EAI Endorsed Transactions
on Scalable Information Systems

Research Article EAI.EU

Forest Information Modeling: A Novel Approach to Sustainable Forest Management

Robertas Damaševičius and Rytis Maskeliūnas*

Center of Excellence Forest 4.0, Center of Real Time Computer Systems, Kaunas University of Technology, Lithuania

Abstract

This research introduces Forest Information Modeling (FIM) as an innovative approach to sustainable forest management. FIM is depicted as a comprehensive digital framework that integrates various data layers and technologies to enhance the efficiency, accuracy, and sustainability of forest management practices. Key components of FIM include ecological data collection, spatial representation, process modeling, decision support systems, and stakeholder engagement, all aimed at facilitating informed decision-making and efficient resource utilization. The study explores the theoretical underpinnings of FIM, its practical applications through case studies, and the construction of FIM using real-world datasets. Practical illustration using the Biomass tree data base is presented. The study addresses the challenges, potential impacts, and future directions of FIM in the context of global forest management and conservation efforts. The findings underscore FIM's potential to transform forest management practices by improving decision-making processes, promoting environmental sustainability, and fostering stakeholder collaboration.

Received on 12 November 2024; accepted on 09 October 2025; published on 22 December 2025

Keywords: Forest Information Modeling, Ecosystem Services, Smart Forestry, Forest 4.0, forest management, information systems, forest data analysis

doi:10.4108/eetsis.7811

1. Introduction

Integration of digital technologies in sustainable forest management has led to the concepts of Smart Forestry and Forest 4.0. These paradigms are increasingly becoming central to the sustainable management of forest resources, driven by advances in digital technologies and network infrastructure. Smart Forestry and Forest 4.0 encompass a holistic approach that integrates advanced system intelligence and a digital ecosystem, designed to enhance the efficiency and sustainability of forest supply chains [1]. The genesis of Forest 4.0 lies in the inspiration drawn from Industry 4.0. It aims to revolutionize sustainable forest management by using smart devices for monitoring purposes, including fire detection, thus addressing the critical challenges of environmental sustainability and climate change [2, 3]. This new era of forestry is characterized by adaptive forest management that amplifies the potential of forest ecosystems to remain resilient and mitigate the effects of climate change, through the use of innovative sensors and monitoring tools [4]. Furthermore, Forest 4.0 aims to develop smart carbon management systems, focusing on both adaptation and mitigation in response to climate change, thus improving the ecological integrity and carbon sequestration capacity of forests [5]. Smart urban forest management further extends this concept by highlighting the role of digital infrastructure, citizen engagement, and innovative monitoring techniques using sensors and IoT technologies to maximize the benefits of urban forests [6]. Climate-smart forestry, a subset of this approach, integrates climate change mitigation and adaptation into forest management strategies, with the objective of maintaining and improving the contributions of forests to people and global agendas [7–9]. This framework is crucial for managing forested landscapes in response to climate change, with a focus on maintaining ecosystem health and vitality [10].

*Corresponding author. Email: robertas.damasevicius@vdu.lt

Smart forestry and forest management have been revolutionized by the integration of digital technologies and network infrastructure, aiming to enhance efficiency, sustainability, and productivity while addressing environmental challenges and promoting biodiversity conservation [11]. An essential component is the use of Internet of Things (IoT) devices with sensors deployed in forests to collect real-time data on environmental conditions such as temperature, humidity, soil moisture and tree growth rates [12]. Remote sensing technologies such as satellite imagery and drones provide detailed information on forest conditions in large areas, facilitating forest health monitoring, change in land cover, inventory assessment, and identification of areas of deforestation or risk of fire, which helps develop effective management plans and interventions for sustainable forestry [13, 14]. Geographic Information Systems (GIS) integrate spatial data with forest-related information to aid decision-making and planning [15]. GIS tools map forest resources, analyze land use patterns, plan operations and monitor habitats, identify suitable tree planting locations, design low impact logging roads, and optimize practices to improve ecosystem services and biodiversity [16]. Big data analytics processes large forest-related datasets to extract insights and patterns [17], optimize management practices, forecast timber yields, and guide harvesting decisions. Artificial intelligence (AI) algorithms analyze complex data to predict forest growth, estimate carbon sequestration, and automate tasks such as species identification, improve efficiency and reduce labor [18]. Blockchain technology ensures transparency and traceability in forestry supply chains, preventing illegal logging, and promoting sustainability [19]. Mobile applications provide real-time information, enable field observations, and facilitate communication among forest workers, improving efficiency and safety [20].

The domain of smart forestry has several gaps. One of the primary gaps is the accurate and comprehensive data collection from vast and often inaccessible forest areas. The reliance on remote sensing technologies, while beneficial, can sometimes provide data that is limited in resolution and specificity, affecting the accuracy of forest monitoring and management [21]. Another significant gap is the integration and analysis of large and diverse datasets. Forest ecosystems are complex, and the data that inform their management comes from varied sources, including satellite imagery, ground-based sensors, and ecological surveys. Developing models that can effectively integrate and analyze this data to provide actionable insights remains a complex task [22]. Moreover, there is a need for more sophisticated models that can predict the impacts of climate change on forests and guide adaptation and mitigation strategies [23]. The gap between technological advancements and their practical application also poses a challenge. While there are numerous innovative technologies available for smart forestry, such as drones for aerial surveillance and IoT devices for real-time monitoring, their deployment on a large scale is often limited by financial, technical, and infrastructural constraints [24].

Forest Information Modeling (FIM) is a digital representation of forest ecosystems, incorporating various data sources and modeling techniques to provide a comprehensive understanding of forest resources, dynamics, and management practices [25]. Similarly to building information modeling (BIM) in construction [26], FIM integrates spatial and non-spatial data, such as forest inventory, remote sensing imagery, ecological models, and socioeconomic factors, into a unified platform. This model enables forest managers, policymakers, and stakeholders to visualize, analyze, and simulate different scenarios for forest management, facilitating informed decision making, sustainable resource use, and biodiversity conservation. FIM improves collaboration, efficiency, and effectiveness in forestry practices by providing a holistic view of forest ecosystems and their interactions with the environment.

