estimation: past, present, and future. Bull. Am. Meteor. Soc. 94 (5), 641–653.

Evans, A.M., Camp, A.E., Tyrrell, M.L., Riely, C.C., 2007. Biotic and abiotic influences on wind disturbances in forests of NW Pennsylvania, USA. For. Ecol. Manag. 245, 44–53.

Everham, E.M., Brokaw, N.V., 1996. Forest damage and recovery from catastrophic wind. Bot. Rev. 62 (2), 113–185.

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37 (12), 4302–4315.

Fink, A.H., Brücher, T., Ermert, V., Krüger, A., Pinto, J.G., 2009. The european storm Kyrill in January 2007: synoptic evolution meteorological impacts and some considerations with respect to climate change. Nat. Hazards Earth Syst. Sci. 9, 405–423.

Evans, A.M., Camp, A.E., Tyrrell, M.L., Riely, C.C., 2007. Biotic and abiotic influences on wind disturbance in forests of NW Pennsylvania, USA. For. Ecol. Manag. 245, 44–53.

Fink, A.H., Brücher, T., Ermert, V., Krüger, A., Pinto, J.G., 2009. The european storm kyrill in january 2007: synoptic evolution, meteorological impacts and some considerations with respect to climate change. Nat. Hazards Earth Syst. Sci. 9 (2), 405–463.

Fridman, J., Valinger, E., 1998. Modelling probability of snow and wind damage using tree, stand, and site characteristics from Pinus sylvestris sample plots. Scand. J. For. Res. 13, 348–356.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 1189–1232.

Gardiner, B., 2021. Wind damage to forests and trees: a review with an emphasis on planted and managed forests. J. For. Res. https://doi.org/10.1080/13416979.2021.1940665.

Gardiner, B.A., Stacey, G.R., Belcher, R.E., Wood, C.J., 1997. Field and wind tunnel assessments of the implications of respacing and thinning for tree stability. Forestry 70, 233–252.

Gardiner, B., Blennow, K., Carnus, J.M., Fleischner, P., Ingemarson, F., Landmann, G., Lindner, M., Marzano, M., Nicoll, B., Orazio, C., Peyron, J.L., Reviron, M.P., Schelhaas, M.-J., Schuck, A., Spielmann, M., Usbeck, T., 2010. Destructive storms in European Forests: Past and Forthcoming Impacts, Final report to European Commission - DG Environment, Joensuu. European Forest Institute, Finland. https://ec.europa.eu/environment/forests/pdf/STORMS%20Final_Report.pdf.

Gardiner, B., Byrne, K., Hale, S., Kamimura, K., Mitchell, S.J., Peltola, H., Ruel, J.C., 2008. A review of mechanistic modelling of wind damage risk to forests. Forestry 81 (3), 447–463. https://doi.org/10.1093/forestry/cpn022.

Gardiner, B., Schuck, A., Schelhaas, M.-J., Orazio, C., Blennow, K., Nicoll, B., 2013. Living with storm damage to forests: What science can tell us. European Forest Institute, Joensuu.

Gregow, H., Laaksonen, A., Alper, M.E., 2017. Increasing large scale windstorm damage in Western, central and northern European forests, 1951–2010. Sci. Rep. https://doi.org/10.1038/srep46397.

Hale, S., Gardiner, B., Peace, A., Nicoll, B., Taylor, P., Pizzirani, S., 2015. Comparison and validation of three versions of a forest wind risk model. Environ. Model. Softw. 68, 27–41. https://doi.org/10.1016/j.envsoft.2015.01.016.

Hanewinkel, M., Zhou, W., Schill, C., 2004. A neural network approach to identify forest stands susceptible to wind damage. For. Ecol. Manag. 196, 227–243. https://doi.org/10.1016/j.foreco.2004.02.056.

Hanewinkel, M., Albrecht, A., Schmidt, M., 2013. Influence of stand characteristics and landscape structure on wind damage. In: Gardiner, B., Schuck, A., Schelhass, C.O., Blennow, K., Nicolll, B. (Eds.), What science can tell us. Living with storm damage to forests. European Forest Institute, Joensuu, pp. 207–224.

