



Data assimilation routines and processing tools

D3.2: Data assimilation routines and processing tools
WP3, T 3.2

AUTHORS: Niamh E. Kelly, Johanna Klapper, Giovanni Santopuoli, Serena Antonucci, Concetta Lisella, Cesar Alvites, Diana Alfieri, Thomas A. M. Pugh, Mart-Jan Schelhaas, Haoming Zhong, Annemarie Eckes-Shephard, Gert-Jan Nabuurs, Pieter Johannes Verkerk.

Data supplied by: Sander Teeuwen, Probos - SYHI data, and Jonas Fridman, SLU - Swedish NFI data.

Technical references

Project Acronym	FORWARDS
Project Title	The ForestWard Observatory to Secure Resilience of European Forests
Project Coordinator	Professor Ruben Valbuena Sveriges Lantbruksuniversitet (SLU) Ruben.valbuena@slu.se
Project Duration	November 2022 – October 2027 (60 months)

Funded by the
European Union



Deliverable No.	D3.2
Dissemination level*	Public
Work Package	WP 3 – Restoration and Climate-Smart Forest management practices
Task	T3.2 – Data processing and forward-looking projections
Lead beneficiary	WR
Contributing beneficiary/ies	EFI, LUND, UNIMOL
Due date of deliverable	31 Oct 2025
Actual submission date	31 Oct 2025

PU = Public

PP = Restricted to other programme participants (including the Commission Services)

RE = Restricted to a group specified by the consortium (including the Commission Services)

CO = Confidential, only for members of the consortium (including the Commission Services)

<i>v</i>	<i>Date</i>	<i>Beneficiary</i>	<i>Authors</i>
1.0	01/10/2025	WR, EFI, LUND and UNIMOL	Niamh Kelly, Johanna Klapper, Giovanni Santopuoli, Serena Antonucci, Concetta Lisella, Cesar Alvites, Diana Alfieri, Thomas A. M. Pugh, Mart-Jan Schelhaas, Pieter Johannes Verkerk, Haoming Zhong, Annemarie Eckes-Shephard, Gert-Jan Nabuurs.
Review	20/10/2025	UNIFI	Giovanni d'Amico
2.0	31/10/2025	WR, EFI, LUND and UNIMOL	Niamh E. Kelly, Johanna Klapper, Giovanni Santopuoli, Thomas A. M. Pugh, Mart-Jan Schelhaas, Haoming Zhong, Annemarie Eckes-Shephard, Gert-Jan Nabuurs, Pieter Johannes Verkerk.

Table of contents

Technical references	1
List of Tables	5
List of Figures	6
Abstract	10
1. Introduction	11
1.1. Background	11
1.2. Report structure	11
1.3. References	13
2. Data Processing Tools	14
2.1. Introduction	14
2.2. Climate-Smart Forestry Indicators Generator (CSF-Ind)	15
2.2.1. CSF-Ind data input	17
2.2.2. CSF-Ind data output	20
2.2.3. Supporting information of CSF-Ind app	23
2.3. Climate-Smart Forestry Assessment Tool (CSF-AT App)	24
2.3.1. CSF-AT app data input	24
2.3.2. CSF-AT app data output	28
2.4. Connecting the CSF-Ind web-app to CSF-AT web-app	29
2.5. Case study demonstrations	30
2.5.1. Practical application of the CSF-Ind: Bosco Pennataro case study	30
2.5.2. Practical application of the CSF-AT: the Collemeluccio case study	38
2.6. References	45
3. Data assimilation routines for spatially explicit data on forest structure and composition to enable forward-looking projections	47

3.1.	Introduction.....	47
3.1.1.	Motivation.....	47
3.1.2.	Synthetic data concept.....	48
3.1.3.	Overall approach	49
3.2.	Case study demonstrations	54
3.2.1.	Case study 1: Netherlands.....	54
3.2.2.	Case study 2: Finland.....	69
3.2.3.	Case study 3: Italy.....	86
3.2.4.	Discussion and Conclusions of the case-studies	98
3.3.	Other developments in EFISCEN-Space	101
3.4.	References.....	102

4. Improved inputs for process-based projections of future climate change impacts on forests 105

4.1.	Introduction.....	105
4.2.	Description of LPJ-GUESS relevant to the implementation.....	106
4.3.	Implementing satellite-driven disturbances in LPJ-GUESS.....	107
4.4.	Example outputs	108
4.5.	Concluding remarks	109
4.6.	References.....	110

List of Tables

Table 1 CALCULABLE INDICATORS WITH CSF-IND WEB-APP. “*” INDICATES THAT OUTPUT DATA OBTAINED FOR GROWING STOK CALCULATION, CAN BE USED AS INPUT DATA FOR ASSESSING CARBON STOCK. SOME OUTPUT CAN BE USED AS INPUT FOR CSF-AT WEB-APP.	21
Table 2 CSF INDICATORS FOR CSF-AT. THE TABLE REPORTS ABBREVIATION, GENERAL DEFINITION, POSSIBLE CALCULATION METHODS, AND MEASUREMENT UNITS. GS, CS, DB, TS, AND SLC ARE MANDATORY INDICATORS, WHILE DW, RN, FD, IF, AND RW ARE OPTIONAL.	25
Table 3 VALUES OF CSF COMPOSITE INDEX AND ITS THREE MAIN PILLARS (MITIGATION, ADAPTATION AND SOCIAL DIMENSION) FOR COLLEMELUCCIO EXPERIMENTAL SITES AND REGIONAL SILVER FIR FORESTS.	42
Table 4: CASE STUDY COUNTRIES AND THE RESPECTIVE REASONS FOR INCLUSION.....	50
Table 5 MATRIX OF DATA SOURCES USED IN THIS STUDY TO TEST AND CREATE SETS OF SYNTHETIC TREE-LEVEL INFORMATION, WITH CASE STUDIES AT THEIR INTERSECTIONS. THE FIRST COLUMN SHOWS SOURCES OF PLOT-LEVEL DATA, THE FIRST ROW SOURCES OF TREE-LEVEL DATA. THE NUMBER IN PARENTHESES REFERS TO THE RESPECTIVE EXPERIMENT WITHIN A CASE STUDY. IF MULTIPLE DATA SOURCES WERE USED WITHIN THE SAME EXPERIMENT, THE NUMBER CAN OCCUR MORE THAN ONCE IN THIS TABLE (E.G., FINLAND (FI-1)).	50

List of Figures

Figure 1 FRONTEND AND BACKEND COMPONENTS OF THE WEB-BASED APPLICATION ARCHITECTURE.	15
Figure 2 WORKFLOW OF CSF-IND APP.	16
Figure 3 CSF-IND WEB-APP USERS INTERFACE: (I) HEADER/NAVIGATION; (II) SIDEBAR CONTROLS; (III) MAIN CONTENT AREA.	17
Figure 4 EXAMPLE OF INPUT DATA AS .XLSX FILE FOR COMPUTING THE GROWING STOCK.....	18
Figure 5 INTERFACE OF CSF-IND WEB-APP SHOWING THE INTERACTIVE SIDEBAR CONTROL TO SELECT CSF PILLARS AND THE RELATIVE INDICATORS THAT USERS WANT TO EVALUATE.....	18
Figure 6 EXAMPLE OF OUTPUT FOR GROWING STOCK INDICATOR: I) THE RESULTS AT PLOT-LEVEL IN TABLE FORMAT, II) THE BAR-CHART REPORTS THE VALUE OF THE INDICATOR PER HECTARE COMPUTED ACROSS THE PLOTS PRESENT IN THE DATASET. THE ARROW INDICATES THE OUTPUT IN THE TABLE WHEN FORETS MANAGEMENT INTENSITY DATA IS NOT AVAILABLE.....	22
Figure 7 EXAMPLE OF CSF-AT APP INTERFACE.	24
Figure 8 EXAMPLE OF OUTPUT GENERATED BY THE APPLICATION WITH THE VALUE OF CSF AT PLOT AND STAND LEVEL.....	26
Figure 9 FRAMEWORK TO ASSESS CLIMATE-SMART FORESTRY. ON THE LEFT, THE STEPWISE PROCEDURE USED TO CALCULATE THE COMPOSITE INDEX; ON THE RIGHT, THE HIERARCHICAL STRUCTURE OF THE COMPOSITE INDEX.....	27
Figure 10 WORKFLOW OF CSF-IND AND CSF-AT WEB-APPS (ABOVE). BELOW ARE REPORTED VARIABLES GENERATED PER EACH INDICATOR BY THE CSF-IND WEB-APP THAT CAN BE USED IN THE CSF-AT WEB-APP.....	29
Figure 11 LOCATION OF BOSCO PENNATARO IN MOLISE REGION. THE RED DOTS INDICATE ALL THE PLOTS ESTABLISHED IN THE STUDY AREA WHICH WERE USED FOR THE CASE STUDY.	30
Figure 12 MAIN OUTPUT FROM CSF-IND. TABLES REPORT THE VOLUME PER MANAGEMENT INTENSITY AND FOR DOMINANT TREE SPECIES, WHILE PIE CHARTS SHOW THE PERCENTAGE OF MANAGEMENT INTENSITY CATEGORIES AND THE TREE DOMINANT SPECIES.	32
Figure 13 EXAMPLE OF THE OUTPUT FOR THE INDICATOR “GROWING STOCK”	33
Figure 14 TABLES AND BAR-CHART EXTRACTED FROM THE CSF-IND PDF REPORT FOR THE INDICATOR “CARBON STOCK”.	34
Figure 15 OUTPUT FOR INDICATOR “DEADWOOD”. TABLES SHOW DEADWOOD VOLUME, WHILE PIE CHARTS SHOW THE PERCENTAGE OF DEADWOOD COMPONENTS IN THE STUDY AREA, AND THE MANAGEMENT INTENSITY.....	36