FIM plays a crucial role in the paradigm of sustainable forest management. Its necessity comes from the need to account for the heterogeneity of agents, context-specific preferences, and the diverse needs within forest ecosystems. Kant [27] emphasizes that FIM is integral in incorporating these diverse elements, which are essential for holistic forest management. Furthermore, FIM is indispensable in decision-making processes related to forest dynamics, landscape, and spatial modeling, as well as in participatory models for collaborative management. Mendoza and Vanclay [28] highlight the importance of FIM in these aspects, underlining its role in facilitating informed and participatory management decisions. Korzukhin et al. [29] argue that FIM is necessary to meet the challenges of ecosystem management. They point out that both the empirical and process models described by FIM can provide diverse ways of understanding and managing forest data, which is crucial in the face of environmental challenges. Albers et al. [30] focus on the use of FIM for sustainability assessment and the feasibility of management actions under uncertainty. Their work underscores the importance of FIM in generating reliable information flows for tropical forest management decisions. Chen et al. [31] demonstrate the utility of FIM in accurately modeling forest dynamics, particularly under changing environmental conditions. They suggest that the incorporation of machine learning and artificial neural networks in FIM can significantly enhance the modeling of forest growth relationships, which is a cornerstone in sustainable forest management. Ansary et al.[32] discuss the technocenological approach in the sustainability management of the forest industry, where FIM is used to design the structural architecture of subsystems, guiding various management decisions. Collectively, these studies provide a rationale for the integration of FIM in sustainable forest management, where FIM emerges as a key tool in addressing the complexities of forest ecosystems, helping decision-making processes, and ensuring the sustainability and resilience of forest resources.

The motivation behind this study stems from the growing need to enhance the sustainability, precision, and resilience of forest management in the face of climate change, biodiversity loss, and increasing demands on forest resources. Traditional forest inventory systems and isolated data analytics approaches lack the integration, scalability, and predictive power required for modern decision-making. Inspired by the success of Building Information Modeling (BIM) in the construction industry, this research introduces Forest Information Modeling (FIM) as a transformative digital framework that unifies ecological, spatial, and process-based data into a coherent model. By using advances in remote sensing, machine learning, and semantic modeling, FIM aims to bridge the gap between forest ecosystem complexity and actionable insights, enabling more informed policy-making, resource allocation, and long-term sustainability strategies.

The main objective of this study is to explore and articulate the concept of FIM as a novel approach to sustainable forest management. This includes examining the role of FIM in improving the efficiency, precision, and sustainability of forest management practices. Key objectives are:

- To understand the technological underpinnings and methodologies of FIM, including its integration with advanced data analytics and machine learning techniques.
- To evaluate the effectiveness of FIM in addressing the challenges of sustainable forest management, such as biodiversity conservation, adaptation to climate change, and resource optimization.
- To identify the benefits and possible impacts of FIM on forest management, particularly in terms of decision making, policy formulation, and stakeholder engagement.

This study focuses on the conceptual framework, methodologies and applications of FIM in the context of sustainable forest management. The scope includes:

- A comprehensive narrative review of current literature and research on FIM and its role in sustainable forestry.
- Analysis of case studies where FIM has been implemented, to understand its practical applications and results.
- Demonstration of the FIM process using a real-world dataset.

This study makes several key contributions to the field of sustainable forest management. First, it introduces the concept of Forest Information Modeling (FIM) as a formal, integrative framework analogous to Building Information Modeling, tailored to the complexities of forest ecosystems. Second, it presents a rigorous mathematical formulation of both the Forest Information Model (FIM) and the Forest Object Model (FOM), enabling structured representation and analysis of ecological units, attributes, processes, and temporal dynamics. Third, it constructs a forest ontology based on real-world biomass data, offering a semantic model for ecological relationships. Fourth, it implements and evaluates a Functional Structural Forest Biomass Model (FSFBM) using a large-scale scientific dataset, applying machine learning (linear regression) to predict tree biomass with high accuracy. Finally, it maps dataset components to conceptual FIM layers, providing a practical blueprint for digital forest management systems that integrate data, models, and decision-support tools.

The article is structured as follows: Section 2 presents a theoretical framework of FIM. Section 3 demonstrates FIM construction using a real-world forest dataset. Finally, the concluding section summarizes the key findings of the study and suggests areas for future research in FIM.

2. Theoretical Framework of Forest Information Modeling (FIM)

In the sections bellow, we provide a formal mathematical definition of FIM by introducing its core concepts and principles.

2.1. Defining FIM: Concepts and Principles

A forest ecosystem, denoted as $\mathcal{F}$, is defined as a spatially distributed, dynamic, and complex system consisting of n distinct ecological units, represented as $\mathcal{U} = \{U_1, U_2, \dots, U_n\}$. Each ecological unit U_i is characterized by a set of state variables, $\mathcal{S}_i = \{S_{i1}, S_{i2}, \dots, S_{im}\}$, where m represents the number of state variables within each unit. The state variables may include measures of tree species composition, biomass, carbon content, and environmental factors, among others.

A Forest Information Model (FIM), denoted as $\mathcal{M}$, is a mathematical representation of the forest ecosystem $\mathcal{F}$. It is defined as a tuple $\mathcal{M} = (\mathcal{U}, \mathcal{S}, \mathcal{P}, \mathcal{T})$, where:

- $\mathcal{U}$ is the set of ecological units within the forest ecosystem.
- $\mathcal{S}$ is the set of state variables characterizing each ecological unit.
- $\mathcal{P}$ represents the set of processes governing the temporal dynamics of the forest ecosystem. These processes can be described by a system of differential equations:

$$\frac{dS_{ij}}{dt} = f_{ij}(\mathcal{S}, \mathcal{P}, \mathcal{T}) \tag{1}$$

where S_{ij} is the j-th state variable of ecological unit U_i, f_{ij} is the functional relationship describing the dynamics, and $\mathcal{T}$ represents time.

- $\mathcal{T}$ is the temporal domain over which the model operates.

The principles underlying FIM include:

1. **Ecological Realism**: The model $\mathcal{M}$ should strive to capture the ecological complexity and realism of the forest ecosystem $\mathcal{F}$ by incorporating relevant ecological processes and interactions.
2. **Data Integration**: FIM should integrate diverse sources of data, including remote sensing, field measurements, and ecological surveys, to parameterize and validate the model.
3. **Spatial Heterogeneity**: Recognizing the spatial heterogeneity of forests, FIM should account for variations in ecological processes and state variables across different ecological units.
4. **Temporal Dynamics**: FIM should simulate the temporal dynamics of the forest ecosystem, allowing for predictions and scenario analysis.
5. **Interdisciplinary Collaboration**: FIM should involve collaboration among ecologists, mathematicians, computer scientists, and domain experts to develop and refine the model.

2.2. Forest Object Model (FOM)

The Forest Object Model (FOM) is a descriptive framework designed to represent, organize, and manage data related to forest ecosystems. It serves as the foundational structure for FIM and provides a formal representation of the forest environment. In this section, we present a comprehensive and elaborate mathematical definition of the FOM.

FOM is defined as a tuple $\mathcal{FOM} = (\mathcal{U}, \mathcal{E}, \mathcal{A}, \mathcal{D})$, where:

- $\mathcal{U}$ represents the set of ecological units within the forest ecosystem. It is defined as:

$$\mathcal{U} = \{U_1, U_2, \ldots, U_n\} \tag{2}$$

where n is the total number of ecological units.

