Predicting Tree Growth and Transpiration in Forests: An Analysis of a Small-Scale Dataset With Pareto Optimized Tsaug Augmentation

Rytis Maskeliūnas[1], Robertas Damaševičius[1], Modupe Odusami[1], Diana Sidabrienė[2], Algirdas Augustaitis[2], Gintautas Mozgeris[2] *

[1] Center of Excellence Forest 4.0, Kaunas University of Technology, Kaunas (Lithuania)
[2] Center of Excellence Forest 4.0, Vytautas Magnus University, Kaunas (Lithuania)

* Corresponding author: rytis.maskeliunas@ktu.lt (R. Maskeliūnas), robertas.damasevicius@ktu.lt (R. Damaševičius), modupe.odusami@ktu.lt (M. Odusami), diana.sidabriene@vdu.lt (D. Sidabrienė), algirdas.augustaitis@vdu.lt (A.Augustaitis), gintautas.mozgeris@vdu.lt (G. Mozgeris).

Received 7 November 2023 | Accepted 28 February 2025 | Published 19 February 2026

Abstract

The study demonstrates the potential of specifically developed data augmentation in estimating tree growth and transpiration by emphasizing the influence of environmental variables, such as photosynthetically active radiation (PAR), air temperature, and relative humidity—on tree growth predictions. The investigation utilizes data obtained from two hemi-boreal semi-natural mixed conifer-deciduous forest sites in the Aukstaitija National Park in Lithuania. Field measurements included xylem sap flow measurements and stem circumference increment growth. The dataset utilized in the analysis consisted of four trees per species and contained information on tree growth, transpiration, and solar angle measurements. Pareto-optimized Tsaug augmentation techniques were employed to diversify the dataset, generating augmented time series to improve diversity and minimize distortion. The results of the correlation analysis indicated significant relationships between environmental variables and tree growth and transpiration. The Prophet based prediction model, notably when trained with augmented data, outperformed other models in predicting tree growth and perspiration variables (MAPE ranging from 0.0017 to 0.01). This was particularly evident for FACP, FAGP, and FADP variables, showcasing substantial improvement with augmented data.

Keywords

Pareto Optimized Data Augmentation, Tree Growth Prediction.

DOI: 10.9781/ijimai.2026.6565

I. Introduction

Forests are vital ecosystems that provide several ecological, economic and social advantages, including carbon sequestration, biodiversity protection, and wood production [1]. Forest ecosystems, on the other hand, are faced with a variety of risks and problems as a result of climate change, change in land use, and other human activities [2]. As a result, effective forest management and conservation methods are becoming increasingly important to preserve the long-term viability of forests and the services they Offer. Nowadays, the ability of forest ecosystems to resist disturbances is often discussed in the context of forest resilience. Resilience refers to their ability to absorb impacts while maintaining desired levels of biodiversity, providing ecosystem services, and sustaining the functioning of wood value chains ([3], [4], [5], [6], [7]). The ability and speed to recover from disturbances are crucial aspects of forest ecosystem management. It is evident that the resilience of forest ecosystems is influenced by the resilience of trees as individual organisms. Therefore, the ecophysiological processes at the individual tree level are of paramount importance in modern forest.

For many decades, researchers have been studying tree growth and transpiration in forests [8]. Eddy covariance measurements, which allow for continuous measurements of tree transpiration and ecosystem-level carbon and water exchange, are one of the most widely used ways to investigate these processes. Recent research has focused on improving the accuracy and reliability of eddy covariance measurements through the development of novel quality control and data processing approaches [9]. Another area of research has been the creation of models that anticipate tree growth and transpiration depending on environmental conditions [10]. The sap flow model, which calculates tree transpiration based on the rate of water transfer via the xylem, is one of the most widely used models. Xylem sap flow measurements are conducted using the heat-ratio method [11],

Please cite this article as: R. Maskeliūnas, R. Damaševičius, M. Odusami, D. Sidabrienė, A. Augustaitis, G. Mozgeris. Predicting Tree Growth and Transpiration in Forests: An Analysis of a Small-Scale Dataset With Pareto Optimized Tsaug Augmentation, International Journal of Interactive Multimedia and Artificial Intelligence, vol. 9, no. 6, pp. 126-141, 2026, http://doi.org/10.9781/ijimai.2026.6565

although specialized sap flow sensors are installed on tree stems. These sensors measure changes in temperature along the stem after applying heat at a specific point, based on the principle that the applied heat is transported by the sap as it moves through the sapwood and the velocity of this transport is positively correlated with the sap flow rate. Since transpiration is sensitive to the water status of the plant, this effect is mediated by stomatal opening, sap flow can serve as an indicator of the water status of the plant. However, it should be noted that these measurements often generate large volumes of data, which can pose challenges during data processing.

Machine learning algorithms have recently been used to forecast tree growth and transpiration based on a variety of environmental parameters such as temperature, humidity, solar radiation, and soil moisture [12]. In addition, studies have been conducted to better understand the links between tree growth and transpiration and environmental factors such as climate change, forest management techniques, and change in land use. Increases in temperature and atmospheric CO2 concentrations, for example, have been shown in studies to cause variations in tree growth and transpiration rates, while forest management measures such as thinning and managed burning can also alter these processes [13]. The influence of climate change on tree growth and transpiration is one of the most significant recent results in this domain. According to research, changes in temperature and precipitation patterns caused by climate change could affect tree growth and transpiration rates [14], thus affecting forest productivity and carbon sequestration. Drought stress, in particular, has a significant impact on tree growth and transpiration in many places, and continuing research aims to understand the processes by which trees adapt to drought stress and devise management techniques to alleviate its effects [15]. Other environmental elements that could affect tree growth and transpiration, in addition to climate change, include soil quality, terrain, and change in land use [16]. Soil nutrient availability, for example, has been proven in studies to influence tree growth rates, while topographic elements such as height and slope can influence local climatic conditions and water availability [17]. Changes in land use, such as logging or conversion to agricultural land, can also have a substantial influence on tree growth and transpiration rates [18].

The use of statistical, machine learning, and deep learning models to forecast tree growth and transpiration in forests demonstrated the potential of how these techniques can potentially give useful information about the intricate interactions between environmental conditions and ecosystem processes [19], [20]. However, for these models to be successful, considerable volumes of high-quality data, computing resources, and model creation and validation skills are required [21]. Ongoing research is focused on improving the accuracy and dependability of these models, as well as creating new strategies to combine data from various sources to improve our knowledge of forest ecosystem dynamics [22]. Deep learning algorithms for forecasting tree growth and transpiration have the potential to significantly improve our understanding of ecosystem dynamics [23]. However, more studies are still needed to improve these models, confirm their precision and dependability, and investigate their potential uses [24].

The accuracy and reliability of forest parameter prediction models are critically dependent on the quality and completeness of environmental data [25]. This dependency often poses significant limitations when the data are incomplete or contain noise. Although much research has been conducted on the environmental parameters that influence tree growth and transpiration, temporal gaps in data and variations in environmental conditions between different regions can further compromise the performance of the model [26]. While the data are very limited, anticipating tree growth and transpiration can have significant consequences for forest management and conservation, such as anticipating the effects of climate change on forest ecosystems and improving forest management methods to increase tree growth and productivity. Advances in data augmentation and pareto-optimized processing have resulted in the applicability of small-scale forest databases that can allow predicting and filling in the gaps of the hourly measurements of environmental parameters [27]. Such augmented databases offer a possibility to create more accurate statistical models that forecast tree growth and transpiration depending on environmental conditions.

The paper continues with a literature review that summarizes previous research on the topic, a methods section that describes the data set and the statistical models and machine learning approaches used for prediction, a results section that presents the findings of the study, and a discussion section that interprets the results and identifies their implications for forest management and conservation. The paper concludes with a summary of the main findings and their implications for future research and forest management practices.

II. Overview of Methods for Predicting Tree Growth and Transpiration

In the realm of tree growth and transpiration prediction, statistical models have proven their worth by integrating various environmental parameters and elucidating complex relationships between trees and their surroundings. Their versatility in handling multiple variables and capturing intricate interactions has been instrumental. These models rely on high-quality data and can be computationally demanding. Machine learning, including traditional artificial neural networks (ANN), random forests (RF) and gradient boost (GB) models, has emerged as a potent alternative, excelling in capturing non-linear relationships and delivering accuracy, especially when datasets are limited. Deep learning, marked by architectures such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and attention mechanisms, is paving the way for more nuanced predictions by effectively handling spatial and temporal complexities. The choice of modeling approach depends on the specific context and data availability, offering researchers a rich toolkit to further our understanding of tree-environment interactions and enhance predictive capabilities.

A. Overview of Statistical Modeling Based Applications

Statistical models that take advantage of environmental parameters, such as temperature ([28]), humidity, solar radiation, and soil moisture, have become valuable tools in forecasting tree growth and transpiration processes ([29]). These models, as discussed in previous studies ([28], [29]), offer a robust framework for understanding and predicting the complex relationships between trees and their environment. One key strength of these statistical models is their ability to integrate a wide range of environmental factors ([30]). They can simultaneously consider multiple variables, making them suitable for capturing intricate interactions and dependencies among environmental parameters ([31]). This capability is fundamental for accurately predicting tree growth and transpiration, as these processes are influenced by a multitude of factors that can vary both spatially and temporally. The effectiveness of these statistical models is highly dependent on the quality and completeness of the available data ([32]). Inaccurate or incomplete environmental data can introduce biases and limitations into the models, potentially leading to less reliable predictions. Many of these models require substantial computing resources for both training and application, particularly when dealing with large datasets or complex model structures. To evaluate the relationships between environmental factors and tree growth or transpiration rates, these models often rely on regression analysis or other statistical approaches. The linear regression (LR) model is a commonly used method in this context ([33]). It calculates the linear associations between one or more environmental factors and tree growth or transpiration, providing a straightforward and interpretable relationship. However, LR has its limitations, as it assumes that the relationships between variables are strictly linear. To account for non-linear connections and accommodate other sources of variation, more sophisticated models have been developed. These include the generalized linear model (GLM) and mixed-effects models ([34], [35]). GLM extends the capabilities of LR by allowing for nonlinear relationships between variables and handling data with non-normal distributions, which is particularly valuable when dealing with ecological data, where relationships may not conform to simple linear patterns. Mixed-effects models are another advance in statistical modeling for tree growth and transpiration analysis [36]. They incorporate fixed and random effects, which account for variability at multiple levels, such as individual tree variation and site-specific effects. The hierarchical approach enables a more nuanced understanding of the factors that influence tree growth and transpiration.

B. Overview of Machine Learning Modeling Based Applications

Machine learning algorithms continue to be a popular choice for forecasting tree growth and transpiration rates based on environmental data. This preference is often due to the relatively modest size of the available datasets in this domain. Traditional models based on artificial neural networks (ANN) are among the most frequently used machine learning methods in such situations [37]–[39]. They excel at capturing and representing complex nonlinear interactions between environmental factors and tree growth or transpiration rates. Several studies have demonstrated the effectiveness of ANN models in predicting tree growth and transpiration ([40], [41]), exceeding standard statistical models in various circumstances. Another common machine learning strategy for forecasting tree growth and transpiration is the random forest (RF) model [42]. RF models are constructed on the basis of decision trees, which are hierarchical structures that recursively partition data based on predictor variable values. RF models leverage multiple decision trees to create an ensemble model that is more accurate and robust than individual decision trees. Numerous studies have shown the success of RF models in predicting tree growth and transpiration, particularly when a substantial number of predictor variables are involved [37], [42], [43]. Gradient boosting (GB) models constitute another machine learning method used to forecast tree growth and transpiration rates [44]. GB models operate by iteratively fitting basic models to the residual errors of previous models, aiming to reduce the overall prediction error [45]. GB models have demonstrated the ability to predict tree growth and transpiration rates with high accuracy, even in cases where the underlying relationships between predictor variables and growth/transpiration rates are intricate and non-linear [46]. In addition to ANN, RF and GB models, several other machine learning methods have found utility in forecasting tree growth and transpiration, starting with Support Vector Machines (SVM) which can be effective in capturing complex relationships between environmental parameters and tree growth/transpiration by finding an optimal hyperplane that best data points into different classes, as they can handle both linear and non-linear relationships through the use of appropriate kernel functions [47]. Some researchers also use K-Nearest Neighbors (K-NN) models to predict tree growth and transpiration by identifying the k-nearest data points in the training dataset and aggregating their results [48]. KNN is particularly useful in this context when there is a spatial component to the data [49]. Finally, XGBoost could be used as an optimized gradient boost framework as it can handle both regression and classification tasks and is known for its speed and precision [50].