- Figure 16 TABLE AND BAR-CHART EXTRACTED FROM THE PDF REPORT FOR FOREST DIVERSITY. THE TABLE SHOWS ONLY THE FIRST 6 COLUMNS OF THE OUTPUT BUT ALL INDICES COMPUTED BY THE WEB-APP ARE REPORTED IN THE .CSV FILES..... 37
- Figure 17 STUDY AREA. THE MAP ON THE LEFT DISPLAYS THE BOUNDARIES OF THE COLLEMELUCCIO FOREST AND THE TWO FIELD PLOTS. THE MAP IN THE CENTER SHOWS THE LOCATION OF THE STUDY AREA WITH ADMINISTRATIVE BOUNDARIES OF THE MUNICIPALITIES IN THE MOLISE REGION. THE MAB RESERVE IS HIGHLIGHTED IN LIGHT GRAY, WHILE THE RED POINT INDICATES THE STUDY AREA. THE BOX ON THE RIGHT SHOWS THE SAMPLING DESIGN OF THE FIELD SURVEY. 40
- Figure 18 TERNARY PLOT SHOWING THE RELATIVE CONTRIBUTIONS OF MITIGATION, ADAPTATION, AND SOCIAL DIMENSION TO THE CSF INDEX FOR THE COLLEMELUCCIO EXPERIMENTAL SITES (MANAGED AND UNMANAGED) AND FOR REGIONAL SILVER FIR FORESTS (ABRUZZO, BASILICATA, CALABRIA)..... 43
- Figure 19 FOREST INVENTORY PLOTS ARE PRESENTED TO EFISCEN-SPACE AS PER-SPECIES DIAMETER DISTRIBUTIONS OF THE STEM NUMBER PER HECTARE. THE EXAMPLE SHOWS THE TREEWISE DISTRIBUTION OF AN NFI PLOT IN THE NETHERLANDS..... 47
- Figure 20 CONCEPTUAL DIAGRAM SHOWING HOW A SYNTHETIC INVENTORY PLOT IS CREATED. PLOT-LEVEL DATA (E.G. SUMMARISED INVENTORY DATA OR PIXEL FROM A REMOTE SENSING-BASED MAP), WHICH ON ITS OWN CANNOT BE USED TO INITIALISE EFISCEN-SPACE IS MATCHED TO TREE-LEVEL DATA FROM A TREE-WISE FOREST INVENTORY (E.G. NFI OR MANAGEMENT INVENTORY). A MATCH IS MADE BASED ON SIMILARITY BETWEEN SUMMARY STATISTICS OF A SELECTION OF PLOT CHARACTERISTICS..... 49
- Figure 21 MAP OF THE NETHERLANDS, WITH THE CASE-STUDY AREA HIGHLIGHTED IN YELLOW. TWO SOURCES OF FIELD-INVENTORY DATA THAT WERE USED ARE ALSO DISPLAYED..... 55
- Figure 22 COMPARISON OF THE DATA SETS USED FOR THE DUTCH CASE STUDY. A) BASAL AREA, B) STEM VOLUME, C) DOMINANT SPECIES GROUP. X-AXES HAVE BEEN TRUNCATED FOR THE VISUALISATION. 57
- Figure 23 SYNTHETIC PLOTS DERIVED FROM TREE-LEVEL NFI-7 DATA PLOTTED AGAINST THE ORIGINAL SYHI DATA FOR BASAL AREA, QMD, STEM DENSITY. A)THE SYNTHETIC DATA WERE ALLOCATED USING BASAL AREA AND QMD; B) USING VOLUME AND QMD. 59
- Figure 24 SPECIES DISTRIBUTION OF THE SYNTHETIC DATA DERIVED FROM NFI-7 DATA COMPARED TO THE ORIGINAL SYHI PLOTS. THE SYNTHETIC DATA WERE ALLOCATED USING THE DOMINANT SPECIES GROUP. 59
- Figure 25 SYNTHETIC PLOTS DERIVED FROM TREE-LEVEL NFI-7 DATA AND NATIONAL STRUCTURE MAPS, USING THE OPTIMAL SETTINGS FROM EXPERIMENT 1; PLOTTED AGAINST THE ORIGINAL SYHI DATA FOR BASAL AREA, QMD, AND STEM DENSITY. 61
- Figure 26 DENSITY CURVE OF BASAL AREA FOR SYNTHETIC AND REFERENCE DATA. SYNTHETIC DATA ARE BASED ON NFI-7 DATA AND NATIONAL STRUCTURE MAPS. THE DISTRIBUTION OF THE SYNTHETIC DATA IS NARROWER AND THE MEDIAN IS HIGHER..... 62