- $\mathcal{E}$ denotes the set of ecological attributes, which characterize each ecological unit U_i within $\mathcal{U}$. It is defined as:

$$\mathcal{E} = \{E_1, E_2, \ldots, E_m\} \tag{3}$$

where m is the total number of ecological attributes. Each ecological attribute E_j is a tuple (N_j, T_j, V_j), where: - N_j represents the name of the attribute. - T_j defines the data type of the attribute, e.g., numeric, categorical, or temporal. - V_j represents the value or values associated with the attribute, depending on its data type.

- $\mathcal{A}$ stands for the set of relationships between ecological units U_i and ecological attributes E_j. It is defined as a function:

$$\mathcal{A} : \mathcal{U} \times \mathcal{E} \rightarrow \mathcal{V} \tag{4}$$

where $\mathcal{V}$ represents the set of possible values that the relationship between an ecological unit and an ecological attribute can take.

- $\mathcal{D}$ represents the temporal domain of the FOM, allowing modeling of temporal dynamics. It defines the time intervals during which data about the forest ecosystem is collected and updated.

The FOM contains forest-related information in a structured and machine-readable format, facilitating automatic processing and analysis. The FOM serves as a structured representation of the forest ecosystem, allowing the integration of ecological data, attributes, and relationships. Allows for the precise characterization of ecological units in terms of their attributes and their changes over time. The FOM provides the foundation for FIM, facilitating data management, analysis, and decision support in forest management, conservation, and research.

The formal mathematical definition of the FOM provides a rigorous framework for representing forest ecosystems in a structured and organized way. It captures the essential components of ecological units, attributes, relationships, and temporal dynamics, enabling the comprehensive modeling and analysis of forest environments.

3. Case Study: Forest Information Modeling using Real-World Dataset

3.1. Dataset

The dataset hosted on PANGAEA [33] is a collection of scientific data curated by a collaboration between the

Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research (AWI), and the Center for Marine Environmental Sciences at the University of Bremen (MARUM). It's supported by various prestigious organizations, including the European Commission, the Federal Ministry of Education and Research (BMBF), the Deutsche Forschungsgemeinschaft (DFG), and the International Ocean Discovery Program (IODP).

The dataset consists of 9613 entries with 27 columns, each representing different attributes of tree biomass [34]. Here's a brief overview of the dataset: The dataset includes a variety of data types, such as numerical (integers and floats) and categorical (objects). Key columns include 'ID', 'Species', 'Tree age', 'DBH' (Diameter at Breast Height), 'H tree' (Tree Height), 'Hcr', 'Dcr', 'Vtot' (Total Volume), 'Vbark' (Bark Volume), 'Origin', various biomass components (like 'Pst', 'Pbark', 'Pbr', 'Pf', 'Pabo', 'Proot', 'Ptot'), 'Location', 'Country', 'Latitude', 'Longitude', 'Altitude (m.s.l.)', 'Tree number / ha', 'Reference', 'Notes', 'Ecoregion', and 'ID_Plot'.

Table 1. Explanation of Dataset Variables

Variable	Description
Species	The species of the tree.
DBH	Diameter at Breast Height - a standard measure of a tree's diameter measured at 1.3 meters above the ground.
Height	The height of the tree.
Age	The age of the tree.
Latitude	The latitude coordinate where the tree is located.
Longitude	The longitude coordinate where the tree is located.
Altitude	The altitude (elevation from sea level) where the tree is located.
Pst	Biomass of the stem.
Pbark	Biomass of the bark.
Pbr	Biomass of the branches.
Pf	Biomass of the foliage.
Proot	Biomass of the roots.
Ptot	Total biomass of the tree.

3.2. Forest Domain Ontology

The forest ontology, as represented in PlantUML, provides a structured visual representation of the relationships and attributes within the forest ecosystem. Each entity in the diagram corresponds to a key concept in the forest dataset, and the relationships between these entities capture how these concepts are interconnected. Here's an explanation of this forest ontology derived from the metadata of the analyzed dataset:

Entities and Their Attributes in the ontology are as follows:

- Tree: This is the central entity, representing individual trees in the forest. Attributes like Age, Diameter at Breast Height (DBH), and Height provide specific details about each tree.
- Species: This entity represents the species classification of each tree. Attributes might include the species' Name and Taxonomic Classification, providing biological and taxonomic context.
- Location: This entity captures the geographic positioning of each tree. Attributes like Latitude, Longitude, and Altitude are crucial for understanding the spatial distribution of trees and the influence of geographic factors on tree growth and health.
- BiomassComponent: This entity represents different components of a tree's biomass (such as stem, bark, branches, foliage, roots), with attributes like Type (of biomass) and Mass (the biomass quantity).
- EnvironmentalFactor: This entity encompasses environmental variables impacting the trees, with attributes like Type (e.g., temperature, precipitation) and Value (the measurement of the environmental factor).

The relationships between entities are as follows:

- Tree and Species: The "is a member of" relationship indicates that each Tree belongs to a certain Species. This connection is vital for ecological studies, biodiversity assessment, and conservation efforts.
- Tree and Location: The "is located at" relationship links each Tree to its Location, signifying the geographical context of each tree, which is important for studying environmental effects on forest growth.
- Tree and BiomassComponent: The "has" relationship signifies that each Tree comprises several BiomassComponents. This relationship is key to understanding biomass distribution and carbon sequestration in forests.
- Tree and EnvironmentalFactor: The "is influenced by" relationship reflects that various EnvironmentalFactors affect each Tree. This relationship is crucial for studying the impact of environmental conditions on tree health and forest ecology.

The forest ontology offers a comprehensive view of the forest ecosystem, capturing the essential elements and their interrelations. It serves as a foundational framework for analyzing the dataset, facilitating ecological research, forest management, and environmental studies. This structured approach allows for a deeper understanding of the complex dynamics within forest ecosystems and supports informed decision-making in forestry and conservation.

Transforming the data model of a forest dataset into a forest ontology involves conceptualizing the dataset's structure and relationships in an ontological framework. An ontology in this context is a structured representation of knowledge as a set of concepts within a domain, and the links between those concepts. In this ontology (Figure 1), Tree is the central entity linked to various attributes and related to other entities like Species and Location. The ontology can be expanded and refined based on the dataset's complexity and the specific requirements of the analysis or application. This structured approach helps in organizing the dataset for better understanding, querying, and analysis, particularly in ecological and environmental research domains.