C. Overview of Deep Learning Modeling Based Applications

Although deep learning models are gaining traction in various fields, their application to forecast tree growth and transpiration rates is an emerging and promising area ([51]). Researchers are exploring a spectrum of deep learning architectures and methods to advance our understanding and prediction capabilities in this domain. Convolutional Neural Networks (CNNs) excel in capturing spatial patterns and features in images, making them valuable tools for image-based analysis in forestry and ecology [52], with application for estimating transpiration rates, for example based on high-resolution thermal images [53], [54]. Recurrent Neural Networks (RNNs) have the ability to model sequences of data effectively and are well suited for scenarios where there are temporal dependencies and dynamic interactions between environmental factors and tree responses [55]. Long-Short-Term Memory (LSTM) networks are suitable for modeling temporal dependencies in time series data, and they can capture long-term relationships between environmental variables and tree growth or transpiration, allowing for more accurate forecasts [56]. Autoencoders as unsupervised deep learning models have the potential to discover meaningful representations of environmental data, where learned representations can be valuable for feature engineering and improving the performance of other forecasting models [57]. Attention mechanisms, such as the Transformer architecture, are increasingly used to model complex dependencies between input variables and tree growth or transpiration rates, as transformer mechanisms can effectively handle spatial and temporal data [58]. Graph Neural Networks (GNNs) are also well suited for modeling relationships in complex environmental networks, as they can capture interactions between different trees in a forest ecosystem and their responses to environmental factors [59]. Researchers are also exploring hybrid models that combine the strengths of multiple deep learning architectures. For example, a combination of CNN for spatial data and LSTM for temporal data can provide a comprehensive approach from environmental [60] to plant growth and transpiration prediction [61].

III. Materials and Methods

A. Study Area

Investigations were carried out using materials from the strict reserve area in the northwest part of the Aukstaitija National Park (682 m above sea level) established in eastern Lithuania ([62], [63], [64], [65]). Two seminatural mixed conifer deciduous forest sites in the hemiboreal region grown on various types of soil, differing in humidity and subsoil water were selected. The sites have over-mature (>120 years), multi-aged and mixed forest dominated by Scots pine [63]. The dendrometric characteristics of the sites considered are presented in [65] and [66].

The first site for pure and mixed pine was established at deep (>120 cm) haplic arenosol soil type (> 120 cm), with a high ratio of sand (~90 percent) within ~80 cm of soil depth, mildly acid with a saturated water amount of ~20 percent in the root zone, the groundwater level is approximately 300 cm. The dry and wet bulk density is ~1.5 g cm^3 and ~1.6 g cm^3 soil in the main root zone.

The second forest site was established at the mesoeutrophic organic peatland forest site saturated with water, with groundwater level at approximately 50 cm depth of the soil. The type of soil is a deep (>80 cm) terric histosol consisting mostly of low moor deep eutrophic peat soil. The dry and wet bulk density differs (0.19 and 0.87 g cm3 soil), and the poorly drained site is mostly saturated with water. The root zone reaches a depth of 40 cm in the soil with a water content of ~400-500 percent [67].

To determine temporary dry periods, we used meteorological, soil water, and plant-related parameters.

B. Field Measurements

Xylem sap flow measurements were done using the heat ratio method [11] using sap flow meter (SFM1) from ICT International (Australia). Sap flow was measured in 5-7 sample pine trees to determine tree crown transpiration in the water-limited MS plot in the pure and mixed stand and in the water saturated GW plot - only in the pure stand. For sap flow measurements, predominant sample trees (Kraft class 1) of each species within the subplots were selected. Sample trees had similar dendrometric characteristics. Sap flow sensors were mounted at ~120 cm stem height and protected with aluminum foil caps.

Sap flow monitoring started in mid-April 2016 using a 15-min measurement interval. 15 min measuring intervals were summarized as hourly values as a basis for all subsequent calculations. The sapwood area for the sample trees was determined by equations, derived by correlation of the basal area and the sapwood depth of the trees analyzed. The results were related and counter-checked to results from literary sources. Sap velocity, sap flow, and total plant water use, were calculated using ICT International Sap Flow Tool Software. For each sample tree, the sap flux was computed as the flow rate per unit of basal area and is presented as mean values of the sap flux for each tree and the considered site (ml per cm2 basal area per time span). Crown transpiration is calculated as the product of the mean daily sap flux by the basal area per sample tree [66].

Stem circumference growth was recorded weekly using mechanical manual band dendrometers. Since April 2016, the increase in stem circumference was continuously recorded using high-resolution electronic logging band dendrometers (DRL26, EMS Brno, Czech Republic). In November, a prolongation was detected in the circumference of the stem of the monitored trees. These data were used to calculate the increment in the diameter /radius of the stem. The volume of the tree was computed as a function of the basal area of the tree, the index of the stem form, and the height. The increase in tree volume was found by multiplying the increase in the base area of the stem by the height of the tree and the index of the stem form, which defined the shape of the tree [66]. The stem form index was assessed in response to the height and diameter of the tree and was made for Scots pine trees at ~.45 [65]. In the MS plot limited to water, the increase in stem was detected in the pure and mixed stand and in the saturated water plot only in the pure stand.

C. Dataset

We chose to test the applicability of pareto-optimized data augmentation data representing three main species (Pinus sylvestris L. Picea abies L. H. Karst, Betula pendula Roth., see Fig. 1) on a small data set of 4 trees per species with around 20000 entries over a three-year period, including tree growth, transpiration, and solar angle, as well as a diurnal cycle in the data that should be eliminated using the solar angle data. This data set allowed us to analyze the correlations between the dependent variables (tree growth and transpiration). Forest time series data contains a substantial amount of information [68], but it is not always obvious due to problems such as unordered timestamps, missing values or timestamps, outliers, and noise in the data. To address this, we modified the data architecture, handling of missing values in the data, and scaling of data features. The imputation approach was used to structure the data. Missing values are replaced with the column mean in this manner. The method of feature scaling involves rescaling the features so that they have the attributes of a conventional normal distribution with a mean of zero and a standard deviation of one.

Fig. 1. Pine tree forest in Nida area, Lithuania.

To summarize and explain the properties of a data set (mean, median, mode, standard deviation, variance, minimum, and maximum values for each variable in the dataset), descriptive statistical analysis was performed. These statistics offer a concise summary of the central tendency, variability, and range of the data set. Box plots are also used as graphical representations of variable distributions to illustrate median, quartiles, and outliers. They summarize the range of values, central tendency, and variability of the data set (C, D, and G), providing information on the characteristics of the data set. Fig. 2 shows the results of descriptive statistics.

D. Correlation Analysis

The correlation analysis employed in this study intends to examine the link between the tabular datasetâ€™s independent characteristics and the dependent variables (tree growth and transpiration), and to assess the strength and direction of a two-variable linear connection. Pearson correlation coefficient was used, which is a measure of the linear association between two variables ranging from -1 to 1. A positive correlation coefficient implies that the two variables have a positive linear relationship, whereas a negative correlation value indicates that the two variables have a negative linear relationship. A coefficient close to 0 suggests that the variables have little or no linear connection. The findings of the correlation analysis are displayed as a correlation matrix and shown as a heatmap, with the correlation coefficient between each pair of variables shown. In the study, correlation analysis is used to determine which independent traits have the strongest and weakest association with the dependent variables. This can aid in identifying possible collinearity or multicollinearity concerns among the independent data and in selecting the most relevant features for forecasting tree growth and transpiration. Note that correlation analysis only examines the linear link between two variables and may miss nonlinear correlations or other forms of dependencies. As a result, various approaches, such as feature selection and machine learning algorithms, are required in conjunction with correlation analysis (see also the following sections). Fig. 3 shows the result of the correlation analysis between dependent variables and independent variables.

E. Pareto Optimized Tsaug Augmentation

The Tsaug library was designed to augment time series data. The augmentation pipeline includes four techniques: TimeWarp, Quantize, Drift, and Reverse. TimeWarp randomly warps the time axis, introducing variations in the temporal alignment of data points. Quantize discretizes values into a specified number of discrete levels, adding a form of quantization noise. Drift introduces random drift into time series, simulating gradual changes over time. Reverse reverses

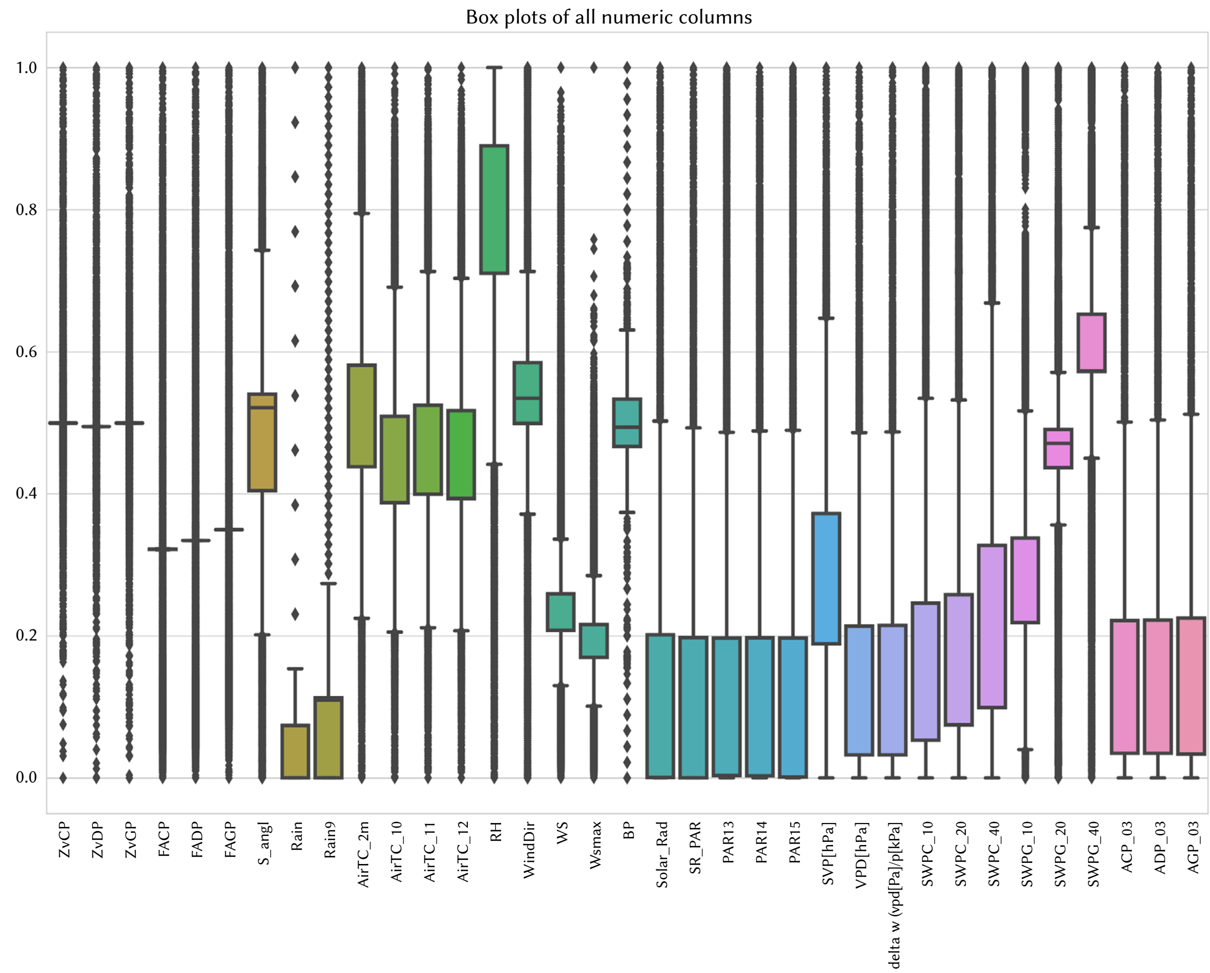


Fig. 2. Results of descriptive statistics.

the time series, capturing temporal patterns in opposite direction. Each technique aims to diversify the dataset by introducing different temporal perturbations. For each technique, a new augmented time series is created for each original time series.

Pareto Optimization is a multi-objective optimization technique used to find solutions that are optimal in terms of multiple conflicting objectives. In the context of Tsaug augmentation, it can be used to simultaneously optimize multiple augmentation objectives, such as maximizing diversity and minimizing distortion, to find a set of augmentations that are Pareto-optimal.