Figure 27 SPECIES DISTRIBUTION OF THE SYNTHETIC DATA COMPARED TO THE ORIGINAL SYHI PLOTS, WHEN USING THE SPECIES GROUPS FROM THE EUROPEAN GENUS MAP COMBINED WITH NFI-7 DATA TO CREATE SYNTHETIC PLOTS.	62
Figure 28 SYNTHETIC DATA DERIVED FROM TREE-LEVEL NFI-7 DATA AND EUROPEAN STRUCTURE MAPS, USING AGB AND THE DOMINANT SPECIES GROUP TO ALLOCATE PLOTS; PLOTTED AGAINST THE ORIGINAL SYHI DATA FOR BASAL AREA, QMD, AND STEM DENSITY.....	64
Figure 29 QMD DISTRIBUTION OF THE DIFFERENT INVENTORY SOURCES. VERTICAL DASHED LINES INDICATE THE MEDIAN VALUES.....	66
Figure 30 DOMINANT SPECIES GROUP DISTRIBUTION OF THE DIFFERENT INVENTORY SOURCES.	66
Figure 31 SYNTHETIC PLOTS DERIVED FROM GERMAN NFI-3 TREE-LEVEL DATA USING VOLUME AND QMD AS CONSTRAINING VARIABLES; PLOTTED AGAINST THE ORIGINAL SYHI DATA FOR BASAL AREA, QMD, AND STEM DENSITY.....	67
Figure 32 DISTRIBUTION OF FINNISH NFI12 PLOTS ACROSS FINLAND (CREATED IN QGIS).	70
Figure 33 DISTRIBUTION OF AVAILABLE METSÄKESKUS PLOTS ACROSS FINLAND (CREATED IN QGIS).	71
Figure 34 REGIONS IN FINLAND, AS USED FOR FOREST RESOURCE REPORTING (CREATED IN RSTUDIO).....	73
Figure 35: SCATTERPLOTS AND PIE CHART OF ORIGINAL AND SYNTHETIC DATA. ORIGINAL DATA = METSÄKESKUS DATA, SYNTHETIC DATA = METSÄKESKUS NFI12 DATA. VARIABLES ARE A) MEAN DBH, B) BASAL AREA, C) STEM DENSITY AND D) DOMINANT SPECIES.....	75
Figure 36: SCATTERPLOTS AND PIE CHART OF ORIGINAL AND SYNTHETIC DATA. ORIGINAL DATA = METSÄKESKUS DATA, SYNTHETIC DATA = METSÄKESKUS MS-NFI DATA. VARIABLES ARE A) MEAN DBH, B) BASAL AREA, C) STEM DENSITY AND D) DOMINANT SPECIES.....	76
Figure 37: REGIONAL COMPARISON OF VOLUME, X-AXIS SHOWS THE OFFICIAL DATA FROM NATURAL RESOURCES INSTITUTE (LUKE). A) ORIGINAL METSÄKESKUS DATA, B) SYNTHETIC DATA WITH NFI12 AND C) SYNTHETIC DATA WITH MS-NFI.....	77
Figure 38: DENSITY PLOTS, SCATTER PLOT AND PIE CHART OF ORIGINAL AND SYNTHETIC DATA. ORIGINAL DATA = METSÄKESKUS, SYNTHETIC DATA = METSÄKESKUS EUROPEAN MAPS. VARIABLES ARE A) MEAN DBH, B) BASAL AREA, C) STEM DENSITY AND D) DOMINANT SPECIES.....	80
Figure 39: VOLUME COMPARISON - REGIONAL DATA AND SYNTHETIC DATA WITH EUROPEAN MAPS.....	81
Figure 40: SCATTERPLOTS, DENSITY PLOT AND PIE CHART OF ORIGINAL AND SYNTHETIC DATA. ORIGINAL DATA = METSÄKESKUS DATA, SYNTHETIC DATA = SWE-NFI NFI12. VARIABLES ARE A) MEAN DBH, B) BASAL AREA, C) STEM DENSITY AND D) DOMINANT SPECIES.	83
Figure 41: SCATTERPLOT, DENSITY PLOTS AND PIE CHART OF ORIGINAL DATA AND SYNTHETIC DATA. ORIGINAL DATA = METSÄKESKUS DATA, SYNTHETIC DATA = SWE-NFI MS-NFI. VARIABLES ARE A) MEAN DBH, B) BASAL AREA, C) STEM DENSITY AND D) DOMINANT SPECIES.....	84

Figure 42: PLOT LOCATIONS OF INFC2015 AS USED IN THE CASE STUDY.	87
Figure 43: SCATTERPLOT, DENSITY PLOTS AND PIE CHART OF ORIGINAL AND SYNTHETIC DATA. ORIGINAL DATA = INFC2015, SYNTHETIC DATA = EUROPEAN MAPS. VARIABLES ARE A) MEAN DBH, B) BASAL AREA, C) STEM DENSITY AND D) DOMINANT SPECIES. DOMINANT SPECIES ARE SUMMARIZED IN 'OTHER SPECIES' IF NUMBER OF OCCURRENCES IS LOW.	89
Figure 44: DENSITY PLOTS AND PIE CHART OF ORIGINAL AND SYNTHETIC DATA. ORIGINAL DATA = INFC2015, SYNTHETIC DATA = NATIONAL/EUROPEAN MAPS. VARIABLES ARE A) MEAN DBH, B) BASAL AREA, C) STEM DENSITY AND D) DOMINANT SPECIES. DOMINANT SPECIES ARE SUMMARIZED IN 'OTHER SPECIES' IF NUMBER OF OCCURRENCES IS LOW.	91
Figure 45: DENSITY PLOTS AND PIE CHART OF ORIGINAL AND SYNTHETIC DATA. ORIGINAL DATA = INFC2015, SYNTHETIC DATA: LEFT COLUMN = EUROPEAN MAPS + TOPOGRAPHIC, RIGHT COLUMN = NATIONAL/EUROPEAN MAPS + TOPOGRAPHIC. VARIABLES ARE A) MEAN DBH, B) BASAL AREA, C) STEM DENSITY AND D) DOMINANT SPECIES. DOMINANT SPECIES ARE SUMMARIZED IN 'OTHER SPECIES' IF NUMBER OF OCCURRENCES IS LOW.	94
Figure 46: PROJECTED OVERALL AVERAGE VOLUME (LEFT) AND AVERAGE VOLUME DEVELOPMENT PER SPECIES GROUP (RIGHT). LINES SHOW RESULTS OF EFISCEN-SPACE RUNS WITH NFI DATA (BASELINE) AND RUNS WITH THE DIFFERENT SYNTHETIC DATASETS. THE GRAPHS TO THE RIGHT DEPICT THE CONTRIBUTION OF EVERY SPECIES TO THE NATIONAL AVERAGE VOLUME. DATASET DESCRIPTIONS IN THE LEGEND REFER TO THE SYNTHETIC DATASETS FROM PREVIOUS EXPERIMENTS AS FOLLOWS: EUROPEAN DATA = IT-1, EUROPEAN DATA + TOPOGRAPHIC = IT-3, NATIONAL/EUROPEAN DATA = IT-2, NATIONAL/EUROPEAN DATA + TOPOGRAPHIC = IT-3.	95
Figure 47: TREE AGE STRUCTURE IN 2010 AND 2023 ACROSS 12 EUROPEAN COUNTRIES BASED ON LPJ-GUESS OUTPUT. NOTE THE LOG SCALE ON THE Y AXIS.	108
Figure 48: FRACTION OF TOTAL CARBON TURNOVER DUE TO NATURAL DISTURBANCES (A) AND HARVEST (B) AVERAGED OVER THE PERIOD 2011-2023, AS ESTIMATED BY LPJ-GUESS FORCED BY SATELLITE AND NFI OBSERVATIONS. HARVEST INCLUDES BOTH CLEARCUT HARVEST AND LESS INTENSIVE EVENTS. PIXEL SIZE IS SCALED BY FOREST AREA.	109

Abstract

The FORWARDS project aims to develop, test and implement a European observatory that will provide important information on the vulnerability of European forests to climate change, as well as evidence-based guidance for Climate-Smart Forest (CSF) and restoration management. Within the FORWARDS project, WP3 aims to identify, quantify and evaluate options to guide forest restoration activities and CSF management across Europe. CSF represents an integrated approach to sustainable forest management that aims to achieve climate change mitigation and adaptation objectives, while simultaneously ensuring the sustainable provision of forest ecosystem services, and considering the social dimension. By optimizing forests' capacity to act as carbon sinks, providing renewable resources, and adapting to new climatic conditions, CSF can contribute significantly to global climate action and sustainable development goals. Our aim for this deliverable was to develop data assimilation routines and processing tools to enable the evaluation of restoration and CSF management options' effectiveness, based on observations and forward-looking projections to provide guidance on CFS and restoration options. For each developed tool or routine, case studies were reported and explained.

We present two user-friendly web-applications, namely CSF-Ind and CSF-AT, for quantifying a set of CSF indicators with a focus on forest management, from local to national level. The web-applications developed in this project represent a starting point for exploring the state and trends of forests and their management, providing outputs on CSF that are useful for forest managers and policy decision makers.

Furthermore, we present a data processing and assimilation routine for synthesising spatially explicit forest data to provide detailed information about the composition, and structure of forests for locations with incomplete data or insufficient data coverage, by combining field and remotely-sensed data products. These data could be used in forest projection models, such as EFISCEN-Space, to simulate forest development in upcoming decades for different management and climate scenarios to evaluate the effectiveness of CSF management practices, as well as the impact of climate change and associated disturbances.

Finally, we describe how data on forest structure can be improved as a basis for process--based -models, using LPJ-GUESS as an example. We focus on improving the timeliness of data used for such projections by leveraging on the latest remote-sensing based data on forest disturbances and harvest to estimate stand age. Moreover, we describe a new feature to directly force stand-replacing disturbances in LPJ-GUESS with observations of disturbance occurrence from satellite information.

1. Introduction

1.1. Background

The FORWARDS project aims to develop, test and implement a European observatory that will provide important information on the vulnerability of European forests to climate change, as well as evidence-based guidance for Climate-Smart Forest (CSF) and restoration management. Within the FORWARDS project, WP3 aims to identify, quantify and evaluate options to guide forest restoration activities and CSF management across Europe.