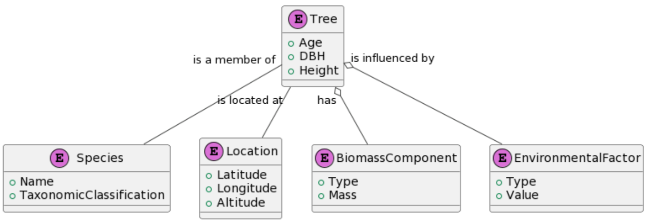

Figure 1. Forest Ontology created from dataset metadata

3.3. Descriptive Statistics

The descriptive statistics of the dataset is presented in Table 2. Here: Tree Age ranges from young to old trees, with some trees as old as 55 years. DBH varies significantly, indicating a wide range of tree sizes. Tree Height ('H tree'): Also shows considerable variation. Biomass Components, like 'Pst', 'Pbark', 'Pbr', 'Pf', 'Pabo', 'Proot', 'Ptot', show wide-ranging values, indicating diversity in biomass distribution among different parts of the trees. Geographical Data defined by 'Latitude' and 'Longitude' indicate a geographical spread of the data, which could be useful for ecological and environmental studies. Tree Density ('Tree number / ha') shows a large variation, which might be indicative of different forest densities.

The histograms from the dataset (Figure 2) provide the distribution patterns for each key attribute, revealing insights into the forest's characteristics. The distribution of tree age is likely skewed, with a higher number of younger trees compared to older ones, a common pattern in natural forests due to higher mortality rates at younger stages and selective

Table 2. Descriptive Statistics of the Biomass Tree Dataset

Statistic	Tree Age	DBH	H Tree	Vtot	Pst	Pbark	Pbr	Pf
Count	9575	9518	8625	7169	7466	4799	8862	8896
Mean	48.80	14.22	12.70	249.00	107.85	12.69	15.70	5.63
Std	39.66	10.61	7.41	512.50	227.57	20.50	39.90	13.03
Min	3.00	0.00	0.13	0.02	0.00	0.00	0.00	0.00
25%	24.00	6.50	6.90	16.90	4.97	1.20	0.98	0.59
50% (Median)	36.00	12.00	11.60	68.40	29.43	4.90	3.86	2.04
75%	63.00	19.50	18.00	262.00	120.85	15.65	13.94	5.70
Max	430.00	98.00	44.20	6984.00	4122.00	280.00	1091.80	305.00

logging practices. This trend is indicative of the forest's regeneration dynamics and can be influenced by various ecological factors and forest management practices. Deviations in this pattern, such as spikes or irregularities, could point to specific environmental events or human interventions that have impacted the age structure of the forest.

The Diameter at Breast Height (DBH) follows a right-skewed distribution, reflecting the natural growth stages of a forest where smaller, younger trees are more abundant than larger, older ones. This distribution pattern is crucial for understanding the growth stages and health of the forest. Similarly, tree height is expected to demonstrate a right-skewed distribution, as shorter trees are more prevalent in a forest setting. Variations in tree height distribution can be attributed to differences in species, site quality, or environmental conditions influencing tree growth.

The total volume (Vtot) of trees is another critical aspect, exhibiting a skewed distribution with many trees having lower volumes, indicative of the overall biomass and maturity of the forest. This attribute, combined with the distributions of different biomass components like stem, bark, branch, and foliage biomass (Pst, Pbark, Pbr, Pf), can offer deeper insights into the forest's biomass allocation patterns. These distributions can vary based on species composition, age structure, and environmental conditions, providing valuable information on how different tree species allocate resources for growth.

The most intriguing aspect of this analysis lies in understanding the forest dynamics, health, and biomass allocation trends. The distribution patterns of age, DBH, height, and biomass components reflect the ecological processes, growth patterns, and environmental impacts on the forest ecosystem. Variations and anomalies in these distributions are not just data points but potential indicators of forest health, management practices, or environmental stressors.

3.4. Species Analysis

The analysis of species distribution (Figure 2) and species-specific characteristics (Table 3) within the dataset reveals fascinating insights into the forest's composition and the varying attributes of different tree species. The dataset exhibits a diverse range of tree species, with Pinus sylvestris L. being the most

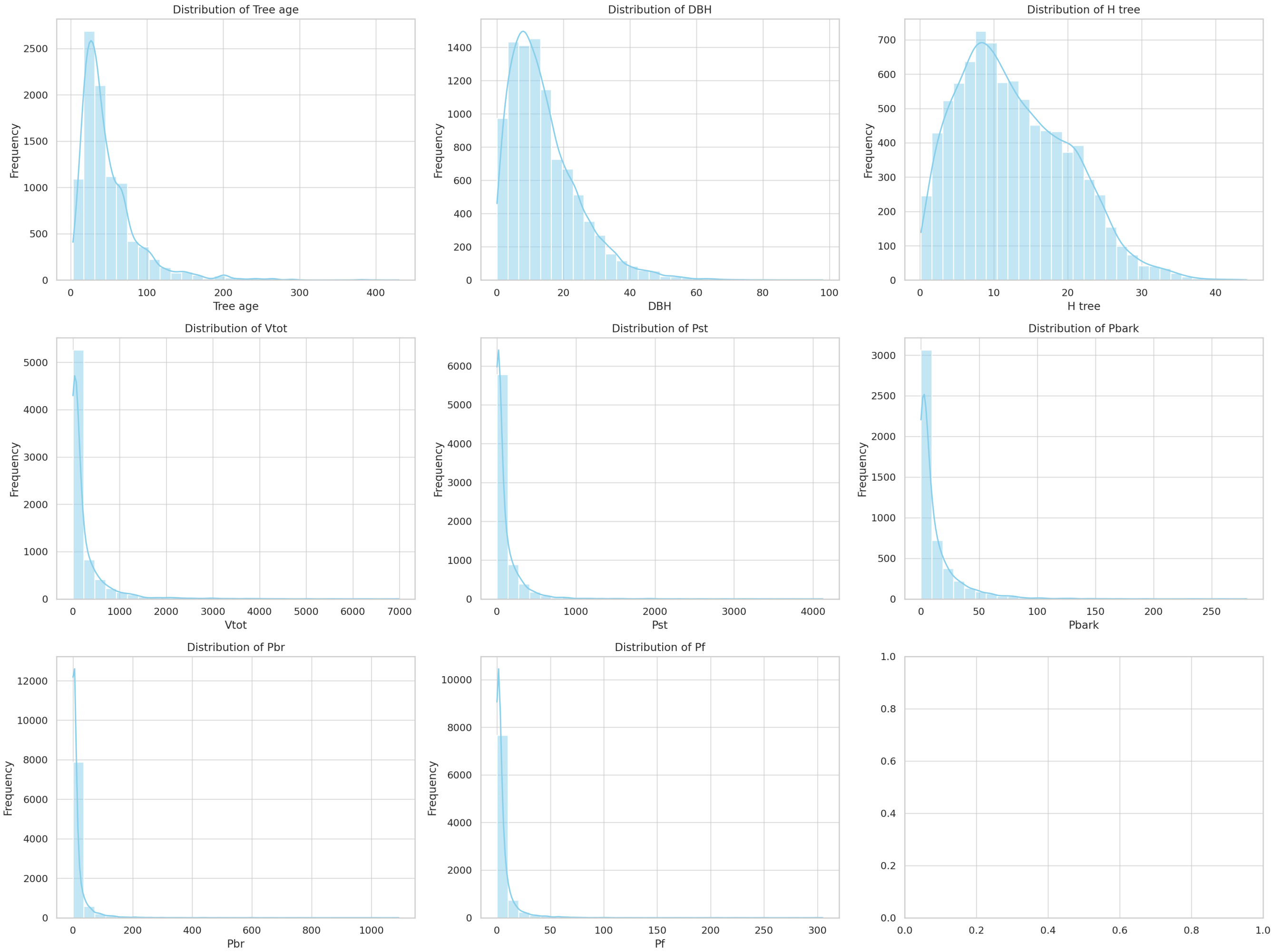

Figure 2. Probability distributions of data

prevalent, accounting for a significant portion of the entries. This is followed by other species like Betula alba L., Picea abies (L.) Karst., and Populus tremula L., each contributing substantially to the dataset. The presence of numerous species, each with varying frequencies, underscores the ecological diversity represented in the dataset. However, it is noteworthy that some species, such as Pinus halepensis Mill. and Betula ermanii Cham., are more rare.