Hanewinkel, M., Kuhn, T., Bugmann, H., Lanz, A., Brang, P., 2014. Vulnerability of uneven-aged forests to storm damage. Forestry 87, 525–534. https://doi.org/10.1093/forestry/cpu008.

Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143, 29–36.

Harris, I., Jones, P.D., Osborn, T.J., Lister, D.H., 2014. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 dataset. Int. J. Climatol. 34, 623–642. https://doi.org/10.1002/joc.3711.

Hart, E., Sim, K., Kamimura, K., Meredieu, C., Guyon, D., Gardiner, B., 2019. Use of machine learning techniques to model wind damage to forests. Agric. For. Meteorol. 265, 16–29. https://doi.org/10.1016/j.agrformet.2018.10.022.

Heaton, J., 2008. Introduction to neural networks for Java. 2nd edition. Heaton Research Inc 440 pp.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., et al., 2017. SoilGrids250m: global gridded soil information based on machine learning. PLoS ONE 12 (2), e0169748.

Hengl, T., Walsh, M.G., Sanderman, J., Wheeler, I., Harrison, S.P., Prentice, I.C., 2018. Global mapping of potential natural vegetation: an assessment of machine learning algorithms for estimating land potential. PeerJ 6, e5457. https://doi.org/10.7717/peerj.5457.

Hijmans, R.J., 2020. raster: Geographic Data Analysis and Modeling. R package version 3.3-13. https://CRAN.R-project.org/package=raster.

Horn, B.K.P., 1981. Hill shading and the reflectance map. Proc. IEEE 69, 14–47.

Hosmer, D.W., Lemeshow, S., Strudivant, R.X., 2013. Applied logistic regression. 3rd ed. John Wiley & Sons, New York.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning with Applications in R. Springer Texts in Statistics 441 pp.

Janda, P., Svoboda, M., Bače, R., Čada, V., Peck, J.E., 2014. Three hundred years of spatio-temporal development in a primary mountain Norway spruce stand in the Bohemian Forest, Central Europe. For. Ecol. Manag. 330, 304–311.

Jewuła, E., 1974. Regionalizacja szkód powodowanych przez wiatry w drzewostanach górskich I podgórskich południowej polski. Sylwan 118 (10), 54–63.

Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P., Kessler, M., 2017. Climatologies at high resolution for the earth's land surface areas. Sci. Data 4, 170122.

Klaus, M., Holsten, A., Hostert, P., Kropp, J.P., 2011. Integrated methodology to assess windthrow impacts on forest stands under climate change. For. Ecol. Manag. 261, 1799–1810.

Kopecka, M., 2011. Destruction of forest habitat in the Tatra National Park, Slovakia. In: Grillo, O., Venora, G. (Eds.), Biodiversity Loss in a Changing Planet. IntechOpen, pp. 257–276 https://doi.org/10.5772/24106.

Kopecký, M., Čížková, Š., 2010. Using topographic wetness index in vegetation ecology: does the algorithm matter? Appl. Veg. Sci. 13, 450–459.

Kramer, M.G., Hansen, A.J., Taper, M.L., Kissinger, E.J., 2001. Abiotic controls on long-term windthrow disturbance and temperate rain forest dynamics in southeast Alaska. Ecology 82 (10), 2749–2768.

Kuhn, M., 2020. caret: Classification and Regression Training. R package version 6.0-86. https://CRAN.R-project.org/package=caret.

Kuhn, M., Johnson, K., 2013. Applied predictive modeling. Springer 600 pp.

Kulakowski, D., Seidl, R., Holeksa, J., Kuuluvained, T., Nagel, T.A., Panayotov, M., Svoboda, M., Thorn, S., Vacchiano, G., Whitlock, C., Wohlgemuth, T., Bebi, P., 2017. A walk on the wild side: disturbance dynamics and the conservation and management of European mountain forest ecosystems. For. Ecol. Manag. 388, 120–131. https://doi.org/10.1016/j.foreco.2016.07.037.

Kwiatkowski, J., Hołdys, T., 1985. Klimat (in Polish). In: Jahn, A. (Ed.), Kakonosze polskie. PAN Wrocław, pp. 87–116.