Let us define the mathematical representation of the Pareto Optimized Tsaug Augmentation. Suppose that we have a set of original time series data, denoted $X = \{x_1, x_2, ..., x_N\}$, where N is the number of time series in the data set. We want to generate a set of augmented time series for each original time series, denoted as $A = \{a_1, a_2, ..., a_N\}$, so that each augmentation technique aims to optimize multiple conflicting objectives.

Let $f_i(a)$ represent the objective function for the i-th augmentation technique, where i can take values from 1 to 4, corresponding to TimeWarp, Quantize, Drift, and Reverse. The goal of Pareto optimization is to find a set of augmented time series **A** that optimizes all these objectives simultaneously. This can be represented as Eq. 1.

$$\begin{array}{lll} \text{Maximize} & \{f_1(a), f_2(a), f_3(a), f_4(a)\} & \text{for each } a \in \mathbf{A} \\ \text{Subject to} & \text{Other constraints (if any)} & \end{array} \tag{1}$$

The pseudocode 1 represents an algorithm for performing Pareto Optimization for Tsaug Augmentation.

Algorithm 1. Pareto Optimization for Tsaug Augmentation

Input: *TS*

Objective: Diversity (*D*), Quality (*Q*), Distortion (*R*)

`// Objective 1: Maximizing Diversity (D)`

1 D ← 0

2 for i ← 1 to N **do**

3 for j ← 1 to M **do**

4 for k ← 1 to K **do**

$D \leftarrow D +$ `Dissimilarity` (TS_i, TS_{ijk})

`// Objective 2: Preserving Data Quality (Q)`

5 Q ← 0

6 for i ← 1 to N **do**

7 for j ← 1 to M **do**

8 for k ← 1 to K **do**

$Q \leftarrow Q +$ `SNR` (TS_i, TS_{ijk})

`// Objective 3: Minimizing Distortion (R)`

9 R ← 0

10 for i ← 1 to N **do**

11 for j ← 1 to M **do**

12 for k ← 1 to K **do**

$R \leftarrow R +$ `Distortion` (TS_i, TS_{ijk})

13 return D, Q, R

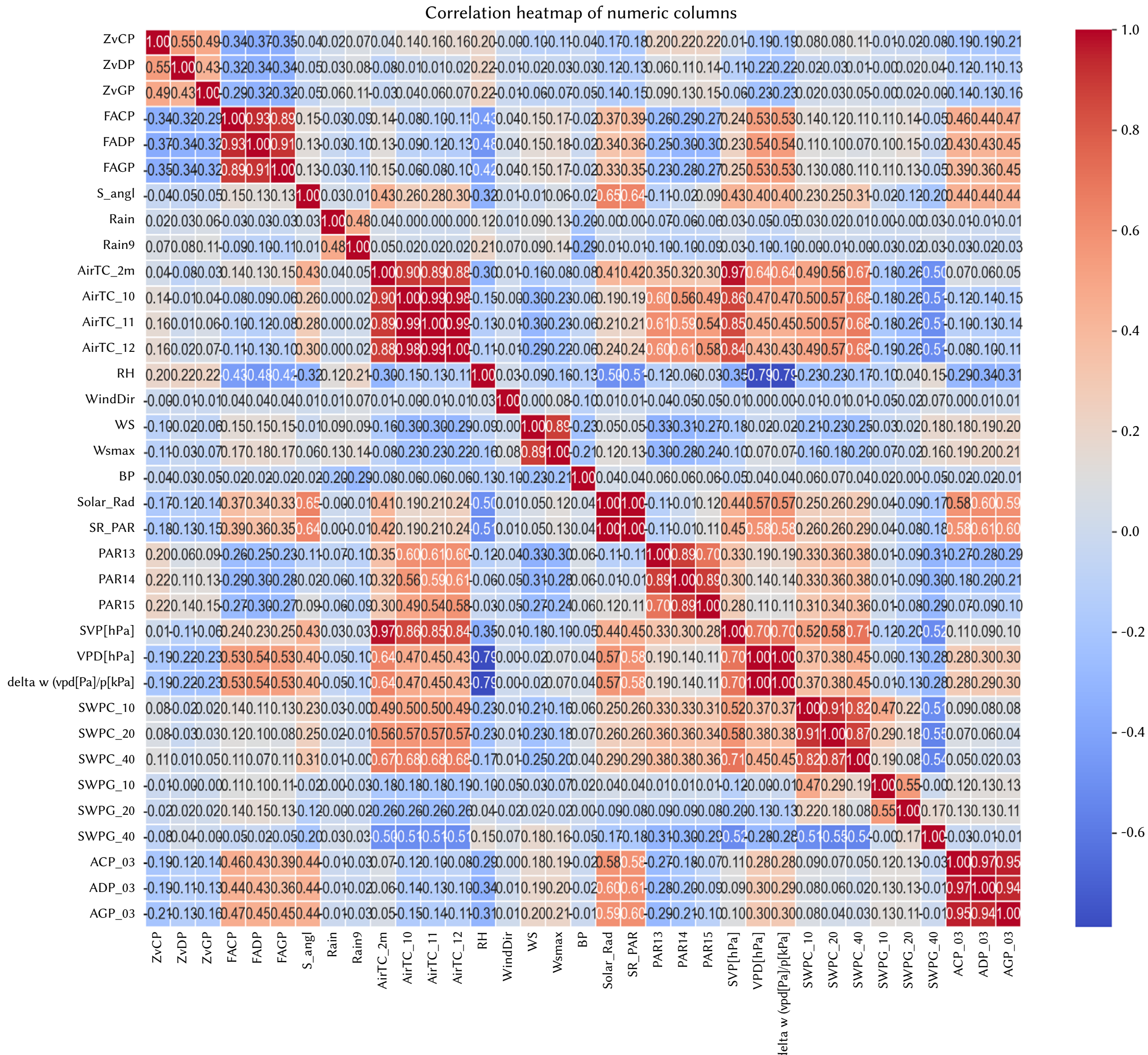


Fig. 3. Correlation heatmap.

The parameters of the algorithm are explained below:

- N represents the total number of original time series instances in your dataset, which is set to 30,301 here.
- M represents the total number of available augmentation techniques, including TimeWarp, Quantize, Drift, and reverse. It is set to 4.
- K denotes the number of augmented versions to be generated for each original time series using each augmentation technique. In your case, it is set to 4.
- P represents the total size of the data set after augmentation. It is set to 122,104 in this specific case.

There are three objectives defined:

- Diversity (D) aims to maximize the diversity of augmented time series in the data set. It quantifies the differences between augmented time series using a dissimilarity metric.
- Quality (Q) aims to preserve the quality of the data after augmentation. It measures data quality using a metric such as signal-to-noise ratio (SNR).
- Distortion (R) aims to minimize the distortion introduced by the augmentation techniques. It quantifies distortion as the difference between the original and augmented time series.

To maximize diversity (D), the algorithm initializes a variable D to 0 to accumulate diversity. It then uses nested loops to iterate over each original time series (i), each augmentation technique (j), and each augmented version (k) generated for that technique. Within the innermost loop, it calculates the dissimilarity between the original time series (TS_i) and the augmented time series (TS_{ijk}) and adds this value to the D variable.

To preserve data quality (Q), the algorithm aims to maximize data quality. It initializes a variable Q to 0 and uses nested loops to iterate over the original time series, enhancement techniques, and augmented versions. Within the innermost loop, it calculates the signal-to-noise ratio (SNR) between the original time series (TS_i) and the augmented time series (TS_{ijk}) and adds this value to the variable Q.

To minimize distortion (R) the algorithm initializes a variable R to 0 and uses similar nested loops to iterate over the original time series, augmentation techniques, and augmented versions. Within the innermost loop, it calculates the distortion between the original time series (TS_i) and the augmented time series (TS_{ijk}) and adds this value to the variable R.

Finally, to perform Pareto optimization, we use the multi-objective optimization algorithm MOEA/D (Multi-Objective Evolutionary Algorithm based on decomposition) [69].

There are 30301 instances of the original time series in the dataset, and each augmentation technique generates augmented versions for each original time series, the total dataset size after augmentation is 122104 instances. Fig. 4 shows the examples of augmentations.

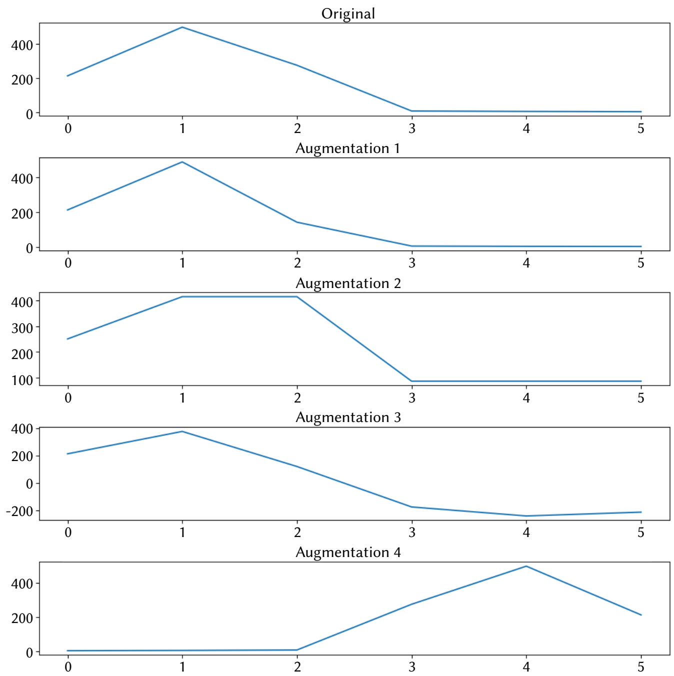


Fig. 4. Augmentation results.

F. Feature Engineering

We chose a principal component analysis (PCA) to choose the most significant characteristics in forecasting tree growth and transpiration from a tabular dataset in this research. PCA is a dimensionality reduction method that converts the original data into a new collection of uncorrelated characteristics known as principal components. It chooses the most essential main components according to how much they contribute to the total variance in the data. Fig. 5 shows the results of the feature importance ranking for the prediction of tree growth, while Fig. 6 shows the results of the feature importance ranking for tree transpiration.

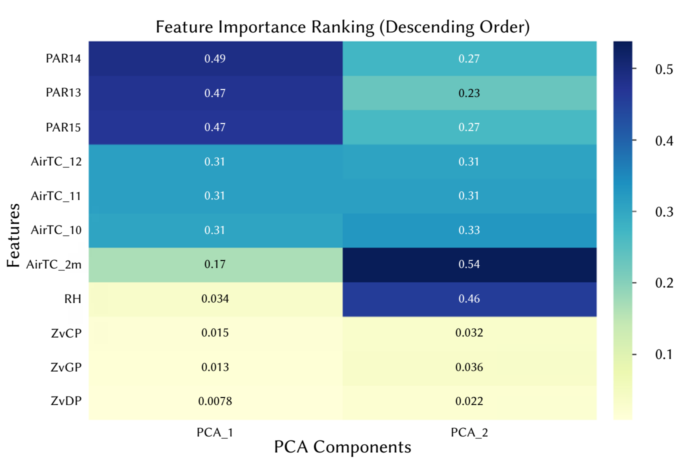


Fig. 5. Feature ranking results.

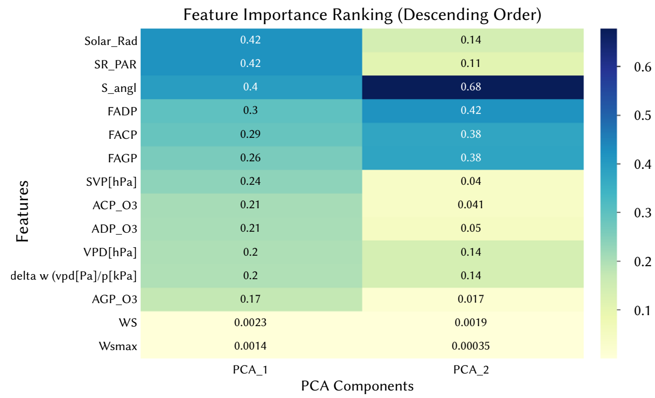


Fig. 6. Feature ranking results.

G. Regression Based Modeling

Many tree growth patterns and environmental interactions often exhibit strong linear relationships that can be effectively captured by linear regression and correlation analyzes. Trees typically respond to environmental variables such as temperature, precipitation, and soil nutrients in ways that can be approximated linearly, especially over smaller ranges or specific conditions. In addition, linear models offer simplicity and interpretability, which is advantageous when dealing with limited or noisy data. Although there may be non-linear interactions, dominant linear trends often provide valuable insights and reliable predictions for tree parameter analysis.