In terms of species-specific characteristics, the dataset allows for an intriguing comparison of average physical attributes across different species. For instance, Abies alba Mill. shows an average age of 72 years, a DBH of around 20.44 cm, and an average height of approximately 18.3 meters. These averages vary distinctly across species, highlighting the unique growth patterns and physical characteristics inherent to each species. The variation in these averages, such as the notably high average age of 140 years for Abies holophylla Maxim. or the substantial DBH of Abies nephrolepis (Trautv.) Maxim., reflects the ecological adaptations and growth conditions specific to each species.

Table 3. Species Analysis: Frequency and Average Characteristics

Species	Frequency	Average Tree Age	Average DBH	Average Height
Pinus sylvestris L.	3995	49.61	14.33	12.72
Betula alba L.	1195	44.24	13.71	13.85
Picea abies (L.) Karst.	901	57.17	17.35	16.03
Populus tremula L.	498	32.84	14.08	13.63
Pinus Pallasiana	434	30.85	11.28	7.66
Tilia cordata Mill.	385	49.55	16.33	15.55
Picea obovata L.	369	32.15	7.01	6.07
Abies alba Mill. Mill.	206	40.24	17.08	14.34
Pinus sibirica Du Tour.	161	39.81	8.95	6.62
Larix cajanderi Mayr.	153	129.28	11.05	9.05

3.5. Physical Characteristics Analysis

The analysis of physical characteristics, specifically focusing on size distribution and age distribution, provides critical insights into the forest's dynamics and structure.

The distribution of tree sizes, as characterized by Diameter at Breast Height (DBH) and height, is a fundamental aspect of forest ecology. The DBH distribution generally reveals the range and commonness of tree sizes within the forest. A right-skewed DBH distribution, commonly observed in natural forests, indicates a larger number of smaller, younger trees compared to fewer, larger, older trees. This pattern is crucial for understanding the growth stages and dynamics of the forest. Similarly, the height distribution of trees, which often correlates with DBH, provides insights into the vertical structure of the forest. The variation in height can be influenced by species diversity, site conditions, and forest management practices.

The age distribution of trees is equally significant as it reflects the forest's age dynamics. It can indicate the regeneration capacity of the forest and its susceptibility to disturbances. A forest with a wide range of tree ages, including a significant proportion of young trees, suggests ongoing regeneration and a healthy age structure. Conversely, a forest dominated by trees of a similar age, especially if older, might be at risk of decline or more susceptible to disturbances like pests, diseases, or extreme weather events.

Combining the insights from size and age distributions allows for a comprehensive understanding of the forest's overall health and sustainability. It enables the identification of potential issues like over-maturity, lack of regeneration, or uneven age structures that could impact the forest's long-term viability. Moreover, these distributions are crucial for planning sustainable forest management practices, including selective logging, regeneration efforts, and conservation strategies.

3.6. Biomass Analysis

The biomass analysis, focusing on component-wise biomass and total biomass estimation, reveals several important aspects of the forest's ecology and the trees within it. The dataset includes various biomass components such as stem (Pst), bark (Pbark), branches (Pbr), foliage (Pf), and roots (Proot). The descriptive statistics (Table 4) for each of these components provide insights into how biomass is distributed across different parts of a tree.

- Stem Biomass (Pst): This component typically constitutes a significant portion of a tree's total biomass. The analysis shows a wide range in stem biomass, with some trees having very low values and others much higher. This variation could be due to differences in species, age, or environmental conditions.
- Bark Biomass (Pbark): Bark biomass, while generally smaller than stem biomass, also shows considerable variability. This could be influenced by species-specific characteristics and the age of the trees.
- Branch Biomass (Pbr) and Foliage Biomass (Pf): These components are usually smaller compared to stem and bark but are crucial for understanding the overall biomass distribution and the tree's canopy structure.
- Root Biomass (Proot): Root biomass is a critical component, especially for understanding carbon sequestration and the tree's stability and health.

The total biomass (Ptot) of a tree is a sum of all these components. The dataset provides values for total biomass, but where it's missing, it can be estimated by summing the individual components. The analysis of total biomass shows the overall biomass capacity of the trees in the dataset. Similar to the individual components, total biomass also exhibits a wide range, reflecting the diversity in tree sizes, species, and ages. It helps in assessing the carbon sequestration potential of forests, understanding the ecological roles of different tree parts, and in forest management for timber, fuel, and other products. The variation in biomass distribution across different tree species and sizes also provides insights into the growth patterns and health of the forest ecosystem.

Table 4. Biomass Analysis of Trees

Statistic	Stem (Pst)	Bark (Pbark)	Branch (Pbr)	Foliage (Pf)	Root (Proot)	Total (Ptot)
Count	7466	4799	8862	8896	1746	1712
Mean	107.85	12.69	15.70	5.63	28.33	173.10
Std	227.57	20.50	39.90	13.03	71.50	401.87
Min	0.00	0.00	0.00	0.00	0.00	0.00
25%	4.97	1.20	0.98	0.59	0.72	3.75
Median	29.43	4.90	3.86	2.04	4.95	28.78
75%	120.85	15.65	13.94	5.70	24.43	164.53
Max	4122.00	280.00	1091.80	305.00	901.00	5134.80

3.7. Geographical Analysis

The geographical analysis of the dataset offers a deep dive into the spatial distribution of trees and the characteristics of different ecological regions (Ecoregions), see Figure 3 and Table 5. The dataset spans a broad geographic range, as evidenced by the latitude and longitude statistics. The mean values for latitude and longitude indicate the central tendency of the dataset's geographical coverage. The standard deviation provides a measure of the spread or dispersion of the tree locations, showing how widely the data is distributed geographically. The minimum and maximum values for latitude and longitude further delineate the geographical boundaries of the dataset.