Lek, S., Guégan, J.F., 1999. Artificial neural networks as a tool in ecological modelling, an introduction. Ecol. Model. 120, 65–73.

Lindner, M., Rummukainen, M., 2013. Climate change and storm damage risk in European forests. In: Barry, G., Schuck, A., Schelhaas, M.-J., Orazio, Ch., Blennow, K., Nicoll, B. (Eds.), Living With Storm Damage to Forests. European Forest Institute, pp. 109–115.

Martin, S.J., 2013. Wind as a natural disturbance agent in forests: a synthesis. Forestry 86, 147–157. https://doi.org/10.1093/forestry/cps058.

Mazurski, K.R., 1986. The destruction of forests in the polish Sudetes Mountains by industrial emissions. For. Ecol. Manag. 17, 303–315.

McMaster, K.J., 2005. Forest blowdown prediction: a correlation of remotely sensed contributing factors. North. J. Appl. Forest. 22, 48–53.

Migoń, P., Pánek, T., Malik, I., Hrádecký, J., Owczarek, P., Šilhán, K., 2010. Complex landslide terrain in the Kamienne Mountains, middle sudetes, SW Poland. Geomorphology 124, 200–214. https://doi.org/10.1016/j.geomorph.2010.09.024.

Mitchell, S., 2013. Wind as a natural disturbance agent in forests: a synthesis. Forestry 86, 147–157. https://doi.org/10.1093/forestry/cps058.

Molnar, C., Casalicchio, G., Bischl, B., 2018. Iml: an R package for interpretable machine learning. J. Open Source Softw. https://doi.org/10.21105/joss.00786.

Molnar, C., 2020. Interpretable Machine Learning. A guide for making black box models explainable. https://christophm.github.io/interpretable-ml-book/.

Musa, A.B., 2013. Comparative study on classification performance between support vector machine and logistic regression. Int. J. Mach. Learn. Cyber. 4, 13–24. https://doi.org/10.1007/s13042-012-0068-x.

Natekin, A., Knoll, A., 2013. Gradient boosting machines, a tutorial. Front. Neurorobot. https://doi.org/10.3389/fnbot.2013.00021.

Oksanen, J., Guillaume Blanchet, F., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Henry, M., Stevens, H., Wagner, H., 2016. vegan: Community Ecology Package. https://github.com/vegandevs/vegan http://cran.r-project.org.

Panayotov, M., Kulakowski, D., Laranjeiro, Dos Sanots, L., Bebi, P., 2011. Wind disturbances shape old Norway spruce-dominated forest in Bulgaria. Forest Ecol. Manag. 262, 470–481.

Pawlik, Ł., 2013. Remodelling of slope surface in the suche mts., SW Poland, as an effect of catastrophic windthrow caused by the Kyrill storm in 2007. In: Decaulne, A. (Ed.), Arbres & dynamiques. 2. Presses Universitaires Blaise Pascal, University of Clermont-Ferrand, France, pp. 49–69.

Pawlik, Ł., Migoń, P., Szymanowski, M., 2016. Local- and regional-scale biomorphodynamics due to tree uprooting in semi-natural and managed montane forests of the Sudetes Mountains, Central Europe. Earth Surf. Process. Landf. 41, 1250–1265.

Peltola, H., Kellomäki, S., Väisänen, H., 1999. Model computations of the impact of climatic change on the windthrow risk of trees. Clim. Chang. 41, 17–36.

Pebesma, E., 2018. Simple features for R: standardized support for spatial vector data. R J. 10 (1), 439–446. https://doi.org/10.32614/RJ-2018-009.

Pellikka, P., Järvenpää, E., 2003. Forest stand characteristics and wind and snow induced forest damage in boreal forest. In: Ruck, B. (Ed.), Proc. of the Int. Conf. on Wind Effects on Trees, Karlsruhe, Germany, 16-18 Sept. 2003, pp. 269–276.

Peltola, H., Kellomäki, S., 1993. A mechanistic model for calculating windthrow and stem breakage of Scots pines at stand edge. Silva Fennica 27, 99–111.