Linear regression was used to develop models that quantify the connection between the dependent variables and the independent characteristics to predict tree growth and transpiration from a tabular dataset. The LR model is based on the assumption that the relationship between the dependent and independent variables is linear and that the errors are normally distributed and independent. The model calculates the slope and intercept of the linear equation that best characterizes the variable connection. The slope is the change in the dependent variable for every unit increase in the independent variable, whereas the intercept is the value of the dependent variable when the independent variable is zero.

Suppose that there are n observations of the tree parameters, each with p predictor variables denoted by the p-dimensional predictor vector x, and a response variable y. Then, the LR model can be represented as Eq. 2.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \epsilon_i \tag{2}$$

where, y_i is the response variable for the i-th observation. $x_{i1}, x_{i2}, ..., x_{ip}$ are the predictor variables p for the i-th observation. $\beta_0, \beta_1, \beta_2, ..., \beta_p$ are the coefficients or parameters of the model that we want to estimate. ϵ_i is the error term or residual for the i-th observation.

The goal of linear regression is to estimate the values of the parameters $\beta_0, \beta_1, \beta_2, ..., \beta_p$ that minimize the sum of the squared errors or residuals, given Eq. 3:

$$RSS(\beta) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \ldots - \beta_p x_{ip})^2 \tag{3}$$

This can be done using the least squares method, which involves finding the values of the parameters that minimize the RSS. Estimates of the parameters can be obtained using Eq. 4:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{4}$$

where, $\hat{\beta}$ is a vector of estimated parameter values.

The application of linear regression models to predict tree growth and transpiration involves several steps. First, the data set is divided into training and testing sets and the model is fitted to the training data.

The model's performance is then evaluated on the testing set using metrics such as mean squared error, mean absolute error, or R-squared. The model coefficients can also be analyzed to understand the relative importance of each feature in predicting the dependent variables.

Similarly, the model can also represent Vector Autoregression (VAR). In a VAR model, we are dealing with a system of multiple time-series variables. The VAR(p) model for a variable vector y_t with p lags can be expressed as Eq. 5.

$$\mathbf{y}_t = \mathbf{c} + \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \cdots + \Phi_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t \tag{5}$$

Where: yt is a vector of time series variables at time t. c is a vector of constants. Φ_1, Φ_2, ..., Φ_p are the coefficient matrices for lags 1 to p. ε_t is a vector of white noise errors at time t.

In this context, the goal is to estimate the coefficient matrices $\Phi_1, \Phi_2, ..., \Phi_p$ that minimize the sum of squared errors (a multivariate analogue of RSS), which would involve solving a system of linear equations. The specific formula for estimating these coefficients depends on the estimation method used, such as ordinary least squares (OLS) or maximum likelihood estimation (MLE).

The dynamic factor model can also be applied in the context of forest growth prediction, where we are dealing with a multivariate time series of forest-related variables. The Dynamic Factor Model for this purpose can be expressed as follows:

At a specific time point t, we have a vector of forest growth-related variables denoted as Y_t, which includes metrics such as tree height, diameter, canopy coverage, etc. This vector is influenced by various underlying factors and the model can be represented as Eq. 6.

$$\mathbf{Y}_t = \boldsymbol{\mu} + \mathbf{F}_t + \boldsymbol{\Lambda}\mathbf{X}_t + \boldsymbol{\epsilon}_t \tag{6}$$

Where: Y_t is a vector of forest growth-related variables at time t. μ represents the mean or baseline state of these variables. F_t is a dynamic factor that captures the common trends or latent factors that affect forest growth over time. Λ is a loading matrix that relates the latent factors to the observed variables. X_t is a vector of additional predictor variables, such as weather conditions, soil quality, or other environmental factors that influence forest growth. ε_t represents the error term, accounting for any unexplained variability or noise in the observed variables.

The primary objective of the Dynamic Factor Model is to estimate the dynamic factor F_t and the loading matrix Λ that best explain the observed forest growth-related variables by first fitting the model to historical data. Once the model parameters are estimated, they can be used to forecast future forest growth based on the expected values of the predictor variables. Note that the Dynamic Factor Model is particularly useful when multiple correlated variables are involved, and it helps identify common underlying factors driving the observed variations in the forest-related data.

IV. Results and Evaluation

In this paper, we use the following metrics to evaluate the performance of our proposed regression model:

- The mean absolute error (MAE) measures the absolute difference between the actual and predicted values. It allows determining the average prediction error.
- The root mean squared error (RMSE) measures the square root of the mean squared errors between the actual and predicted values. It allows determining the general prediction accuracy.
- The mean absolute percentage error (MAPE) measures the percentage difference between the actual and predicted values. It allows determining the accuracy of the model when dealing with different scales of data.

Table I, Table II, Table III, and Table IV show the results of the prediction of tree growth and the prediction of tree perspiration using the linear regression model, VAR model, Dynamic factor and Prophet model, respectively.

TABLE I. Performance Result of Linear Regression Model in Predicting Tree Growth and Tree Perspiration

Variables	MAE	RMSE	MAPE
FACP	1.09	1.09	0.08
FAGP	0.16	0.16	0.01
FADP	0.52	0.52	0.03
ZvCP	0.11	0.11	1.30
ZvDP	0.15	0.15	2.50
ZvGP	0.19	0.19	4.25

TABLE II. Performance Result of VAR Model in Predicting Tree Growth and Tree Perspiration

Variables	MAE	RMSE	MAPE
FACP	0.418	0.418	0.032
FAGP	0.678	0.678	0.0427
FADP	0.316	0.316	0.0313
ZvCP	0.003	0.0030	0.035
ZvDP	0.002	0.002	0.0342
ZvGP	0.002	0.002	0.048

TABLE III. Performance of Dynamic Factor Model in Predicting Tree Growth and Tree Perspiration

Variables	MAE	RMSE	MAPE
FACP	0.25	0.32	1.69
FAGP	0.31	0.42	0.51
FADP	0.30	0.39	0.70
ZvCP	0.71	1.23	3.33
ZvDP	0.65	1.04	11.64
ZvGP	0.67	1.05	2.02

TABLE IV. Performance Result of Prophet Model in Predicting Tree Growth and Tree Perspiration

Variables	MAE	RMSE	MAPE
FACP	8.130	9.760	0.006
FAGP	6.840	7.990	0.004
FADP	6.840	7.990	0.004
ZvCP	0.08	0.09	0.009
ZvDP	0.03	0.04	0.006
ZvGP	0.04	0.04	0.008

Fig. 7 shows the tree growth performance using linear regression.

Fig. 8 shows the performance of tree perspiration using linear regression.

Fig. 9 shows the tree growth performance using the VAR model.

Fig. 10 shows the performance of tree perspiration using the VAR model.

Fig. 11 shows the tree growth performance using the dynamic factor model.

Fig. 12 shows the performance of tree perspiration using the dynamic factor model.

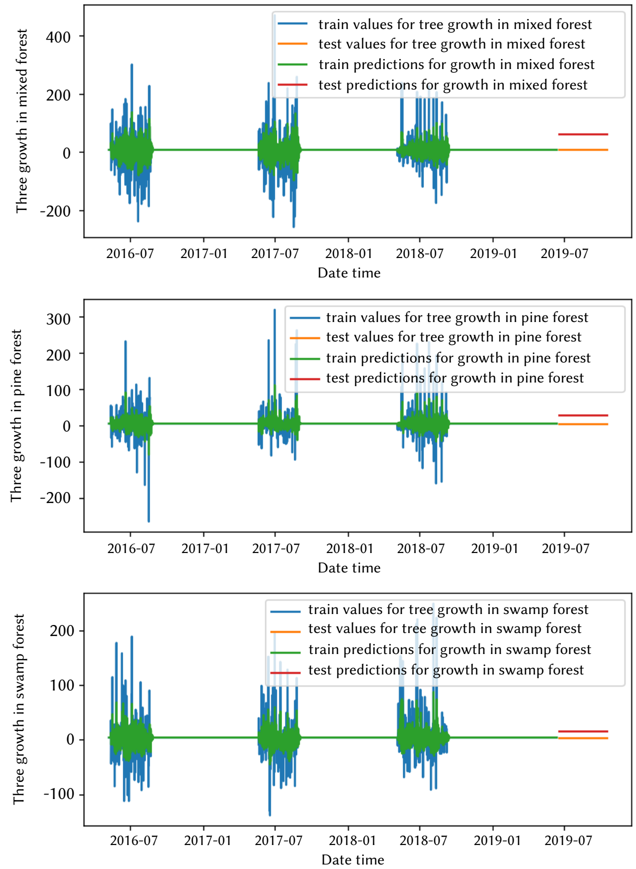


Fig. 7. Tree growth performance using linear regression model.

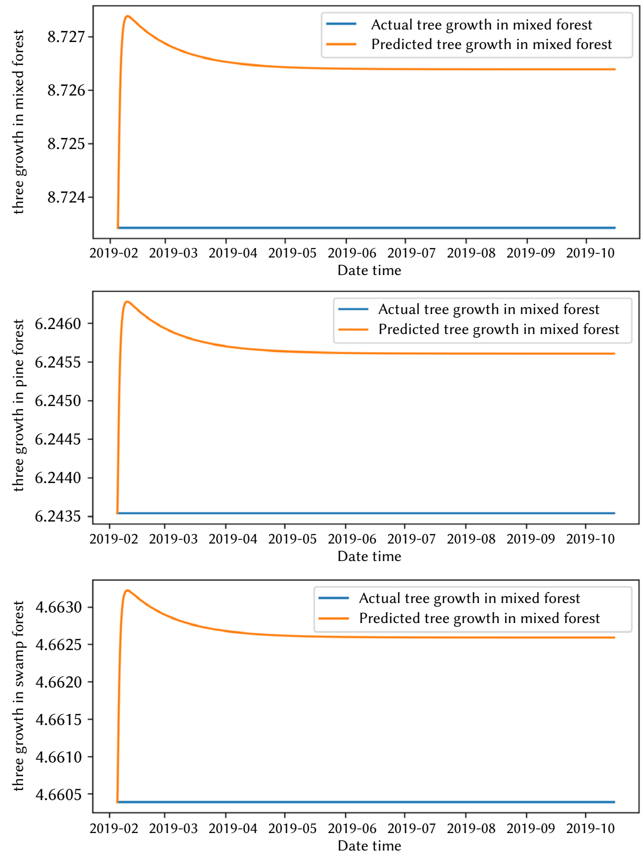


Fig. 9. Tree growth performance using VAR model.

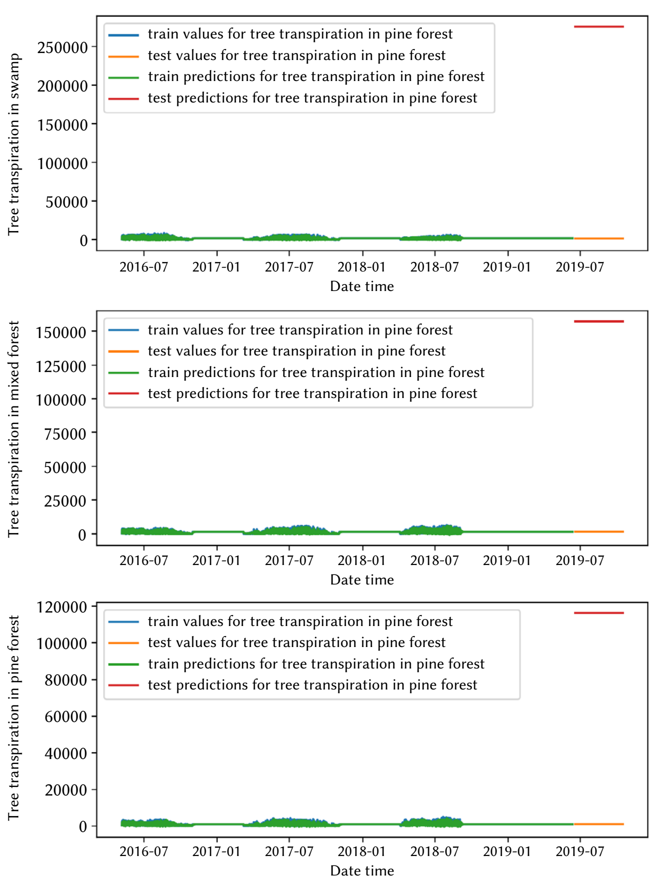


Fig. 8. Tree perspiration performance using linear regression model.