Grouping the data based on Ecoregions allows for an analysis of region-specific characteristics. By examining the mean values of various attributes (like tree age, DBH, height, biomass components) for each Ecoregion, we can discern how these characteristics vary across different ecological zones. Ecoregions are typically defined based on similarities in the environment, climate, soil, and other ecological factors, which in turn influence the growth patterns and physical attributes of trees. For instance, trees in one Ecoregion might have a higher average age or larger DBH compared to those in another, reflecting differences in growth conditions, forest management practices, or species composition. Biomass component analysis across Ecoregions can reveal how biomass allocation varies in different ecological settings, which is vital for understanding carbon storage, habitat quality, and overall forest health in those regions.

Table 5. Mean Characteristics by Ecoregion

Ecoregion	Mean Tree Age	Mean DBH	Mean Height	Mean Total Biomass
80402	19.50	6.85	10.70	20.20
80404	43.49	19.94	10.55	-
80405	87.67	32.60	23.77	834.10
80408	25.00	5.15	5.50	-
80409	33.00	10.52	11.73	39.63

This map (Figure 3) is a valuable tool for visualizing the geographical spread and ecological diversity of the dataset. It helps in understanding the distribution of different Ecoregions and the trees associated with them, which is crucial for ecological studies and forest management planning. The geographical analysis of the dataset not only highlights the spatial distribution of the trees but also provides insights into the ecological diversity and region-specific characteristics within the data.

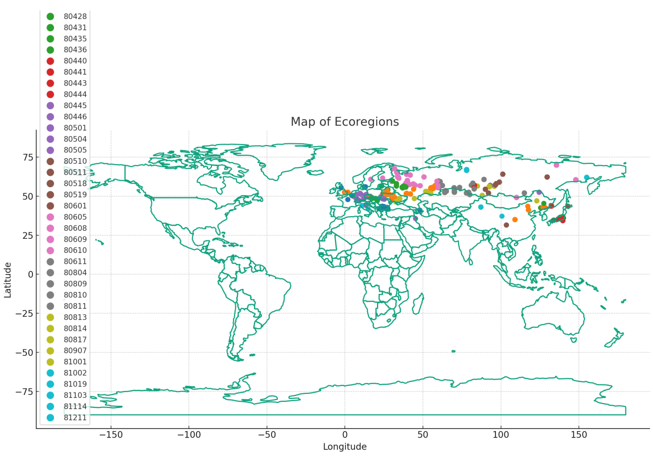

Figure 3. Map of ecoregions

3.8. Correlation Analysis

The heatmap analysis of the dataset's correlation matrix (Figure 4) reveals several intriguing insights into the relationships between various tree attributes. A notable observation is the strong positive correlations between tree physical characteristics such as Diameter at Breast Height (DBH), Tree Height (H tree), and Total Volume (Vtot). This relationship aligns with the basic understanding of tree growth, where larger trees in terms of diameter and height naturally exhibit greater volume. Similarly, there is a discernible positive correlation between these physical attributes and biomass components like stem biomass (Pst), bark biomass (Pbark), branch biomass (Pbr), and foliage biomass (Pf). This suggests that as trees grow larger, they accumulate more biomass across different components. The relationship between Tree Age and other characteristics like DBH, H tree, and Vtot is particularly interesting. While it is generally expected that older trees are larger and have more volume, the strength of this correlation can provide insights into how trees grow and accumulate biomass over time in different environmental conditions. The correlation between Tree Age and various biomass components can also be enlightening, revealing patterns in how biomass allocation in trees changes throughout their lifespan.

Heatmap analysis offers valuable information about how biomass is distributed within a tree. Strong correlations among different biomass components, such as between stem and branch biomass, can indicate common patterns in biomass distribution. On the other hand, any weak or negative correlations can also be revealing. For example, a weak correlation between Tree Age and DBH or H tree could suggest that factors other than age, such as environmental conditions or genetic traits, play a significant role in determining a tree's size.

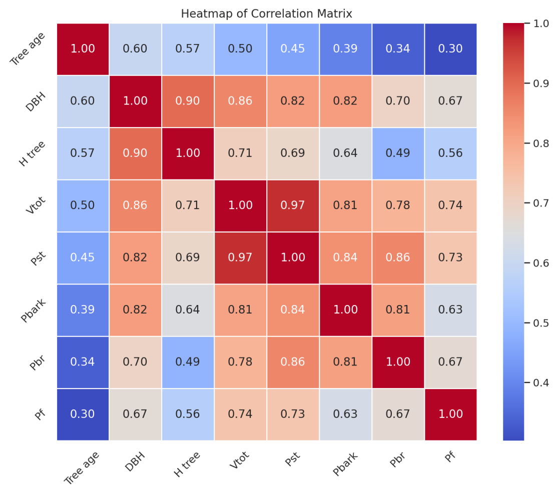

Figure 4. Correlation heatmap

3.9. FOM construction

The Forest Object Model (FOM) for the given dataset can be defined within the mathematical framework as follows:

Ecological Units $\mathcal{U}$: In this dataset, each tree or observation can be considered an ecological unit. Therefore, the set of ecological units $\mathcal{U}$ can be represented as $\mathcal{U} = U_1, U_2, \ldots, U_n$, where n is the total number of trees or observations in the dataset.

Ecological Attributes $\mathcal{E}$: The ecological attributes include characteristics such as species, age, DBH, height, and various biomass components. Therefore, $\mathcal{E}$ can be represented as $\mathcal{E} = E_1, E_2, \ldots, E_m$, where m is the total number of attributes recorded for each tree. Each attribute E_j is a tuple (N_j, T_j, V_j): N_j represents the name of the attribute (e.g., 'Species', 'DBH'). T_j defines the data type of the attribute (e.g., numeric for 'DBH', categorical for 'Species'). V_j represents the value associated with the attribute for a given tree.

Relationships $\mathcal{A}$: The relationships between ecological units and attributes are defined as a function $\mathcal{A} : \mathcal{U} \times \mathcal{E} \to \mathcal{V}$, where $\mathcal{V}$ is the set of values for the attributes. For instance, the relationship function maps a tree U_i to its DBH or species. Temporal Domain $\mathcal{D}$:

The temporal domain in this context represents the time intervals during which data were collected. If the dataset includes the year of data collection or tree age, this can be incorporated into $\mathcal{D}$ to model the temporal dynamics of the forest ecosystem. The FOM thus provides a comprehensive and structured representation of the forest data, capturing the complexities of ecological units, their attributes, the relationships among them, and the temporal aspects.

3.10. Construction of FIM

Table 6 demonstrates how different aspects of the dataset align with the functional components of FIM. This table organizes the components of the dataset according to the various layers of FIM. It shows how each aspect of the dataset aligns with the conceptual layers of Forest Information Modeling, providing a clear picture of the dataset's role in the broader context of forest ecosystem modeling and management.