Pettit, J.L., Pettit, J.M., Janda, P., Rydval, M., Čada, V., Schurman, J.S., Nagel, T.A., Bače, R., Saulnier, M., Hofmeister, J., Matula, R., Kozák, D., Frankovič, M., Turcu, D.O., Mikoláš, M., Svoboda, M., 2021. Both cyclone-induced and convective storms drive disturbance patterns in European primary beech forests. J. Geophys. Res. Atmos. 126, e2020JD033929. https://doi.org/10.1029/2020JD033929.

Placek, A., 2011. Rzeźba strukturalna Sudetów w świetle wyników pomiarów wytrzymałości skał I analiz numerycznych modelu wysokości. Rozpr. Nauk. Inst. Geogr. Rozw. Reg. Uniw. Wroc. 16, 1–190.

Probst, P., Boulesteix, A.-L., 2017. To tune or not to tune the number of trees in random forest?. ArXiv preprint arXiv:1705.05654 https://arxiv.org/abs/1705.05654

Probst, P., Wright, M., Boulesteix, A.-L., 2019. Hyperparameters and tuning strategies for random forest. WIREs 9, e1301. https://doi.org/10.1002/widm.1301.

Regmi, N.R., Giardino, J.R., McDonald, E.V., Vitek, J.D., 2014. A comparison of logistic regression-based models of susceptibility to landslides in western Colorado, USA. Landslides 11, 247–262. https://doi.org/10.1007/s10346-012-0380-2.

R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rodriguez, E., Morris, C.S., Belz, J.E., Chapin, E.C., Martin, J.M., Daffer, W., Hensley, S., 2005. An assessment of the SRTM topographic products, technical report JPL D-31639. Jet Propulsion Laboratory, Pasadena, California, p. 143.

Różycka, M., Migoń, P., 2017. Tectonic geomorphology of the Sudetes Mountains (Central Europe) – a review and re-appraisal. Ann. Soc. Geol. Pol. 87, 275–300. https://doi.org/10.14241/asgp.2017.016.

Ruel, J.-C., 1995. Understanding windthrow: silvicultural implications. For. Chron. 71, 434–445.

Schaetzl, R.J., Johnson, D.L., Burns, S.F., Small, T.W., 1989. Tree uprooting: review of terminology, process, and environmental implications. Can. J. For. Res. 19, 1–11.

Schelhaas, M.-J., Nabuurs, G.-J., Schuck, A., 2003. Natural disturbances in the european forests in the 19th and 20th centuries. Glob. Chang. Biol. 9, 1620–1633.

Schindler, D., Bauhus, J., Mayer, H., 2012. Wind effects on trees. Eur. J. Forest Res. 131, 159–163. https://doi.org/10.1007/s10342-011-0582-5.

Schindler, D., Grebhan, K., Albrecht, A., Schönborn, J., 2009. Modelling the wind damage probability in forests in southwestern Germany for the 1999 winter storm 'Lothar'. Int. J. Biometeorol. 53, 543–554. https://doi.org/10.1007/s00484-009-0242-3.

Schmuck, A., 1969. The climate of the Sudetes (in Polish, with English summary). Problemy Zagospodarowania Ziem Górskich 5, 93–150.

Seidl, R., Schelhaas, M.-J., Lexer, M.J., 2011. Unraveling the drivers of intensifying forest disturbance regimes in Europe. Glob. Chang. Biol. 17, 2842–2852. https://doi.org/10.1111/j.1365-2486.2011.02452.x.

Seidl, R., Thom, D., Kautz, M., Martin-Benito, D., Peltoniemi, M., Vacchiano, G., Wild, J., Ascoli, D., Petr, M., Honkaniemi, J., Lexer, M.J., Trotsiuk, V., Mairota, P., Svoboda, M., Fabrika, M., Nagel, T.A., Reyer, C.P.O., 2017. Forest disturbance under climate change. Nat. Clim. Chang. 7, 395–402.

Senf, C., Seidl, R., 2020. Mapping the forest disturbance regimes of Europe. Nat. Sustain. https://doi.org/10.1038/s41893-020-00609-y.

Sobik, M., Błaś, M., Migała, K., Godek, M., Nasiółkowski, T., 2013. Klimat (in Polish). In: Knapik, R., Raj, A. (Eds.), Przyroda Karkonoskiego Parku Narodowego (in Polish). Karkonosze National Park, Jelenia Góra, pp. 147–186.