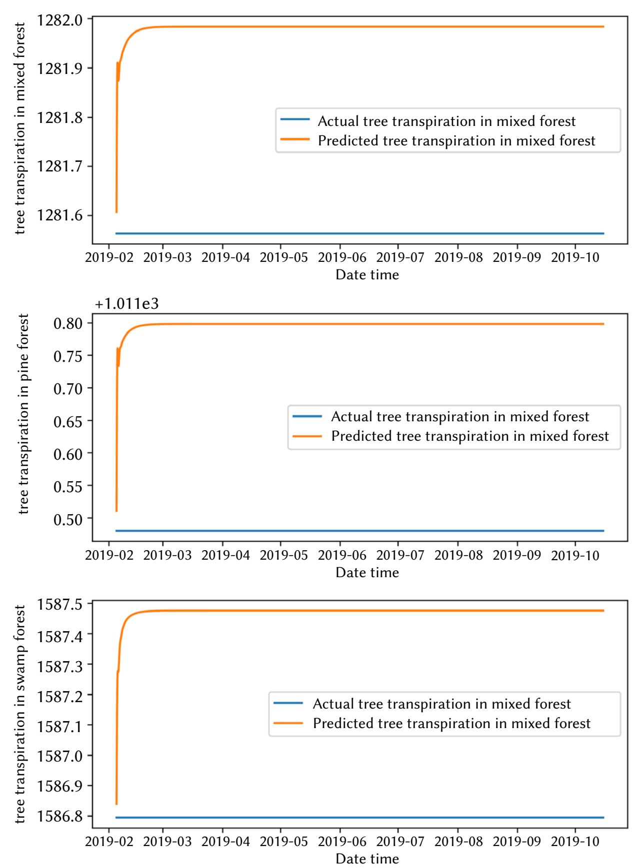


Fig. 10. Tree perspiration performance using VAR model.

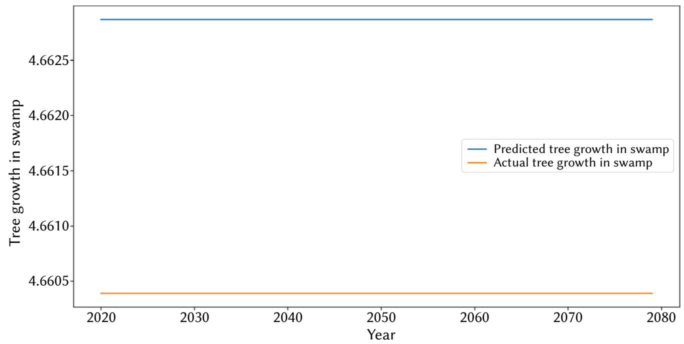


Fig. 11. Tree growth performance using Dynamic Factor Model.

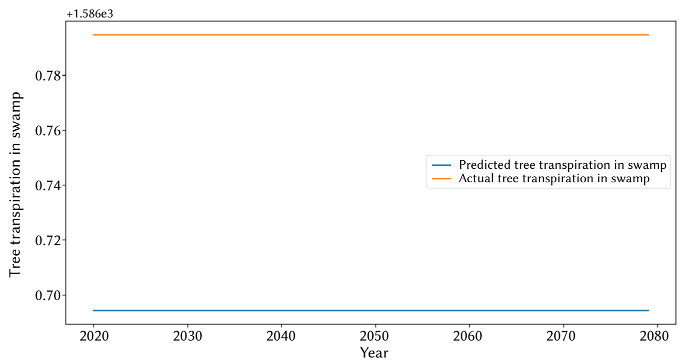


Fig. 12. Tree perspiration performance using Dynamic Factor Model.

Prophet modeling was used to create a data frame for the next 12 months and also to make a forecast based on the data frame. The seasonal trend for the FACP and ZvCP variables before augmentation is shown in Fig. 13 and 14, respectively. The forecast for 3 years for tree growth and perspiration is shown in Fig. 15 and Fig. 16, respectively.

Environmental variables such as PAR, air temperature, and relative humidity play a more significant role in predicting tree growth. These results suggest that changes in these environmental factors have a strong influence on the variations observed in the dependent variables. The dependent variables themselves (ZvCP, ZvDP, ZvCP) have lower importance values, indicating that they are less influential in explaining their own variations compared to the environmental factors. In Fig. 4 it is shown that Solar _Rad, SR _PAR, S _angl seem to be the most important factors in predicting these FACP, FADP, FAGP variables, with solar-related variables playing a significant role. From Table I, the linear regression model performs exceptionally well for predicting FAGP. The MAE and RMSE values of 0.16 indicate very accurate predictions, and the MAPE of 0.01% suggests that the model predictions are almost identical to the actual values. The performance of the model for ZvGP appears to be the least accurate among the variables listed. The MAE and RMSE values of 0.19 indicate that, on average, the predictions are off by about 0.19 units, which is relatively higher compared to other variables. The MAPE of 4.25% is the highest among all variables, suggesting a higher percentage error in the predictions for ZvGP.From Table I, the performance of the Linear Regression model varies for different variables. FAGP appears to have the most accurate predictions with very low errors, while ZvGP has the highest relative error with a MAPE of 4.25%. ZvCP and ZvDP also have relatively higher MAPE values. Also in Table II, the efficacy of the VAR model exhibits variability between different factors. The model demonstrates strong performance in the domains of ZvCP, ZvDP, and ZvGP, as seen by its MAE and RMSE values, as well as moderately MAPE values. However, FACP and FAGP exhibit marginally elevated errors, as evidenced by the higher values of MAE, RMSE, and MAPE. From Table III, While some variables such as ZvGP are predicted

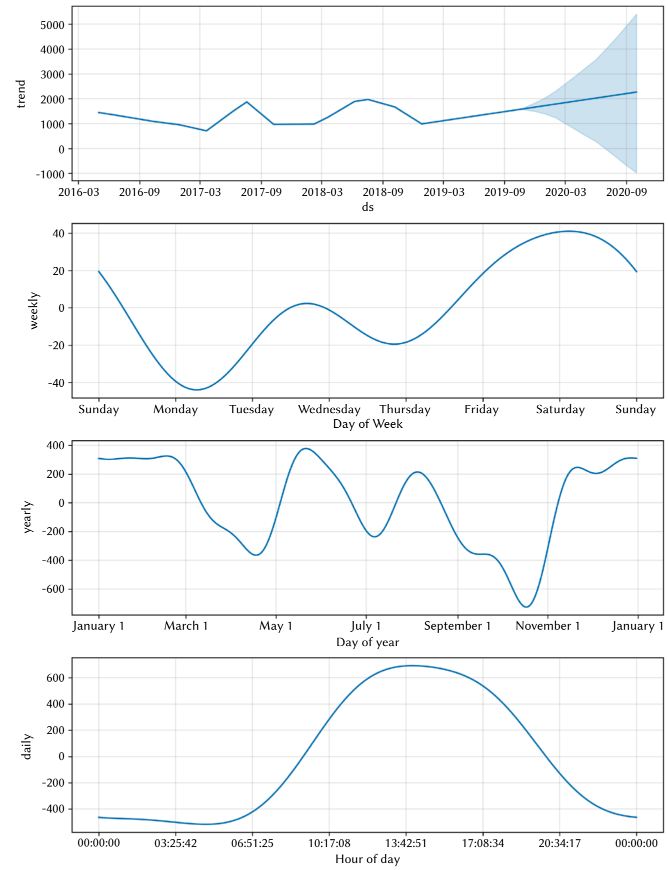


Fig. 13. Seasonal trend for tree transpiration in mixed forest.

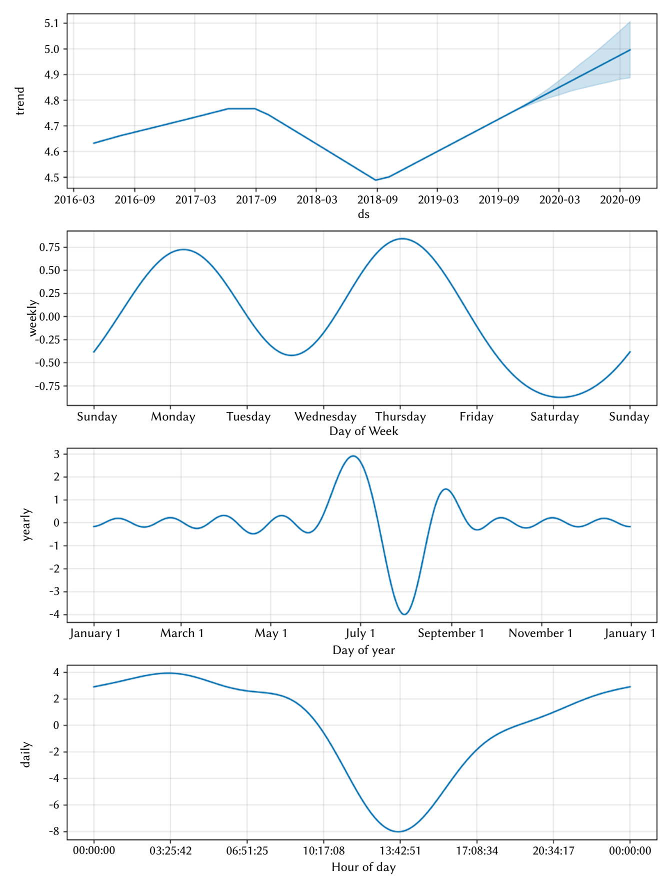


Fig. 14. Seasonal trend for tree growth in swamp forest.

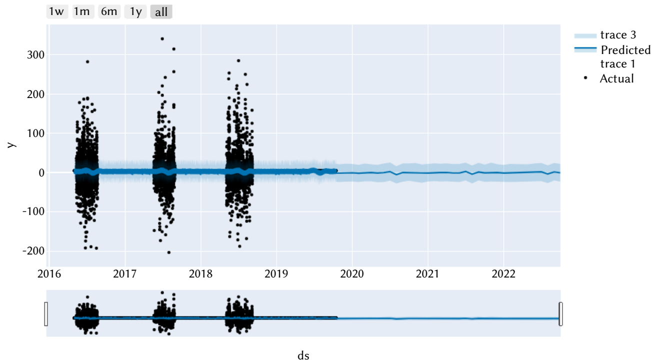


Fig. 15. Prediction of tree growth in swamp forest using prophet model.

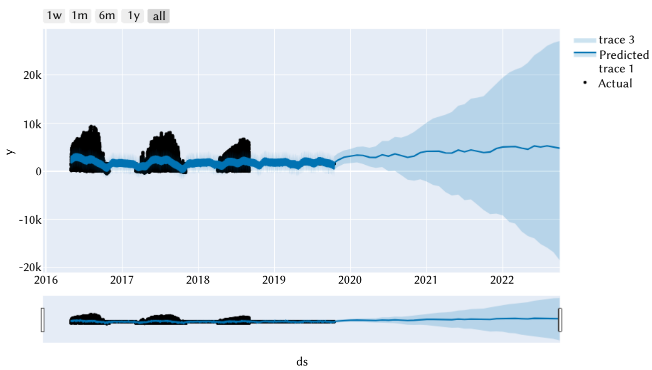


Fig. 16. Prediction of tree transpiration in swamp forest using prophet model.

with extremely high precision, others such as FACP and FAGP have relatively higher errors. From Table IV, the prophet model generally has lower errors and higher precision when predicting ZvCP, ZvDP, ZvGP compared to FACP, FAGP, FADP. From Fig. 5 to Fig. 13 and Fig. 14, ZvGP and FAGP are clearly visualized for the next 3 years, respectively, when compared to Fig. 5 to Fig. 10. The visualization of the prediction using the prophet model shows a more accurate value in the prediction value than the other models, also its MAPE value is lower than values obtained by the other models; therefore, the augmented data is used on the prophet model in order to obtain a more accurate prediction values. Table V shows the performance of the prophet model with augmented data.