3.11. Constructing a Functional Structural Forest Biomass Model

Constructing a Functional Structural Forest Biomass Model (FSFBM) using the dataset variables and their values involves creating a mathematical representation that captures the relationships between different forest attributes and biomass. This model integrates functional aspects (like growth processes) with structural characteristics (like size and shape) of the forest. Assuming the dataset includes variables such as tree age, Diameter at Breast Height (DBH), tree height (H), and various biomass components (stem, bark, branches, foliage, roots), the FSFBM can be formulated as follows. The FSFBM integrates the growth dynamics (modeled as a function of tree age and other ecological factors) with the structural attributes of trees (like DBH and height) to estimate the biomass distribution in different components of the forest ecosystem. This model allows for an understanding of how biomass is allocated in a forest, which is crucial for ecological studies, carbon sequestration analysis, and forest management. This mathematical formulation captures the essence of the Functional Structural Forest Biomass Model, providing a structured approach to understanding the complex relationships within forest ecosystems:

Table 6. Mapping of Dataset to FIM Layers

FIM Layer	Dataset Components
Ecological Data Layer	**Biophysical Data:** Data on tree physical characteristics (DBH, height, etc.). **Biodiversity Data:** Information on tree species composition. **Environmental Data:** Not explicitly covered unless environmental factors like temperature or soil type are included in the dataset.
Spatial Representation Layer	**Geospatial Information:** Geographic locations (latitude, longitude) of trees. **3D Modeling:** Not directly applicable unless the dataset includes data for 3D spatial structure.
Ecological Process Modeling Layer	**Dynamic Models:** Data can be used to build models for forest growth, but the dataset itself does not include dynamic models. **Temporal Dynamics:** Tree age data may contribute to understanding temporal aspects.
Decision Support Layer	**Scenario Analysis and Risk Assessment:** The dataset provides foundational data for such analyses but does not include them directly.
Stakeholder Engagement Layer	**Collaborative Platforms and Public Outreach:** Not covered by the dataset; requires additional tools and platforms.
Monitoring and Feedback Layer	**Sensor Networks and Data Integration:** The dataset could be a part of such systems but does not encompass them directly.

Let B_{stem}, B_{bark}, B_{branches}, B_{foliage}, B_{roots} be the biomass of stem, bark, branches, foliage, and roots respectively.

$$B_{\text{component}} = f_{\text{component}}(\text{DBH}, H) \tag{5}$$

where $f_{\text{component}}$ is a function of biomass for each component based on DBH and height.

$$B_{\text{total}} = B_{\text{stem}} + B_{\text{bark}} + B_{\text{branches}} + B_{\text{foliage}} + B_{\text{roots}} \tag{6}$$

Total biomass is the sum of all biomass components.

$$B_{\text{component}}(t) = g_{\text{component}}(\text{Age}, \text{DBH}(t), H(t)) \tag{7}$$

where $g_{\text{component}}$ models growth dynamics as a function of age, DBH, and height over time.

The Linear Regression model has been successfully applied to the dataset, and we have evaluated its performance. The R^2 score of 0.781 indicates that the model explains approximately 78.1% of the variance in the biomass data, which is a relatively strong fit, While the value of Root Mean Squared Error (RMSE) was 0.448. These results suggest that the model is quite effective in predicting biomass, although there's always room for improvement. Figure 5 visualizes the coefficient values of the selected features from the Linear Regression model. Each bar represents the magnitude and direction (positive or negative) of the feature's impact on the model's predictions. This analysis helps in understanding which of the specified features are most influential in predicting biomass.

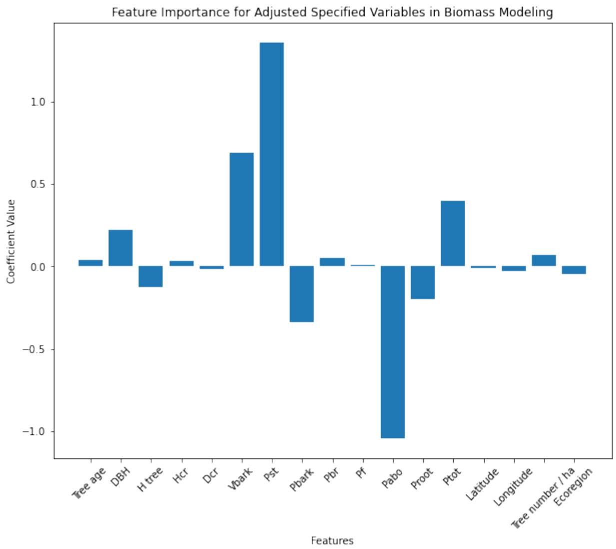

Figure 5. Coefficients of the Linear Regression model of Forest Biomass

4. Discussion and Conclusions

In both forestry management and environmental conservation, the integration of trees into digital systems for efficient management and sustainability has become increasingly important. Trees, differing from non-living elements of the environment, exhibit dynamic growth and require continual care and monitoring. Therefore, effective management of forest resources and their multifunctional uses is crucial.

In reference to BIM, we proposed FIM (Forest Information Modeling), serving as a bridge between forest ecosystems in reality and digital applications across multiple professions, including forestry science, ecology, environmental management, and urban planning. A comprehensive FIM model includes information tags and geometric representations for individual trees, understory vegetation, and other forest elements. This paper describes various data collection methods, such as remote sensing, LiDAR scanning, and ground-based surveys. It also discusses how topological geometry in the FIM model can be utilized for simulating forest dynamics and ecosystem processes.

FIM brings together diverse knowledge and data, facilitating accurate evaluation of forest conditions and management strategies. For forest managers and conservationists, FIM can predict resource needs and ecological benefits over the lifecycle of the forest. For environmental scientists and urban planners, FIM provides insights into spatial requirements for forest conservation and urban green space planning. Arborists and ecologists can assess risks related to tree health and forest resilience. In urban forestry studies, FIM's accurate geometric data can be used to evaluate the impact of trees on microclimates, including cooling and carbon sequestration. Landscape architects and environmental planners can use FIM for species selection, habitat design, and assessing the integration of natural and built environments.

Despite these advantages, FIM also faces challenges similar to those encountered with BIM. Firstly, adopting FIM entails higher learning and training costs for stakeholders. A forest conservation organization, for example, may need to invest in adapting their data management systems to align with FIM standards. Secondly, the standardization of forest data in FIM creates barriers to access and interpretation, especially for the general public. Lastly, legal frameworks and contractual arrangements concerning the ownership, access, and use of forest data in FIM are still underdeveloped.

Acknowledgement

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.

Data availability

This study uses the PANGEA dataset [33].

Conflicts of interest

The authors declare no conflict of interest.

References

[1] Feng, Y. and Audy, J. (2020) Forestry 4.0: a framework for the forest supply chain toward industry 4.0. *Gestão & Produção* **27**(4).

[2] Sahal, R., Alsamhi, S., Breslin, J. and Ali, M. (2021) Industry 4.0 towards forestry 4.0: Fire detection use case †. *Sensors (Basel, Switzerland)* **21**(3).