CSF is a holistic approach to guiding forest management, building on the concepts of sustainable forest management (SFM) with a strong focus on climate and ecosystem services (ES) (Nabuurs et al. 2017, Jandl et al. 2018, Yousefpour et al. 2018; Bowditch et al. 2020, Verkerk et al. 2020; Alfieri et al., 2024;). CSF represents an integrated approach to SFM that aims to achieve climate change mitigation and adaptation objectives, simultaneously ensuring the sustainable provision of forest ES, and considering the social dimension (Weatherall et al., 2022; Verkerk et al. 2020). CSF shifts traditional forestry practices towards a more dynamic and responsive framework capable of addressing the multifaceted challenges posed by a changing climate. Key components of CSF include (i) enhancing carbon sequestration in forest biomass and soils, (ii) the use of wood products as substitutes for more carbon-intensive materials, (iii) fostering forest resilience to climate impacts (e.g., pests, diseases, extreme weather events), and (iv) supporting biodiversity conservation and restoration practices. By optimizing forests' capacity to act as carbon sinks, providing renewable resources, and adapting to new climatic conditions, CSF can contribute significantly to global climate action and sustainable development goals (Verkerk et al., 2020). This approach necessitates robust policy frameworks, innovative silvicultural practices, and active stakeholder engagement to facilitate the transition towards more resilient and climate-beneficial forest ecosystems. In recent years, the interest to promote and assess CSF has increased. Definitions of CSF were developed, and a set of indicators was defined as a tool for assessing and monitoring it (Alfieri et al., 2024; Bowditch et al., 2020; Nabuurs et al., 2018; Verkerk et al. 2020).

To provide guidance on CSF and restoration, the availability of data assimilation routines, spatially explicit forest data and scenario simulations are essential. Our aim was therefore to develop data assimilation routines and processing tools to enable the evaluation of restoration and CSF management options' effectiveness, based on observations and forward-looking projections. This Deliverable report describes improved data assimilation routines, extended model and processing tools and their access.

1.2. Report structure

Chapter 2 presents two web-applications, namely CSF-Ind and CSF-AT, for quantifying a set of CSF indicators with a focus on forest management, from local to national level. The development of such data processing tools is useful for evaluating the state and trends of forests and their management in the context of CSF. The development of user-friendly tools enabling management of a huge quantity of data can represent a viable way to explore forest structure and assess CSF. The web-applications developed in this project represent a starting point for exploring forest data, providing outputs on CSF that are useful for forest managers and policy decision makers.

Chapter 3 presents a data processing and assimilation routine for synthesising spatially explicit forest data to provide up-to-date and detailed information about the location, composition, and structure of forests by combining field and remotely-sensed data products. Forest projection models, such as EFISCEN-Space, are often used for the simulation of forest development in the upcoming decades for different management and climate change scenarios to evaluate the effectiveness of CSF management practices, as well as the impact of climate change and associated disturbances. Such forest models are commonly initialized with data from National Forest Inventories (NFIs), or similar forest inventories, which contain information on the distribution of tree diameters and species in individual sample plots. However, these data are not always complete, they do not always provide sufficient coverage for local assessments (~one small plot per 300-500 ha is normal), and may not represent the most recent state of the forest. This chapter tests alternative methods to address completeness, coverage, and timeliness by combining field-based forest inventories and remote sensing products, as a basis for forest modelling.

Chapter 4 describes how data on forest structure can be improved as a basis for process-based models, using LPJ-GUESS as an example. This chapter focuses on improving the timeliness of data used for such projections by leveraging on the latest remote-sensing based data on forest disturbances and harvest to estimate stand age. It also describes a new feature to directly force stand-replacing disturbances in LPJ-GUESS with observations of disturbance occurrence from satellite information.

Case studies were reported and explained for each developed tool. Altogether, the chapters present improved data assimilation routines, extended model and processing tools that will be used in a next step to develop good practice guidance on effective CSF and restoration options.

1.3. References

- Alfieri, D., Tognetti, R., & Santopuoli, G. (2024).** Exploring climate-smart forestry in Mediterranean forests through an innovative composite climate-smart index. *Journal of Environmental Management*, 368, 122002.
- Bowditch, E., Santopuoli, G., Binder, F., del Río, M., la Porta, N., Kluvankova, T., Lesinski, J., Motta, R., Pach, M., Panzacchi, P., Pretzsch, H., Temperli, C., Tonon, G., Smith, M., Velikova, V., Weatherall, A., & Tognetti, R. (2020).** What is Climate-Smart Forestry? A definition from a multinational collaborative process focused on mountain regions of Europe. *Ecosystem Services*, 43. <https://doi.org/10.1016/j.ecoser.2020.101113>.
- Jandl, R., Ledermann, T., Kindermann, G., Freudenschuss, A., Gschwantner, T., & Weiss, P. (2018).** Strategies for Climate-Smart Forest Management in Austria. *Forests*, 9(10), 592. <https://doi.org/10.3390/f9100592>.
- Nabuurs, G. J., Delacote, P., Ellison, D., Hanewinkel, M., Hetemäki, L., & Lindner, M. (2017).** By 2050 the mitigation effects of EU forests could nearly double through climate smart forestry. *Forests*, 8(12), 484.
- Nabuurs, G. J., Verkerk, P. J., Schelhaas, M. J., González Olabarria, J. R., Trasobares, A., & Cienciala, E. (2018).** Climate-Smart Forestry: mitigation impacts in three European regions (Vol. 6, p. 31). Helsinki, Finland: EFI.
- Verkerk, P. J., Costanza, R., Hetemäki, L., Kubiszewski, I., Leskinen, P., Nabuurs, G. J., ... & Palahí, M. (2020).** Climate-smart forestry: the missing link. *Forest Policy and Economics*, 115, 102164.
- Yousefpour, R., Augustynczyk, A. L. D., Reyer, C. P., Lasch-Born, P., Suckow, F., & Hanewinkel, M. (2018).** Realizing mitigation efficiency of European commercial forests by climate smart forestry. *Scientific reports*, 8(1), 345.
- Weatherall, A., Nabuurs, G., Velikova, V., Santopuoli, G., Neroj, B., Bowditch, E., ... & Tognetti, R. (2022).** Defining climate-smart forestry. In *Climate-Smart Forestry in Mountain Regions* (Vol. 40, pp. 35-58). Springer.

2. Data Processing Tools

2.1. Introduction

Assessing CSF represents one of the main challenges of the last decade. Several efforts have been made to identify indicators for its evaluation in Europe. Indicators are quantifiable measures that reflect the state of a forest ecosystem or its management practices. The use of indicators in the forestry sector originates from the evaluation of SFM, developing and adopting a set of pan-European criteria and indicators (C&I) (SoEF, 2020; Wijewardana, 2008). The C&I were designed as a policy instrument to evaluate progress towards SFM, guide national forest policies, and facilitate transparent reporting across European countries. The SFM framework was built around six pan-European criteria, and associated with these six criteria, a set of quantitative and qualitative indicators was developed. Starting from the pan-European C&I set, a set of CSF indicators was defined for three main CSF pillars: (I) adaptation, with indicators focusing on making forests more resilient to the impacts of climate change; (II) mitigation, with indicators measuring the contribution of forests and wood products to reducing greenhouse gas emissions; and (III) social dimension, with indicators focusing on the social and economic aspects (Alfieri et al., 2024; Santopuoli et al., 2021; Bowditch et al., 2020). Some studies highlight that the most common indicators used in the field of CSF include carbon stock, forest damage, tree species composition, age structure or diameter distribution, roundwood production, common forest bird species, soil conditions, regeneration, net revenue, wood energy, and forest fragmentation (Alfieri et al., 2024; Santopuoli et al., 2021). The indicators are specific, quantifiable, and comparable over time, reflecting broader ecological concepts, so they have interpretative power needed to inform decision-making. Nevertheless, uncertainties about the practical use of indicators for the evaluation of CSF management performance are still present. To address these issues, Alfieri et al. (2024), developed an index called "composite climate-smart index" (I_{CSF}) designed to quantitatively evaluate CSF indicators implementation, selecting the most important indicators after a review. In this context, more efforts are needed to translate the CSF theoretical concept into practical applications, developing a replicable, scalable, and user-friendly method to assess the effectiveness of CSF application over time (Alfieri et al., 2025). Accordingly, two web-based applications were developed: Climate Smart Forestry Indicators Generator (CSF-Ind) and Climate Smart Forestry Assessment Tool (CSF-AT).