TABLE V. Performance of Prophet Model in Predicting Tree Growth and Tree Perspiration Using Augmented Data

Variables	MAE	RMSE	MAPE
FACP	0.007	0.01	0.001
FAGP	0.007	0.01	0.001
FADP	0.007	0.01	0.001
ZvCP	2.13	2.79	0.0017
ZvDP	2.13	2.79	0.0017
ZvGP	2.13	2.79	0.0017

The analysis indicates that environmental variables such as PAR, air temperature, and relative humidity play a significant role in predicting tree growth. These factors seem to have a strong influence on the variations observed in the dependent variables.The dependent variables themselves (ZvCP, ZvDP, ZvCP) are found to have lower importance values, indicating that they are less influential in explaining their own variations compared to the environmental factors. The Linear Regression model performs exceptionally well for predicting FAGP. The low MAE and RMSE values of 0.16% indicate very accurate predictions, and the MAPE of 0.01%. However, the performance of the model for ZvGP appears to be the least accurate among the variables. The MAE and RMSE values of 0.19% indicate that, on average, the predictions are off by about 0.19%, which is relatively higher compared to other variables. The MAPE of 4.25% is the highest among all variables, suggesting a higher percentage error in the predictions for ZvGP. The performance of the linear regression model varies for different variables, with FAGP having the most accurate predictions and ZvGP having the highest relative error. ZvCP and ZvDP also have relatively higher MAPE values. The VAR Model demonstrates strong performance in the prediction of ZvCP, ZvDP, and ZvGP, as evidenced by its low MAE and RMSE values, as well as moderately low MAPE values. However, FACP and FAGP exhibit marginally elevated errors in their predictions, as evidenced by the higher values of MAE, RMSE, and MAPE. This suggests that the VAR model may not perform as well for these two variables as it does for the others.The Dynamic Factor Model shows much higher MAE and RMSE values for all variables compared to the previous models, indicating that its predictions have larger errors. The MAPE values for FACP, FAGP, and FADP are relatively low, indicating a reasonable percentage error in the predictions. However, for ZvCP and ZvDP, the MAPE values are higher, suggesting a higher percentage error in the prediction of these variables. the Linear Regression model seems to perform well for some variables like FAGP but not as well for others like ZvGP. The VAR Model shows good performance for certain variables and relatively higher errors for others.

On the other hand, the Dynamic Factor Model appears to have larger errors in its predictions for all variables. Likewise, the difference in performance indicates that prophet models are more suitable for predicting the variables related to tree growth. From the visualization of the proposed models forecast for tree growth and tree perspiration, the prophet model gave a nearer value of prediction than the other models, as shown in Fig. 14.

Table V shows a significant improvement in the performance of the prohet model (lower MAE, RMSE, and MAPE) when augmented data are used. The impact of augmentation varies between variables. The variables FACP, FAGP, and FADP show substantial improvements in MAE and RMSE, indicating that the augmented data have helped the model to make more accurate predictions for these variables. On the other hand, the variables ZvCP, ZvDP, and ZvGP, while still showing improvements, have higher errors even after augmentation. The variables ZvCP, ZvDP, and ZvGP seem to benefit less from data augmentation compared to FACP, FAGP, and FADP. Although the use of data augmentation has improved the performance of the Prophet model in predicting tree growth and perspiration, as evidenced by lower MAE, RMSE and MAPE values. The impact of augmentation varies across different variables,and further investigation may be needed to understand why certain variables benefit more than others and whether additional augmentation techniques can be applied to further enhance model accuracy. Fig. 17 provides the comparison of prophet model with augmented data and non-augmented data.

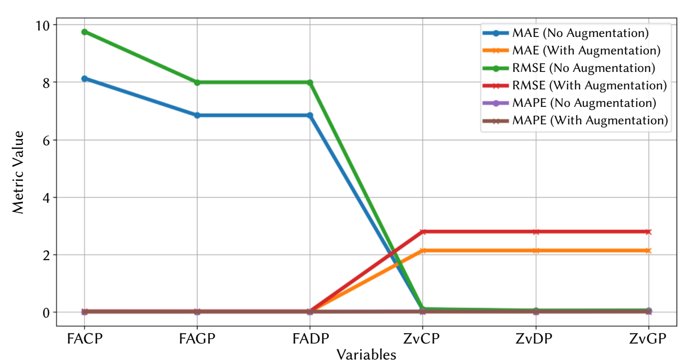


Fig. 17. Prophet Model with Augmented data and non- Augmented data.

Fig. 18 highlights variations in model performance in different variables and metrics. The Prophet model exhibits lower errors, with MAE and RMSE values as low as 0.03 and 0.04, respectively, for ZvDP, and a MAPE of just 0.006 for FACP, indicating superior prediction accuracy compared to the other models. In contrast, the Dynamic Factor Model consistently shows higher errors, for instance, an MAE of 0.71 and RMSE of 1.23 for ZvCP, suggesting it may be less effective for this dataset. Linear regression and VAR models exhibit moderate performance; for example, Linear regression has an MAE and RMSE of 0.16 for FAGP, and VAR shows an RMSE of 0.003 for ZvCP, indicating decent accuracy.

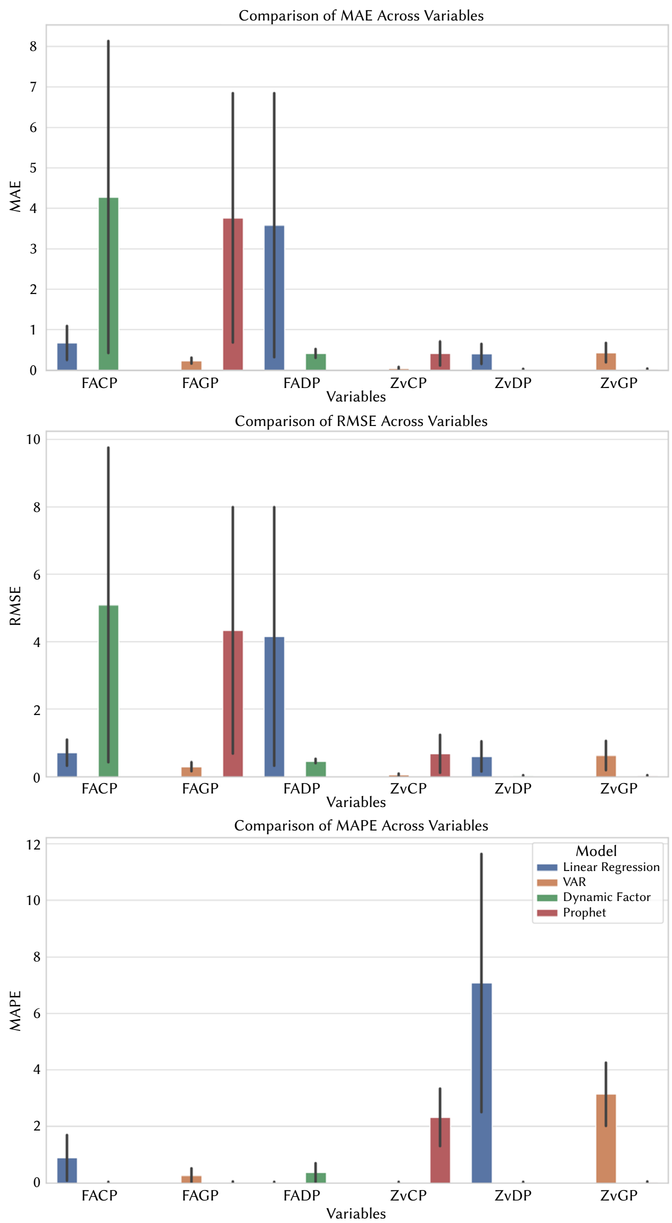


Fig. 18. Comparison of models across different forest types.

V. Discussion and Conclusions

Compared to other augmentation approaches (see Table VI), our Tsaug-based Paareto augmentation approach helped predict tree growth and transpiration. Environmental variables such as PAR, air temperature and relative humidity were found to play a major effect in the forecast of tree growth, highlighting the need to monitor and regulate these variables for forest health. The significance values of the dependent variables (ZvCP, ZvDP, ZvGP) were lower, suggesting that focusing on controlling environmental conditions may be more beneficial in managing tree growth and transpiration. The linear regression model performed exceptionally well in predicting FAGP but had larger errors in predicting ZvGP. The VAR model demonstrated variable-specific performance, with good predictions for certain variables and larger errors for others. Although the dynamic factor model had greater errors overall, it did give insights into model performance. The need for customized tactics depending on the accuracy of projections for certain variables has implications for forest management and conservation.

The results show that environmental factors such as PAR, air temperature, and relative humidity have a greater impact on the prediction of tree development. This implies that changes in these environmental conditions have a significant impact on tree health and development. To sustain healthy forests, forest managers must monitor and control certain environmental factors. In terms of variable importance, this study found that the dependent variables (ZvCP, ZvDP and ZvGP) themselves had lower important values, which means that they are less relevant in explaining their fluctuations compared to environmental influences. This implies that concentrating on regulating and optimizing climatic conditions may be more helpful in managing tree growth and transpiration.Depending on the model's performance, different variables may necessitate variable-specific management tactics.

Tsaug-based Pareto augmentation also addresses the issue of overfitting, particularly with small original datasets, by leveraging a multi-objective optimization approach. This technique balances the trade-off between the diversity and fidelity of augmented data, ensuring that the synthetic data generated is both varied and representative of the underlying data distribution. By carefully selecting augmented samples that contribute to different Pareto optimal solutions, the method improves the generalizability of the model. This reduces the likelihood of overfitting, making the model more robust and better equipped to handle new unseen data, thus improving overall performance and reliability.

Naturally, this study has limitations, such as the narrow time range of the dataset and the absence of an in-depth investigation of the causative processes underlying environmental influences on tree growth and sweat. Long-term data analysis, new modeling approaches, and better knowledge of the biological mechanisms involved in tree growth and sweating might aid future studies. This would lead to more accurate forecasts and a better understanding of the dynamics of forest ecosystems. The complexity and novelty of the augmentation techniques also pose challenges in interpreting the results and understanding the underlying processes driving the predictions, thus the development of forest-specific explainable AI models is also becoming a necessity. Finally, such approaches might not scale well to much larger datasets or different types of forests, limiting the applicability of the findings to other contexts; therefore, additional and more wide-scale benchmarks will be conducted in our future research.

Funding Information

This research paper has received funding from Horizon Europe Framework Programme (HORIZON), call Teaming for Excellence (HORIZON-WIDERA-2022-ACCESS-01-two-stage) - Creation of the centre of excellence in smart forestry "Forest 4.0" No. 101059985. This research has been co-funded by the European Union under the project FOREST 4.0 - "Ekscelencijos centras tvariai miško bioekonomikai vystyti" No.10-042-P-0002.

TABLE VI. Comparison of Data Augmentation Techniques

Technique	Description	Advantages	Limitations
SMOTE	Synthetic Minority Oversampling Technique	Balances class distribution, reduces bias	May introduce noise, less effective with highly imbalanced data
GANs	Generative Adversarial Networks	Generates high-quality, realistic data	Computationally intensive, requires large training sets
Time-Series Augmentation	Techniques like jittering, scaling, and cropping for temporal data	Captures temporal patterns, improves robustness	May distort original signal if not applied carefully
Rotation and Scaling	Geometric transformations of spatial data	Simple to implement, preserves spatial relationships	Limited to spatial data, may not capture all variability
Noise Injection	Adding noise to existing data points	Enhances model robustness, simulates real-world variations	Risk of degrading data quality if excessive
Synthetic Data Generation	Creating entirely new data points based on statistical properties	Expands dataset size, preserves data characteristics	May not fully capture complex relationships
Tsang-Based Pareto Augmentation	Multi-objective optimization to enhance data diversity	Balances fidelity and diversity, reduces overfitting	More complex implementation, parameter tuning needed

Credit Authorship Contribution Statement

Rytis Maskeliūnas: Conceptualization, Methodology, Formal analysis, Supervision, Writing – original draft, Project administration.

Robertas Damaševičius: Conceptualization, Methodology, Validation, Visualization, Writing – review & editing.

Modupe Odusami: Software, Formal analysis, Investigation, Visualization, Writing – original draft.

Diana Sidabrienė: Investigation, Data curation, Resources.

Algirdas Augustaitis: Conceptualization, Data curation, Resources.

Gintautas Mozgeris: Validation, Formal analysis, Writing – review & editing.

Data Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declaration Of Conflicts Of Interest

The authors declare no conflicts of interest.

Acknowledgments

This research paper has received funding from Horizon Europe Framework Programme (HORIZON), call Teaming for Excellence (HORIZON-WIDERA-2022-ACCESS-01-two-stage) - Creation of the centre of excellence in smart forestry “Forest 4.0” No. 101059985. This research has been co-funded by the European Union under the project "FOREST 4.0 - Ekscelencijos centras tvariai miško bioekonomikai vystyti" (Nr. 10-042-P-0002).

The Use of AI tools

The article has been improved for clarity and language refinement using the latest versions of the WriteFull and Grammarly AI tools. Although these tools have improved the overall quality of the text, the ideas, content, and responsibility for the final output remain solely with the authors. Any remaining errors or omissions are not attributable to the use of these tools.