Suvanto, S., Henttonen, H.M., Nöjd, P., Mäkinen, H., 2016. Forest susceptibility to storm damage is affected by similar factors regardless of storm type: comparison of thunder storms and autumn extra-tropical cyclones in Finland. For. Ecol. Manag. 381, 17–28. https://doi.org/10.1016/j.foreco.2016.09.005.

Suvanto, S., Peltoniemi, M., Tuominen, S., Strandström, M., Lehtonen, A., 2019. High-resolution mapping of forest vulnerability to wind for disturbance-aware forestry. For. Ecol. Manag. 453, 117619. https://doi.org/10.1016/j.foreco.2019.117619.

Svoboda, M., Janda, P., Nagel, T.A., Fraver, S., Rejzek, J., Bače, R., 2012. Disturbance history of an old-growth sub-alpine Picea abies stand in the bohemian Forest, Czech Republic. J. Veg. Sci. 23, 86–97.

Svoboda, M., Janda, P., Bače, R., Fraver, S., Nagel, T.A., Rejzek, J., Mikoláš, M., Douda, J., Boublík, K., Šamonil, P., Čada, V., Trotsiuk, V., Teodosiu, M., Bouriaud, O., Biriş, A.I., Sýkora, O., Uzel, P., Zelenka, J., Sedlák, V., Lehejček, J., 2013. Landscape-level variability in historical disturbance in primary Picea abies mountain forests of the eastern Carpahians, Romania. J. Veg. Sci. https://doi.org/10.1111/jvs.12109.

Šamonil, P., Antolík, L., Svoboda, M., Adam, D., 2009. Dynamics of windthrow events in a natural fir-beech forest in the Carpathian Mountains. For. Ecol. Manag. 257, 1148–1156. https://doi.org/10.1016/j.foreco.2008.11.024.

Thurnhofer-Hemsi, K., López-Rubio, E., Molina-Cabello, M.A., Najarian, K., 2020. Radial Basis Function kernel optimization for Support Vector Machine classifiers. arXiv.org. https://arxiv.org/abs/2007.08233.

Touzani, S., Granderson, J., Fernandes, S., 2018. Gradient boosting machine for modeling the energy consumption of commercial buildings. Energy Build. 158, 1533–1543.

Trotsiuk, V., Svoboda, M., Janda, P., Mikolas, M., Bace, R., Rejzek, J., Samonil, P., Chaskovskyy, O., Korol, M., Myklush, S., 2014. A mixed severity disturbance regime in the primary Picea abies (L.) karst. Forests of the ukrainian Carpathians. For. Ecol. Manag. 334, 144–153.

Usbeck, T., Wohlgemuth, T., Dobbertin, M., Pfister, Ch., Bürgi, A., Rebetez, M., 2010. Increasing storm damage to forests in Switzerland from 1858 to 2007. Agric. For. Meteorol. 150, 47–55.

Xi, W., Peet, R.K., 2011. The complexity of catastrophic wind impacts on temperate forests. In: Lupo, A. (Ed.), Recent Hurricane Research. Climate, Dynamics, and Societal Impacts. IntechOpen, pp. 503–534 https://doi.org/10.5772/16167.

Yamamoto, S.-I., 2000. Forest gap dynamics and tree regeneration. J. For. Res. 5, 223–229.

Valinger, E., Fridman, Jonas, 2011. Factors affecting the probability of windthrow at stand level as a result of Gudrun winter storm in southern Sweden. For. Ecol. Manag. 262, 398–403. https://doi.org/10.1016/j.foreco.2011.04.004.

Vicena, I., Parez, J., Konopka, J., 1979. Ochrona lesa proti polomům. Ministerstvo lesního a vodního hospodárství CSR, Prague, Czech Republic.

Vitalo, R., Stephenson, D.B., 2009. Serial clustering of intense European storms. Willis Research Network. University of Exeter, UK . http://www.willisresearchnetwork.com.

Zielonka, T., Malcher, P., 2009. The dynamics of a mountain mixed forest under wind disturbances in the Tatra Mountains, central Europe – a dendroecological reconstruction. Can. J. For. Res. 39, 2215–2223.