Web-based applications provide users with an accessible and automated data processing tool. In this work, the web-based apps that were developed enable users to generate CSF indicators and to evaluate CSF-practices through a user-friendly tool without requiring prior coding skills.

The developed web-based apps connect users with both frontend and backend components (Figure 1).

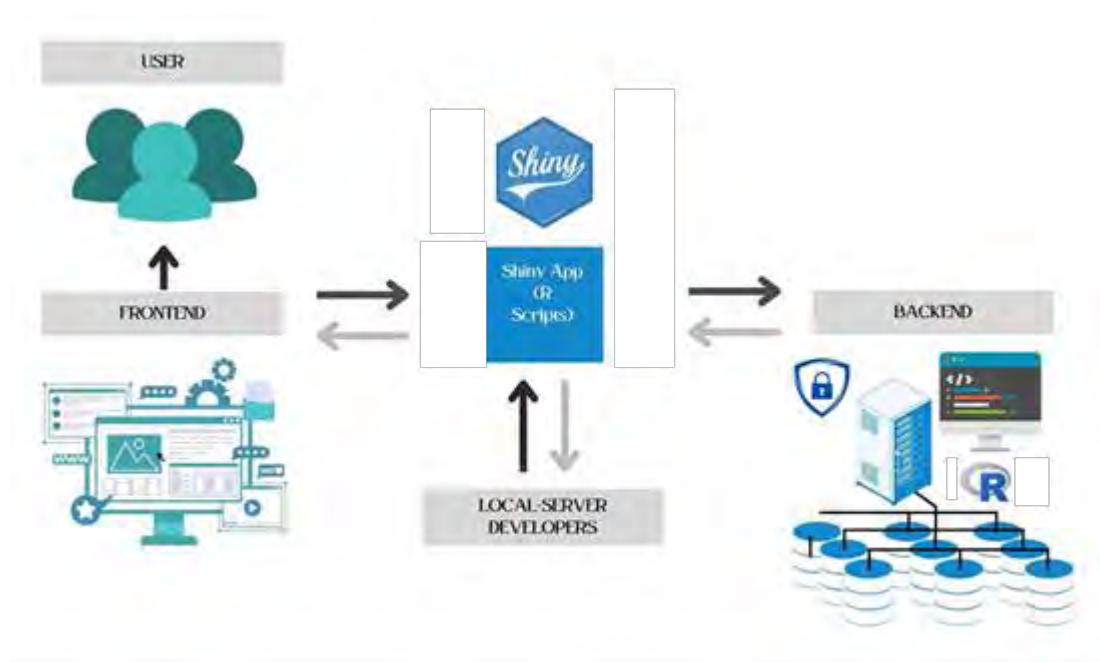


FIGURE 1 FRONTEND AND BACKEND COMPONENTS OF THE WEB-BASED APPLICATION ARCHITECTURE.

Frontend components consist of presentation layers, such as a Shiny User Interface, which are interpreted and executed by the client-side (browser). The user interface should be user-friendly, easy to use, and efficient, requiring minimal input from the user to achieve the desired output. Backend components include the scripts and algorithms produced in programming languages, which are executed on the server to handle data processing and database management. To develop both web applications, the Shiny (Chang et al., 2024) frontend and R (R core team, 2025) backend were used. Developers implemented the code list, such as scripts in the R environment, allowing the final users (e.g., researchers, forest technicians, general public) to benefit from the Shiny user interface.

2.2. Climate-Smart Forestry Indicators Generator (CSF-Ind)

In the CSF-Ind app (<https://ricercaforestale-unimol.shinyapps.io/CSF-Ind>), it is possible to use forest inventory data (e.g., field surveys or remote/proximal sensing data) at tree, plot, or stand level to assess CSF-indicators under the adaptation and mitigation pillars. It allows to set-up the analysis for single plots as well as several plots (multi-plot), displaying preliminary results directly on the App (Figure 2). However, there is also the possibility to download the outputs in table format or as a PDF report.

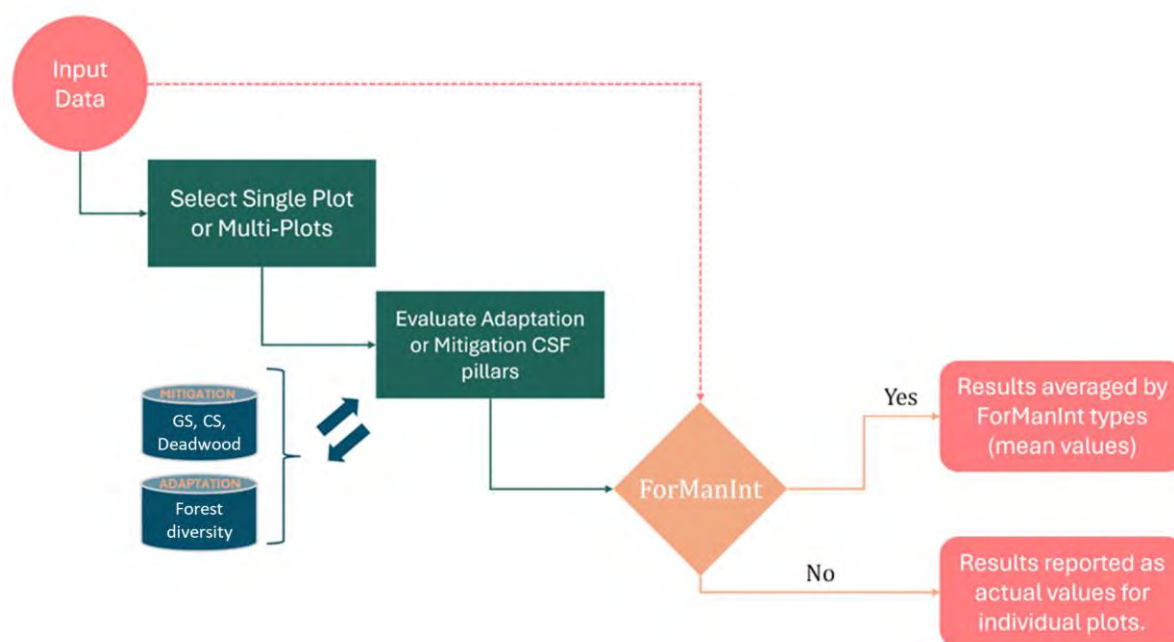


FIGURE 2 WORKFLOW OF CSF-IND APP.

The CSF-Ind is composed of a user interface where users can customize some parameters and select the variables required to calculate CSF indicators. The layout structure of the interactive dashboard of the CSF-Ind web-app is composed of three parts: (i) Header/navigation; (ii) Sidebar controls; (iii) Main content area (Figure 3).

i) In the Header/navigation page, the Header includes the name of the web-app and the logo of the web-app developers, namely the FORWARDS project and UNIMOL. On the left part, there are the interactive elements such as the three pillars of CSF (mitigation, adaptation and social dimension) which the users can select for their evaluation. Despite the social dimension being included in the web-app, in the current version it is not working, but it is scheduled to be updated over time.

ii) The Sidebar controls is an interactive part within which users can import input data, select indicators and forest inventory variables through a dropdown menu, to set-up the analysis. Moreover, there is the possibility to download the outputs in two formats, such as a report as PDF file, and a table as .csv file.

iii) The main content area is the area where outputs of the analysis are shown. It includes an interactive EU map, output data in table format, and charts that summarise the results of the analysis.

The input data required and the output data obtained from the web-based app are explained in detail in the next sections.