[3] Okewu, E., Misra, S., Maskeliunas, R., Damaşeviçius, R. and Fernandez-Sanz, L. (2017) Optimizing green computing awareness for environmental sustainability and economic security as a stochastic optimization problem. *Sustainability (Switzerland)* **9**(10).

[4] Torresan, C., Benito Garzón, M., O'Grady, M., Robson, T.M., Picchi, G., Panzacchi, P., Tomelleri, E. *et al.* (2021) A new generation of sensors and monitoring tools to support climate-smart forestry practices. *Canadian Journal of Forest Research* **51**(12): 1751–1765.

[5] Reischl, D.A. and Uhl, E. (2021) Carbon smart forestry under climate change. *The Project Repository Journal* **10**(1): 96–99.

[6] Nitoslawski, S.A., Galle, N.J., van den Bosch, C.K. and Steenberg, J. (2019) Smarter ecosystems for smarter cities? a review of trends, technologies, and turning points for smart urban forestry. *Sustainable Cities and Society* **51**.

[7] Hallberg-Sramek, I., Reimerson, E., Priebe, J., Nordström, E.M., Mårald, E., Sandström, C. and Nordin, A. (2022) Bringing "climate-smart forestry" down to the local level—identifying barriers, pathways and indicators for its implementation in practice. *Forests* **13**: 98.

[8] Bowditch, E. *et al.* (2020) What is climate-smart forestry? a definition from a multinational collaborative process focused on mountain regions of europe. *Ecosystem Services* .

[9] Santopuoli, G., Temperli, C., Alberdi, I., Barbeito, I., Bosela, M., Bottero, A., Klopčič, M. *et al.* (2021) Pan-european sustainable forest management indicators for assessing climate-smart forestry in europe. *Canadian Journal of Forest Research* **51**(12): 1741–1750.

[10] Nitschke, C.R. and Innes, J.L. (2008) Integrating climate change into forest management in south-central british columbia: An assessment of landscape vulnerability and development of a climate-smart framework. *Forest Ecology and Management* **256**(3): 313–327.

[11] Mondejar, M.E., Avtar, R., Diaz, H.L.B., Dubey, R.K., Esteban, J., Gómez-Morales, A., Hallam, B. *et al.* (2021) Digitalization to achieve sustainable development goals: Steps towards a smart green planet. *Science of The Total Environment* **794**: 148539.

[12] Torresan, C., Benito Garzón, M., O'grady, M., Robson, T.M., Picchi, G., Panzacchi, P., Tomelleri, E. *et al.* (2021) A new generation of sensors and monitoring tools to support climate-smart forestry practices. *Canadian Journal of Forest Research* **51**(12): 1751–1765.

[13] Lechner, A.M., Foody, G.M. and Boyd, D.S. (2020) Applications in remote sensing to forest ecology and management. *One Earth* **2**(5): 405–412.

[14] Tang, L. and Shao, G. (2015) Drone remote sensing for forestry research and practices. *Journal of forestry research* **26**: 791–797.

[15] Sonti, S. (2015) Application of geographic information system (gis) in forest management. *Journal of Geography & Natural Disasters* **5**(3): 1000145.

[16] Mekonnen, M., Sewunet, T., Gebeyehu, M., Azene, B. and Melesse, A.M. (2016) Gis and remote sensing-based forest resource assessment, quantification, and mapping in amhara region, ethiopia. *Landscape dynamics, soils and hydrological processes in varied climates* : 9–29.

[17] Gabrys, J. (2020) Smart forests and data practices: From the internet of trees to planetary governance. *Big data & society* **7**(1): 2053951720904871.

[18] Sellén, D. (2016), Big data analytics for the forest industry: A proof-of-concept built on cloud technologies, Master Thesis, Mid Sweden University.

[19] Nikolakis, W., John, L. and Krishnan, H. (2018) How blockchain can shape sustainable global value chains: an evidence, verifiability, and enforceability (eve) framework. *Sustainability* **10**(11): 3926.

[20] Picchio, R., Proto, A.R., Civitarese, V., Di Marzio, N. and Latterini, F. (2019) Recent contributions of some fields of the electronics in development of forest operations technologies. *Electronics* **8**(12): 1465.

[21] Lechner, A.M., Foody, G.M. and Boyd, D.S. (2020) Applications in remote sensing to forest ecology and management. *One Earth* **2**(5): 405–412.

[22] Kanga, S. (2023) Advancements in remote sensing tools for forestry analysis. *Sustainable Forestry* **6**(1).

[23] Li, X., Chen, W.Y., Sanesi, G. and Lafortezza, R. (2019) Remote sensing in urban forestry: Recent applications and future directions. *Remote. Sens.* **11**: 1144.

[24] Talbot, B., Pierzchala, M. and Astrup, R. (2017) Applications of remote and proximal sensing for improved precision in forest operations. *Croatian Journal of Forest Engineering : Journal for Theory and Application of Forestry Engineering* **38**: 327–336.

[25] Blanco, J.A., Ameztegui, A. and Rodríguez, F. (2020) Modelling forest ecosystems: a crossroad between scales, techniques and applications. *Ecological Modelling* **425**: 109030.

[26] Borrmann, A., König, M., Koch, C. and Beetz, J. (2018), Building information modeling: Why? what? how?, Building Information Modeling.

[27] Kant, S. (2003) Extending the boundaries of forest economics. *Forest Policy and Economics* **5**: 39–56.

[28] Mendoza, G. and Vanclay, J. (2008) Trends in forestry modelling. *Cab Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources* **3**.

[29] Korzukhin, M., Ter-Mikaelian, M. and Wagner, R. (1996) Process versus empirical models: which approach for forest ecosystem management? *Canadian Journal of Forest Research* **26**: 879–887.

[30] Albers, H., Fisher, A. and Hanemann, W. (1996) Valuation and management of tropical forests. *Environmental and Resource Economics* **8**: 39–61.

[31] Chen, J. *et al.* (2020) Using machine learning to synthesize spatiotemporal data for modelling dbh-height and dbh-height-age relationships in boreal forests. *Forest Ecology and Management* **466**: 118104.

[32] Ansary, M. *et al.* (2020) Technocenosis paradigm sustainability management of a forest enterprise. *IOP Conference Series: Earth and Environmental Science* **595**.

[33] Schepaschenko, D., Shvidenko, A., Usoltsev, V.A., Lakyda, P., Luo, Y., Vasylyshyn, R., Lakyda, I. *et al.* (2017), Biomass tree data base, PANGAEA.

[34] Schepaschenko, D., Shvidenko, A., Usoltsev, V., Lakyda, P., Luo, Y., Vasylyshyn, R., Lakyda, I. *et al.* (2017) A dataset of forest biomass structure for eurasia. *Scientific Data* **4**(1).