References

[1] S. Díaz, A. Hector, D. A. Wardle, “Biodiversity in forest carbon sequestration initiatives: not just a side benefit,” *Current Opinion in Environmental Sustainability*, vol. 1, pp. 55–60, Oct. 2009, doi: https://doi.org/10.1016/j.cosust.2009.08.001.

[2] E. Doblas-Miranda, J. Martínez-Vilalta, F. Lloret, A. Álvarez, A. Ávila, F. J. Bonet, L. Brotons, J. Castro, J. C. Yuste, M. Díaz, P. Ferrandis, E. García-Hurtado, J. M. Iriondo, T. F. Keenan, J. Latron, J. Llusià, L. Loepfe, M. Mayol, G. Moré, D. Moya, J. Peñuelas, X. Pons, R. Poyatos, J. Sardans, O. Sus, V. R. Vallejo, J. Vayreda, J. Retana, “Reassessing global change research priorities in mediterranean terrestrial ecosystems: how far have we come and where do we go from here?,” *Global Ecology and Biogeography*, vol. 24, pp. 25–43, Sept. 2014, doi: https://doi.org/10.1111/geb.12224.

[3] W. N. Adger, T. P. Hughes, C. Folke, S. R. Carpenter, J. Rockström, “Social-ecological resilience to coastal disasters,” *Science*, vol. 309, no. 5737, pp. 1036–1039, 2005, doi: https://doi.org/10.1126/science.1112122.

[4] C. Folke, “Resilience: The emergence of a perspective for social–ecological systems analyses,” *Global Environmental Change*, vol. 16, no. 3, pp. 253–267, 2006, doi: https://doi.org/10.1016/j.gloenvcha.2006.04.002. Resilience, Vulnerability, and Adaptation: A Cross-Cutting Theme of the International Human Dimensions Programme on Global Environmental Change.

[5] R. Seidl, “The shape of ecosystem management to come: Anticipating risks and fostering resilience,” *BioScience*, vol. 64, pp. 1159–1169, 11 2014, doi: https://doi.org/10.1093/biosci/biu172.

[6] R. Seidl, W. Rammer, T. Spies, “Disturbance legacies increase the resilience of forest ecosystem structure, composition and functioning,” *Ecological applications*, vol. 24, pp. 2063–2077, 08 2014, doi: https://doi.org/10.1890/14-0255.1.

[7] R. Seidl, T. Spies, D. Peterson, Stephens, J. Hicke, “Searching for resilience: addressing the impacts of changing disturbance regimes on forest ecosystem services,” *Journal of Applied Ecology*, vol. 53, pp. 120–129, 2016, doi: https://doi.org/10.1111

[8] Z. Miao, F. R. A. Widagdo, L. Dong, F. Li, “Prediction of branch growth using quantile regression and mixed-effects models: An example with planted larix olgensis henry trees in northeast china,” *Forest Ecology and Management*, vol. 496, 2021.

[9] N. Nattharom, S. Roongtawanreongsri, S. Bumrungsri, “Growth prediction for rubber tree and intercropped forest trees to facilitate environmental services valuation in south thailand,” *Biodiversitas*, vol. 21, no. 5, pp. 2019–2034, 2020.

[10] M. Vaz Monteiro, T. Levanič, K. J. Doick, “Growth rates of common urban trees in five cities in great britain: A dendrochronological evaluation with an emphasis on the impact of climate,” *Urban Forestry and Urban Greening*, vol. 22, pp. 11–23, 2017.

[11] S. Burgess, M. Adams, N. Turner, C. Beverly, C. Ong, A. Khan, T. Bleby, “An improved heat pulse method to measure low and reverse rates of sap

flow in woody plants," *Tree physiology*, vol. 21, pp. 589–598, 09 2001, doi: https://doi.org/10.1093/treephys/21.9.589.

[12] F. Ellsäßer, A. Röll, J. Ahongshangbam, P.-A. Waite, Hendrayanto, B. Schuldt, D. Hölscher, "Predicting tree sap flux and stomatal conductance from drone-recorded surface temperatures in a mixed agroforestry system—a machine learning approach," *Remote Sensing*, vol. 12, p. 4070, Dec. 2020, doi: https://doi.org/10.3390/rs12244070.

[13] A. J. Hansen, R. P. Neilson, V. H. Dale, C. H. Flather, L. R. Iverson, D. J. Currie, S. Shafer, R. Cook, P. J. Bartlein, "Global change in forests: Responses of species, communities, and biomes," *BioScience*, vol. 51, no. 9, p. 765, 2001, doi: https://doi.org/10.1641/0006-3568(2001)051[0765:gcifro]2.0.co;2.

[14] J. Archambeau, M. B. Garzón, F. Barraquand, M. Miguel, C. Plomion, S. C. González-Martínez, "Combining climatic and genomic data improves range-wide tree height growth prediction in a forest tree," *American Naturalist*, vol. 200, no. 4, pp. E141–E159, 2022.

[15] N. A. Ryazanova, N. V. Polyakova, Z. K. Shigapov, "Dependence of phenorhythms of north american maples on the meteorological parameters in the southern urals (ufa)," *Contemporary Problems of Ecology*, vol. 15, no. 6, pp. 690–698, 2022.

[16] J. Shi, F. Xu, W. Xiang, "Fusing tree-ring and permanent sample plot data to model influences of climate and thinning on tree growth in larch plantations in northeast china," *Forest Ecology and Management*, vol. 531, 2023.

[17] D. Achat, C. Deleuze, G. Landmann, N. Pousse, J. Ranger, L. Augusto, "Quantifying consequences of removing harvesting residues on forest soils and tree growth – a meta-analysis," *Forest Ecology and Management*, vol. 348, pp. 124–141, July 2015, doi: https://doi.org/10.1016/j.foreco.2015.03.042.

[18] R. M. Marchin, A. A. Broadhead, L. E. Bostic, R. R. Dunn, W. A. Hoffmann, "Stomatal acclimation to vapour pressure deficit doubles transpiration of small tree seedlings with warming," *Plant, Cell & Environment*, vol. 39, pp. 2221–2234, Aug. 2016, doi: https://doi.org/10.1111/pce.12790.

[19] L. Xing, N. Cui, C. Liu, L. Zhao, L. Guo, T. Du, C. Zhan, Z. Wu, S. Wen, S. Jiang, "Estimation of daily apple tree transpiration in the loess plateau region of china using deep learning models," *Agricultural Water Management*, vol. 273, p. 107889, Nov. 2022, doi: https://doi.org/10.1016/j.agwat.2022.107889.

[20] Y. Li, J. Ye, D. Xu, G. Zhou, H. Feng, "Prediction of sap flow with historical environmental factors based on deep learning technology," *Computers and Electronics in Agriculture*, vol. 202, p. 107400, Nov. 2022, doi: https://doi.org/10.1016/j.compag.2022.107400.

[21] N. Bali, A. Singla, "Emerging trends in machine learning to predict crop yield and study its influential factors: A survey," *Archives of Computational Methods in Engineering*, vol. 29, pp. 95–112, Mar. 2021, doi: https://doi.org/10.1007/s11831-021-09569-8.

[22] R. Dubayah, J. B. Blair, S. Goetz, L. Fatoyinbo, M. Hansen, S. Healey, M. Hofton, G. Hurtt, J. Kellner, S. Luthcke, J. Armston, H. Tang, L. Duncanson, S. Hancock, P. Jantz, S. Marselis, P. L. Patterson, W. Qi, C. Silva, "The global ecosystem dynamics investigation: High-resolution laser ranging of the earth's forests and topography," *Science of Remote Sensing*, vol. 1, p. 100002, June 2020, doi: https://doi.org/10.1016/j.srs.2020.100002.

[23] F. Granata, "Evapotranspiration evaluation models based on machine learning algorithms—a comparative study," *Agricultural Water Management*, vol. 217, pp. 303–315, May 2019, doi: https://doi.org/10.1016/j.agwat.2019.03.015.

[24] S. J. Gumiere, M. Camporese, A. Botto, J. A. Lafond, C. Paniconi, J. Gallichand, A. N. Rousseau, "Machine learning vs. physics-based modeling for real-time irrigation management," *Frontiers in Water*, vol. 2, Apr. 2020, doi: https://doi.org/10.3389/frwa.2020.00008.

[25] C. A. Silva, A. T. Hudak, L. A. Vierling, E. L. Loudermilk, J. J. O'Brien, J. K. Hiers, S. B. Jack, C. Gonzalez-Benecke, H. Lee, M. J. Falkowski, *et al.*, "Imputation of individual longleaf pine (pinus palustris mill.) tree attributes from field and lidar data," *Canadian journal of remote sensing*, vol. 42, no. 5, pp. 554–573, 2016.

[26] N. Schuwirth, F. Borgwardt, S. Domisch, M. Friedrichs, M. Kattwinkel, D. Kneis, M. Kuemmerlen, S. D. Langhans, J. Martínez-López, P. Vermeiren, "How to make ecological models useful for environmental management," *Ecological Modelling*, vol. 411, p. 108784, 2019.

[27] Q. Wen, L. Sun, F. Yang, X. Song, J. Gao, X. Wang, H. Xu, "Time series data augmentation for deep learning: A survey," *arXiv preprint arXiv:2002.12478*, 2020.

[28] C. R. Ford, C. E. Goranson, R. J. Mitchell, R. E. Will, R. O. Teskey, "Modeling canopy transpiration using time series analysis: A case study illustrating the effect of soil moisture deficit on pinus taeda," *Agricultural and Forest Meteorology*, vol. 130, pp. 163–175, June 2005, doi: https://doi.org/10.1016/j.agrformet.2005.03.004.

[29] Y. Wang, G. Cao, Y. Wang, A. A. Webb, P. Yu, X. Wang, "Response of the daily transpiration of a larch plantation to variation in potential evaporation, leaf area index and soil moisture," *Scientific Reports*, vol. 9, Mar. 2019, doi: https://doi.org/10.1038/s41598-019-41186-1.

[30] R. Seidl, P. M. Fernandes, T. F. Fonseca, F. Gillet, A. M. Jönsson, K. Merganičová, S. Netherer, A. Arpaci, J.-D. Bontemps, H. Bugmann, *et al.*, "Modelling natural disturbances in forest ecosystems: a review," *Ecological Modelling*, vol. 222, no. 4, pp. 903–924, 2011.

[31] M. Austin, "Spatial prediction of species distribution: an interface between ecological theory and statistical modelling," *Ecological Modelling*, vol. 157, pp. 101–118, Nov. 2002, doi: https://doi.org/10.1016/s0304-3800(02)00205-3.

[32] G. McDermid, R. Hall, G. Sanchez-Azofeifa, S. Franklin, G. Stenhouse, T. Kobliuk, E. LeDrew, "Remote sensing and forest inventory for wildlife habitat assessment," *Forest Ecology and Management*, vol. 257, pp. 2262–2269, May 2009, doi: https://doi.org/10.1016/j.foreco.2009.03.005.

[33] F. Freese, *Linear regression methods for forest research*, vol. 17. US Department of Agriculture, Forest Service, Forest Products Laboratory, 1964.

[34] J. J. Faraway, *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. CRC press, 2016.

[35] M. Fortin, "Population-averaged predictions with generalized linear mixed-effects models in forestry: an estimator based on gauss- hermite quadrature," *Canadian Journal of Forest Research*, vol. 43, no. 2, pp. 129–138, 2013.

[36] F. J. Oddi, F. E. Miguez, L. Ghermandi, L. O. Bianchi, L. A. Garibaldi, "A nonlinear mixed-effects modeling approach for ecological data: Using temporal dynamics of vegetation moisture as an example," *Ecology and evolution*, vol. 9, no. 18, pp. 10225–10240, 2019.

[37] F. Ellsäßer, A. Röll, J. Ahongshangbam, P.-A. Waite, B. Schuldt, D. Hölscher, "Predicting tree sap flux and stomatal conductance from drone-recorded surface temperatures in a mixed agroforestry system—a machine learning approach," *Remote Sensing*, vol. 12, no. 24, p. 4070, 2020.

[38] T. J. Fernandes, A. D. D. Campo, R. Garcia-Bartual, M. Gonzalez-Sanchis, "Coupling daily transpiration modelling with forest management in a semiarid pine plantation," *iForest-Biogeosciences and Forestry*, vol. 9, no. 1, p. 38, 2015.

[39] R. Whitley, B. Medlyn, M. Zeppel, C. Macinnis-Ng, D. Eamus, "Comparing the penman–monteith equation and a modified jarvis–stewart model with an artificial neural network to estimate stand-scale transpiration and canopy conductance," *Journal of Hydrology*, vol. 373, no. 1-2, pp. 256–266, 2009.