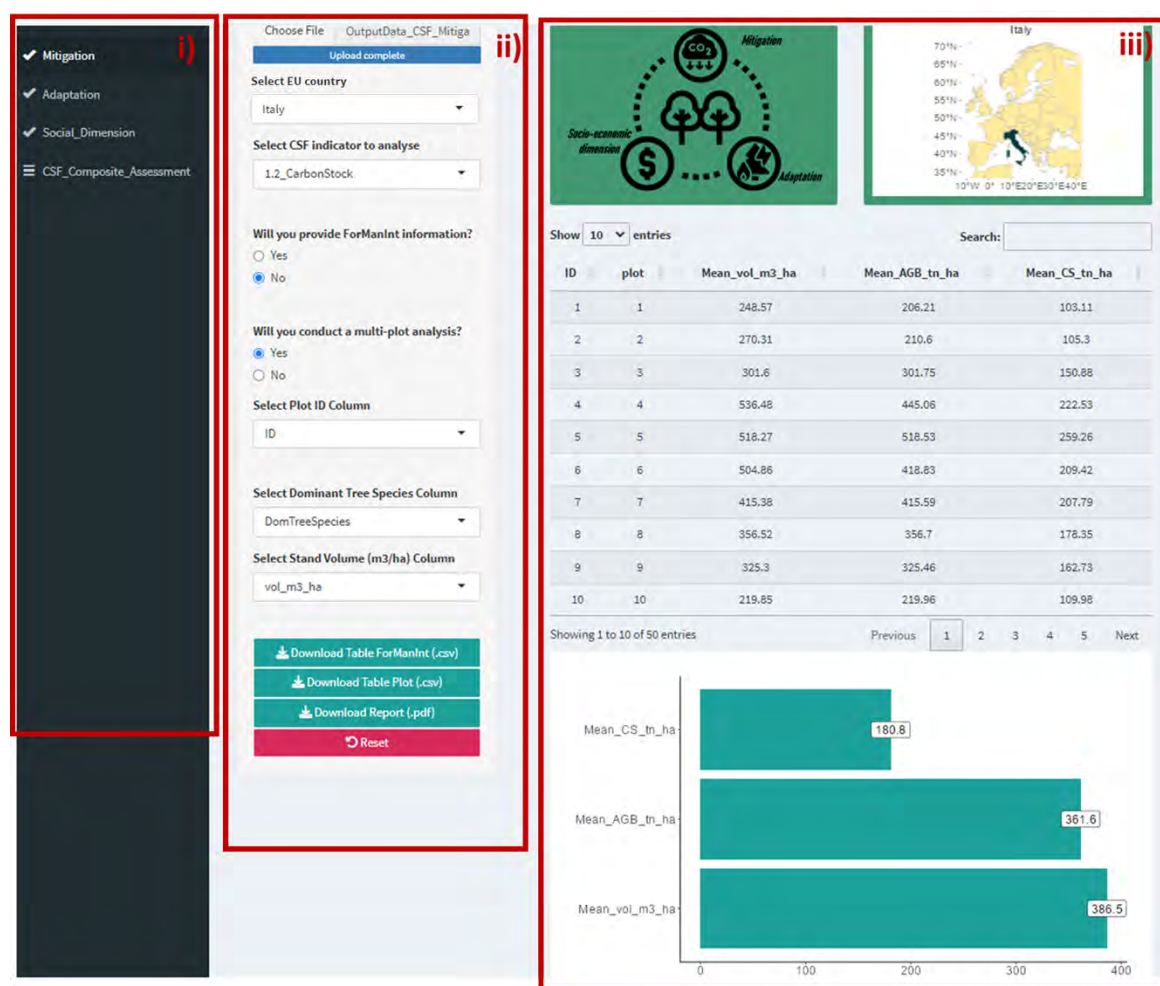


FIGURE 3 CSF-IND WEB-APP USERS INTERFACE: (I) HEADER/NAVIGATION; (II) SIDEBAR CONTROLS; (III) MAIN CONTENT AREA.

2.2.1. CSF-Ind data input

After launching the CSF-Ind app, users have to select which pillar must be detected; adaptation or mitigation (Figure 3). According to some recent studies (Bowditch et al., 2020; Santopuoli et al., 2021; Alfieri et al., 2024), there are some indicators that are highly important for mitigation and adaptation evaluation. For this reason, the CSF-Ind app allows users to select which indicator to evaluate and the forest inventory variables required for the calculations. In the sidebar control (Figure 3- ii), it is mandatory to upload, in the section “Choose file”, an Excel input file (.xlsx, see the example in Figure 4), and select the country where the data were collected in the section “Select EU country”. In the current version, Italian allometric equations have been used for developing the web-app. However, additional equations can be uploaded, even by external users, after the web-based app release. In the section “Select CSF indicator to analyse”, it is necessary to select from the dropdown list, which indicator should be evaluated (Figure 5). Thereafter, the users have to specify (i) whether or not data concerning Forest Management Intensity (ForManInt) are available, (ii) if the data refer to a single or more plots (i.e., multi-plot), and (iii) enter various required variables, either by choosing from dropdown columns or adjusting via slider buttons. This aspect is extremely important since the users have to identify the correspondences between required variables and forest inventory variables from the Excel file to run the web-app.

ID_plot	ID_tree	Species	DBH	Tree_height	ForManInt
4	16	abies_alba	14	10.401	A
4	19	abies_alba	8.5	6.177	A
4	39	abies_alba	9	5.122	A
4	40	abies_alba	9.5	5.623	A
6	7	abies_alba	8	4.55	P
6	8	abies_alba	9.5	4.911	P
6	9	abies_alba	12	8.496	P
6	10	abies_alba	11.5	8.112	P
6	15	abies_alba	6	4.13	P
6	19	abies_alba	19.5	10.638	P
6	24	abies_alba	8	4.02	P

FIGURE 4 EXAMPLE OF INPUT DATA AS .XLSX FILE FOR COMPUTING THE GROWING STOCK.

FIGURE 5 INTERFACE OF CSF-IND WEB-APP SHOWING THE INTERACTIVE SIDEBAR CONTROL TO SELECT CSF PILLARS AND THE RELATIVE INDICATORS THAT USERS WANT TO EVALUATE.

The data required to launch the CSF-Ind app are different for each indicator, and they are summarised and described in Table 1.

To calculate the **growing stock indicator**, the data necessary are:

- **IDPlot**: Number to identify the field plot;
- **PlotArea (m²)**: Size of plot area reported in m². It is possible to select the size of plots with a slider button. A unique value for all plots is required.
- **IDTree**: Number to identify each single tree within the plots;
- **TreeDiameter (cm)**: Diameter reported in cm per each tree measured at breast height (1.3 m height);
- **TreeHeight (m)**: Total tree height for each tree in meters;
- **TreeSpecies**: the binomial Latin name of species per each tree. The tree species name composed of at least two words needs to be separated by “_”, such as “fagus_sylvatica”, “pinus_sylvestris”. In some cases, species are grouped at genus level (e.g. “acer”). Tree species names allowed in the column are listed in the supporting information section (Section 2.2.3).

The variables above reported are also necessary to compute the **forest diversity indicators** in the **adaptation** pillar.

To calculate the **carbon stock indicator**, the mandatory data have to be inserted at plot-level, and they are:

- **IDPlot**: Number to identify the field plot;
- **Volume per hectare**: the volume computed in the plot reported per hectare (m³ ha⁻¹);
- **Dominant tree species**: the tree species dominant in the plot (which we defined as the tree species with highest % of basal area within the plot).

However, to compute the carbon stock, it is possible to directly use the output (.csv) generated through the growing stock’s CSF-ind.

An important component that influences the three pillars of CSF is the deadwood volume, because it indicates the ability of the forest ecosystem to maintain biodiversity and ecological functions. It consists of several components: standing dead tree (sdt), lying dead tree (ldt), coarse woody debris (cwd), snag, and stump (Lombardi et al., 2008).

To compute the **deadwood indicator** (Puletti et al., 2019), it is important to collect the following information:

- **Deadwood type**: indicating the component of deadwood which is referred: sdt, ldt, cwd, snag, and stump; this is the nomenclature that is recognized by the web-app;
 - **For sdt and ldt**
 - **DBH (cm)**: diameter at breast height (1.30m) in cm;
 - **Height (m)**: total height/length of sdt/ldt in m;
 - **For snag and cwd**
 - **D half (cm)**: diameter at half-length in cm;
 - **Length (m)**: total length of the element in m;
 - **For stump**
 - **Height cut (m)**: total height of section from the ground in m;
 - **D cut (cm)**: diameter at the top section of the stump in cm.