[40] P. Jutras, S. O. Prasher, G. R. Mehuys, "Artificial neural network prediction of street tree growth patterns," *Transactions of the ASABE*, vol. 53, no. 3, pp. 983–992, 2010.

[41] P. Semenzato, D. Cattaneo, M. Dainese, "Growth prediction for five tree species in an italian urban forest," *Urban Forestry and Urban Greening*, vol. 10, no. 3, pp. 169–176, 2011.

[42] L. Li, S. Chen, C. Yang, F. Meng, N. Sigrimis, "Prediction of plant transpiration from environmental parameters and relative leaf area index using the random forest regression algorithm," *Journal of Cleaner Production*, vol. 261, p. 121136, 2020.

[43] H. Chen, J. J. Huang, E. McBean, "Partitioning of daily evapotranspiration using a modified shuttleworth-wallace model, random forest and support vector regression, for a cabbage farmland," *Agricultural Water Management*, vol. 228, p. 105923, 2020.

[44] H. L. Munro, C. R. Montes, K. J. Gandhi, "A new approach to evaluate the risk of bark beetle outbreaks using multi-step machine learning methods," *Forest Ecology and Management*, vol. 520, p. 120347, 2022.

[45] V. Shikuku, *Artificial Intelligence Applications in Water Treatment and Water Resource Management*. IGI Global, 2023.

[46] F. Ghafarian, R. Wieland, D. Lüttschwager, C. Nendel, "Application of extreme gradient boosting and shapley additive explanations to predict temperature regimes inside forests from standard open-field meteorological data," *Environmental Modelling & Software*, vol. 156, p. 105466, 2022.

[47] D. Gupta, N. Gujre, S. Singha, S. Mitra, "Role of existing and emerging technologies in advancing climate-smart agriculture through modeling: A review," *Ecological Informatics*, p. 101805, 2022.

[48] N. Khan, M. A. Kamaruddin, U. U. Sheikh, Y. Yusup, M. P. Bakht, "Environment-based oil palm yield prediction using k-nearest neighbour regression," in *2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, 2022, pp. 1–6, IEEE.

[49] S. Shataee, S. Kalbi, A. Fallah, D. Pelz, "Forest attribute imputation using machine-learning methods and aster data: comparison of k-nn, svr and random forest regression algorithms," *International journal of remote sensing*, vol. 33, no. 19, pp. 6254–6280, 2012.

[50] S. K. Kiangala, Z. Wang, "An effective adaptive customization framework for small manufacturing plants using extreme gradient boosting-xgboost and random forest ensemble learning algorithms in an industry 4.0 environment," *Machine Learning with Applications*, vol. 4, p. 100024, 2021.

[51] C. Su, X. Wu, X. Tang, J. Hu, "Growth height prediction for the trees under overhead lines based on deep learning algorithm," in *2018 International Conference on Power System Technology, POWERCON 2018 - Proceedings*, 2019, pp. 3693–3699.

[52] M. L. Borowiec, R. B. Dikow, P. B. Frandsen, A. McKeeken, G. Valentini, A. E. White, "Deep learning as a tool for ecology and evolution," *Methods in Ecology and Evolution*, vol. 13, no. 8, pp. 1640–1660, 2022.

[53] V. Sagan, M. Maimaitijiang, P. Sidike, K. Eblimit, K. T. Peterson, S. Hartling, F. Esposito, K. Khanal, M. Newcomb, D. Pauli, *et al.*, "Uav-based high resolution thermal imaging for vegetation monitoring, and plant phenotyping using ici 8640 p, flir vue pro r 640, and thermomap cameras," *Remote Sensing*, vol. 11, no. 3, p. 330, 2019.

[54] T. Wen, J.-H. Li, Q. Wang, Y.-Y. Gao, G.-F. Hao, B.-A. Song, "Thermal imaging: The digital eye facilitates high-throughput phenotyping traits of plant growth and stress responses," *Science of The Total Environment*, p. 165626, 2023.

[55] J. Wang, M. Bretz, M. A. A. Dewan, M. A. Delavar, "Machine learning in modelling land-use and land cover-change (lulcc): Current status, challenges and prospects," *Science of the Total Environment*, vol. 822, p. 153559, 2022.

[56] M. Y. Chia, Y. F. Huang, C. H. Koo, J. L. Ng, A. N. Ahmed, A. El-Shafie, "Long-term forecasting of monthly mean reference evapotranspiration using deep neural network: A comparison of training strategies and approaches," *Applied Soft Computing*, vol. 126, p. 109221, 2022.

[57] I. H. Sarker, "Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions," *SN Computer Science*, vol. 2, no. 6, p. 420, 2021.

[58] L. Zhao, S. Walkowiak, W. G. D. Fernando, "Artificial intelligence: A promising tool in exploring the phytomicrobiome in managing disease and promoting plant health," *Plants*, vol. 12, no. 9, p. 1852, 2023.

[59] H. Zeng, Q. Zhu, Y. Ding, H. Hu, L. Chen, X. Xie, M. Chen, Y. Yao, "Graph neural networks with constraints of environmental consistency for landslide susceptibility evaluation," *International Journal of Geographical Information Science*, vol. 36, no. 11, pp. 2270–2295, 2022.

[60] Y. Li, J. Ye, D. Xu, G. Zhou, H. Feng, "Prediction of sap flow with historical environmental factors based on deep learning technology," *Computers and Electronics in Agriculture*, vol. 202, p. 107400, 2022.

[61] H. Li, Y. Mao, Y. Wang, K. Fan, H. Shi, L. Sun, J. Shen, Y. Shen, Y. Xu, Z. Ding, "Environmental simulation model for rapid prediction of tea seedling growth," *Agronomy*, vol. 12, no. 12, p. 3165, 2022.

[62] A. Augustaitis, I. Augustaitienė, G. Činga, J. Mažeika, R. Deltuvas, R. Juknys, A. Vita, "Did the ambient ozone affect stem increment of scots pines (pinus sylvestris l.) on territories under regional pollution load? step iii of lithuanian studies," *The Scientific World Journal*, vol. 7, pp. 58–66, 2007, doi: https://doi.org/10.1100/tsw.2007.55.

[63] A. Augustaitis, I. Augustaitienė, A. Kliučius, G. Pivoras, B. Bendoravičius, D. Šopauskienė, D. Jasinevičiene, I. Baužienė, I. Eitminavičiutė, K. Arbačiauskas, R. Mažeikytė, "N deposition, balance and benefit in the forest ecosystem of main landscape types of lithuania," *International Journal of Environmental Studies*, vol. 65, no. 3, pp. 339–359, 2008, doi: https://doi.org/10.1080/00207230701862801.

[64] A. Augustaitis, I. Augustaitienė, A. Kliučius, G. Pivoras, D. Sopauskiene, R. Girgždienė, "The seasonal variability of air pollution effects on pine conditions under changing climates," *European Journal of Forest Research*, vol. 129, pp. 431–441, 05 2010, doi: https://doi.org/10.1007/s10342-009-0319-x.

[65] A. Augustaitis, I. Augustaitienė, M. Baugarten, S. Bičenkienė, R. Girgždienė, G. Kulbokas, E. Linkevičius, V. Marozas, M. Mikalaju¯ nas, G. Mordas, G. Mozgeris, E. Petrauskas, A. Pivoras, G. Šidlauskas, V. Ulevičius, A. Vitas, R. Matyssek, "Tree-ring formation as an indicator of forest capacity to adapt to the main threats of environmental changes in lithuania," *Science of The Total Environment*, vol. 615, pp. 1247–1261, 2018, doi: https://doi.org/10.1016/j.scitotenv.2017.09.169.

[66] M. Baumgarten, B. D. Hesse, I. Augustaitienė, V. Marozas, G. Mozgeris, S. Byčenkienė, G. Mordas, A. Pivoras, G. Pivoras, D. Juonytė, V. Ulevičius, A. Augustaitis, R. Matyssek, "Responses of species-specific sap flux, transpiration and water use efficiency of pine, spruce and birch trees to temporarily moderate dry periods in mixed forests at a dry and wet forest site in the hemi-boreal zone," *Journal of Agricultural Meteorology*, vol. 75, no. 1, p. 13–29, 2019, doi: https://doi.org/10.2480/agrmet.d-18-00008.

[67] Z. Gulbinas, M. Samuila, "Integrated monitoring in small wooded catchments in lithuania," *Geological Quarterly*, vol. 46, no. 1, pp. 81–97, 2012.

[68] L. Nunes, J. Tomé, M. Tomé, "Prediction of annual tree growth and survival for thinned and unthinned even-aged maritime pine stands in portugal from data with different time measurement intervals," *Forest Ecology and Management*, vol. 262, no. 8, pp. 1491–1499, 2011.

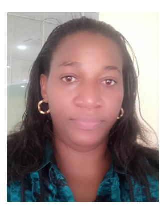
[69] S. Z. Martinez, C. A. C. Coello, "A multi-objective evolutionary algorithm based on decomposition for constrained multi-objective optimization," in *2014 IEEE Congress on evolutionary computation (CEC)*, 2014, pp. 429–436, IEEE.

Rytis Maskeliūnas

Rytis Maskeliūnas received the Ph.D. degree in computer science, in 2009. He is currently a Professor with the Department of Multimedia Engineering, Kaunas University of Technology, Kaunas, Lithuania and an Invited Professor with the Faculty of Applied Mathematics, Silesian University of Technology, Poland. He is author or coauthor of more than 200 refereed scientific articles and serves as a Reviewer/Committee Member for various refereed journals. His main areas of scientific research are multimodal signal processing, modeling, development and analysis of associative, multimodal interfaces, also targeted at elderly, and people with major disabilities.

Robertas Damaševičius

Robertas Damaševičius received the Ph.D. degree in informatics engineering from the Kaunas University of Technology, Lithuania, in 2005. He is currently a Professor with the Department of Software Engineering, Kaunas University of Technology, Lithuania. His research interests include sustainable software engineering, human–computer interfaces, assisted living, data mining, and machine learning. He is the author of more than 300 articles as well as a monograph published by Springer. He is also the Editor-in-Chief of the Information Technology and Control Journal.



Modupe Odusami

Modupe Odusami is currently pursuing a Ph.D degree in informatics engineering from Kaunas University of Technology. Her research interest includes Machine learning, multimodal imaging, and computer vision. She is the author or co-author of over 40 research papers and contributes to the academic community as a journal reviewer.

Diana Sidabrienė

Diana Sidabrienė is currently pursuing a Ph.D. degree in Forestry at Vytautas Magnus University, where she focuses on the impact of climate change on tree physiology. Her impactfull research focuses on understanding how different tree species adapt to varying climatic conditions and how these adaptations influence their eco-physiological responses. Diana is an author of 12 research papers. She has contributed to several important studies on forest resilience and has presented her findings at multiple international conferences. Her work aims to develop sustainable forest management practices that enhance tree health and productivity in the face of global climate shifts.

Algirdas Augustaitis

Algirdas Augustaitis (born 3 December 1962 in Kaunas, Lithuania) - Lithuanian forester, ecologist, professor at the Faculty of Forestry and Ecology, Vytautas Magnus University, Faculty of Agriculture. Graduated from the Lithuanian Academy of Agriculture in 1985. From 1985 he worked at the Academy (which later became a university). In 1992 he was appointed Head of the Forest Monitoring Laboratory. In 2007 he was awarded the title of Professor. In 2011-2012 he was Director of the Institute of Environment. Research interests: forest science, ecology, forest development under global change. In 1993, he developed a comprehensive monitoring programme for natural forest ecosystems, and from 2002 to 2012 he was the coordinator of this programme in Lithuania. Since 1997 he has been a member of the International Union of Forest Scientists. He has published more than 100 scientific articles. With co-authors, he prepared the monographs "The State of Lithuanian Forests and Factors Conditioning It" (1999), "Monitoring of Forest Ecosystems in Lithuania" (1999), "Comprehensive Monitoring of Conditionally Natural Ecosystems: Ten-Year Results" (2006).

Gintautas Mozgeris

Gintautas Mozgeris received the Ph.D. degree in forestry from the Lithuanian University of Agriculture in 1995. He is currently a Professor with the Department of Forest Sciences, Vytautas Magnus University, Kaunas, Lithuania. He is author or coauthor of more than 70 refereed scientific articles. His main areas of scientific research are spatial modelling, photogrammetry and remote sensing, land use change and habitat modelling, conservation planning, scenario development, forestry decision support systems.