If the user collected data in forests with different management intensity and wants to compute the indicators while retaining that information, optional data (information) are needed, specifically, the forest management intensity as reported below:

ForManInt (Forest management intensity based on Dunker et al., 2012), is a data column, which needs to contain one of the following options (percentage refers to the percentage of harvested timber volume):

- P - Passive - Unmanaged forest nature reserve (no harvesting, 0%)
- L - Low - Close-to-nature forestry (single or group selection, <20%)
- M - Medium - Combined objective forestry (shelterwood - coppice with standards 40-70%)
- H - High - Intensive even-aged forestry (clearcut, 70-100%)
- I - Intensive - Short rotation forestry (short-rotation forestry, 100%).
- NA - Not Available

For potential future improvements of the web-app, additional input variables, such as deadwood decay class and geographical information, could be useful. The deadwood decay class information can be useful to evaluate changes of deadwood quality status over time. While, the geographical data can be useful to develop CSF interactive maps.

- **deadwood decay class:** all deadwood components should be classified into five-decay classes related to characteristics of deadwood (Hunter, 1990; Gasparini, 2022):
 - RD - recently dead - Bark still attached.
 - WD - weakly decayed - Bark is loose but not fragmented.
 - MD - medium decayed - Bark is fragmented. Wood consistency is reduced but the log still has a hard core.
 - VD - very decayed - Bark absent. Rotten throughout the log.
 - D - decomposed - Bark absent. Wood consistency is lost (dust).
- **Geographical data**
 - **CoordsTree:** xCoordinates, yCoordinates for each tree (WGS84 or ETRS89-extended / LAEA Europe);
 - **CoordsPlot (Lat., Long.):** provide the coordinates (WGS84 or ETRS89-extended / LAEA Europe) of the centroid of each field plot;
 - **AOI_stand:** Area of Interest (AOI) boundary (.shp): shapefile of the forest stands;
 - **AOI_plots:** Single Plots boundary (.shp): shapefile of the sampling plots.

2.2.2. CSF-Ind data output

The CSF-Ind app will return outputs on mitigation and adaptation pillars of CSF. In particular, the web-app will provide results for three indicators under the mitigation aspect, namely: **growing stock** (GS), **carbon stock** (CS), and **deadwood** (DW). Moreover, for the adaptation pillar, the results obtained are composed of 15 indices of **forest diversity** related to tree species composition and forest structural data - e.g., Shannon, Simpson, and Gini index. The output data for each pillar/indicator are explained in Table 1.

TABLE 1 CALCULABLE INDICATORS WITH CSF-IND WEB-APP. “*” INDICATES THAT OUTPUT DATA OBTAINED FOR GROWING STOK CALCULATION, CAN BE USED AS INPUT DATA FOR ASSESSING CARBON STOCK. SOME OUTPUT CAN BE USED AS INPUT FOR CSF-AT WEB-APP.

Climate Smart Forestry (CSF) dataset for CSF-Ind				Input for CSF-AT
Pillars	Indicator	Input Data	Output data	
Mitigation	Growing Stock (GS)	IDPlot IDTree PlotArea TreeHeight TreeDiameter TreeSpecies ForManInt	* Vol_m3_ha : forest volume (m ³) regarding the volume of stem and large branches (trees $\varnothing \geq 5$ cm) * DomTreeSpecies : Dominant Tree Species defined as the tree species with highest % of basal area within the plot	Growing Stock (GS)
	Carbon Stock (CS)	IDPlot IDTree PlotArea StandVolume (Vol_m3_ha) DomTreeSpecies ForManInt	AGB_tn_ha : aboveground biomass (Mg ha ⁻¹) CS_tn_ha : Carbon Stock (Mg ha ⁻¹) CO2_tn_ha : carbon dioxide (Mg ha ⁻¹)	Carbon Stock (CS)
	Deadwood (DW)	IDPlot IDTree Height, Length, Height cut (m) DBH, D Half, D cut (cm) Deadwood type	Dead_vol_m3_plot : the total deadwood volume in the plot (m ³) Dead_vol_m3_ha : the deadwood volume per hectare (m ³ ha ⁻¹)	Deadwood (DW)
			SI_sp : Simpson of tree species; SH_sp : Shannon-Weiner of tree species	Tree species composition (TS)
Adaptation	Forest diversity	IDPlot IDTree PlotArea TreeHeight TreeDiameter BasalArea TreeSpecies ForManInt	SI_dbh : Simpson of diameter measured at breast height (DBH); SH_dbh : Shannon-Weiner of diameter measured at breast height (DBH); GI_dbh : Gini of diameter measured at breast height (DBH); Mean_dbh : mean of diameter measured at breast height (DBH); SD_dbh : standard deviation of diameter measured at breast height (DBH)	Diameter distribution (DB)
			Mean_SLC : ratio between the mean of tree height and the mean of diameter measured at breast height (DBH)	Slenderness Coefficient (SLC)
			Mean_th : mean of tree heights (m); Sum_ba : sum of basal area (m ²); N_sp : number of tree species; GI_ba : Gini of basal area;	
			CV_ba : coefficient of variation of basal area; SD_th : standard deviation of tree heights; SH_th : Shannon-Weiner of tree heights	

The results are displayed directly in the user interface, where it is possible to visualize the result at plot level. Specifically, the calculated indicators will be shown in: i) table format and ii) bar-chart in the main area of the CSF-Ind interface (Figure 6).

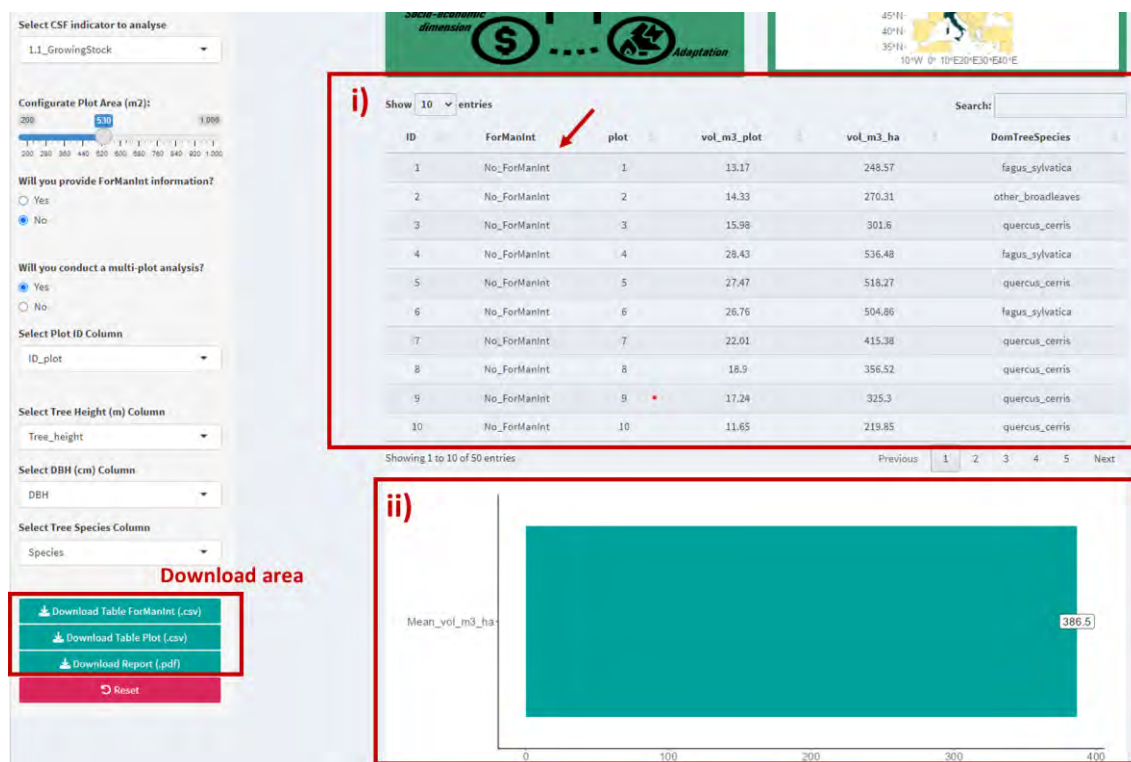


FIGURE 6 EXAMPLE OF OUTPUT FOR GROWING STOCK INDICATOR: i) THE RESULTS AT PLOT-LEVEL IN TABLE FORMAT, ii) THE BAR-CHART REPORTS THE VALUE OF THE INDICATOR PER HECTARE COMPUTED ACROSS THE PLOTS PRESENT IN THE DATASET. THE ARROW INDICATES THE OUTPUT IN THE TABLE WHEN FORETS MANAGEMENT INTENSITY DATA IS NOT AVAILABLE.

Additionally, from the user interface, for each indicator (GS, CS, DW, Forest Diversity) the users can choose to download the data in different formats:

- in PDF (.pdf), the data are downloaded as a report in table and graph formatted by markdown (Baumer & Udwin, 2015);
- in csv table (.csv), the data can be grouped at forest management intensity or plot level (Figure 6).

When forest management intensity data is not available, “No_ForManInt” will be shown in the ForManInt column (see the arrow reported in Figure 6). The .csv table of Growing Stock (**GS**) could be used to compute the Carbon Stock (**CS**) in this web-app; likewise, some indicators computed in the adaptation pillars as **forest diversity indicator**, could be reconducted to the indicators used as input in CSF-AT app (Table 1). A more detailed explanation of linkage between CSF-Ind and CSF-AT is described in Section 2.4.

2.2.3. Supporting information of CSF-Ind app

List of tree species names allowed

In the current version of the app, there are allometric equations for different tree species. As mentioned, the app was implemented with Italian equations, and the list of tree species came from national forest inventory documents (Tabacchi et al., 2011; Gasparini et al., 2022). Users have to use scientific nomenclature with underscore between genus and species such as “abies_alba”. In some cases, species are grouped at the genus level (i.e., carpinus) because there is not a specific equation, while users can select “other_broadleaves” or “other_coniferous” in the cases where both genus and species are missed in tree list, as reported below:

- abies_alba
- acer
- alnus
- carpinus
- castanea_sativa
- cupressus
- eucalyptus
- exotic_pines_group
- fagus_sylvatica
- fraxinus
- larix_decidua
- picea_abies
- pinus_pinaster
- pinus_cembra
- pinus_halepensis
- pinus_nigra
- pinus_nigra_laricio
- pinus_pinea
- pinus_sylvestris
- pseudotsuga_menziesii
- quercus_cerris
- quercus_ilex
- quercus_pubescens
- robinia_pseudoacacia
- salix
- other_broadleaves (all broadleaf species not included above)
- other_coniferous (all coniferous species not included above)
- small_conifers

2.3. Climate-Smart Forestry Assessment Tool (CSF-AT App)

The Climate-Smart Forestry Assessment Tool (CSF-AT) is an interactive and user-friendly web-based application that allows users to assess the degree of “climate-smartness” of forest stands. It provides a structured assessment of the three main pillars of CSF (i.e., mitigation, adaptation, and socio-economic dimension). The app automatically calculates a Composite Climate-Smart Index by aggregating a set of ecological and structural indicators through a hierarchical method (Alfieri et al., 2024). Users can upload their own forest data containing precalculated values of relevant indicators (such as carbon stock, growing stock, deadwood, or regeneration) in CSV format (Figure 7, panel A). Once data is uploaded, the users can start the calculation of the “smartness” level of the forest stand (Figure 7, panel B). The results can be visualized both in tabular format (Figure 7, panels B1 and B2) and graphical format, through an interactive ternary plot (Figure 7, panel C), which enables users to compare the smartness level of multiple forest stands simultaneously.

A more detailed explanation of the input data structure, the methodological framework for the composite index calculation, and the result outputs of the CSF-AT app are provided in the following sections.

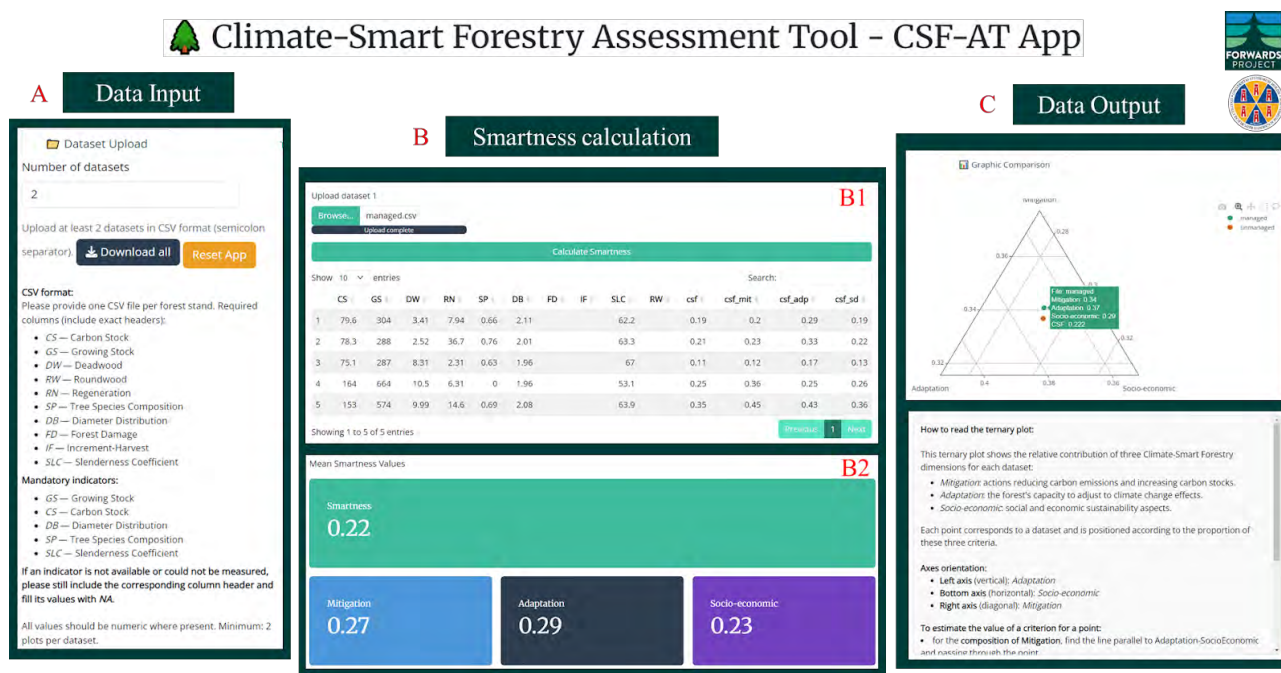


FIGURE 7 EXAMPLE OF CSF-AT APP INTERFACE.

2.3.1. CSF-AT app data input

CSF-AT was developed using 10 indicators, five mandatory and five optional. A description of the CSF indicators with their relative possible calculation methods is provided in Table 2. The data input process is managed in the first section of the application, titled “Dataset Upload” (Figure 7, panel A). In this section, users define the number of datasets they wish to analyse (a minimum of two datasets is required). Based on this choice, the corresponding upload panels are generated, enabling the upload of CSV files containing the indicator values.

Each uploaded CSV file must comply with the following specifications:

- Each CSV corresponds to a single forest stand.
- Each CSV must include at least two sample plots.
- The file must contain columns for the required indicators (see Table 2 below).
- Column order is not relevant, but column names must match exactly.
- If values for optional indicators are not available, enter “NA”.
- All indicator values must be calculated in advance.
- All entries must be numeric values.

TABLE 2 CSF INDICATORS FOR CSF-AT. THE TABLE REPORTS ABBREVIATION, GENERAL DEFINITION, POSSIBLE CALCULATION METHODS, AND MEASUREMENT UNITS. GS, CS, DB, TS, AND SLC ARE MANDATORY INDICATORS, WHILE DW, RN, FD, IF, AND RW ARE OPTIONAL.

Indicator	Abbreviation	General definition	Possible calculation methods	Unit
Growing stock	GS	Volume of living trees per hectare	Based on stem volume equations or inventory data	m ³ ha ⁻¹
Carbon stock	CS	Organic carbon stored in aboveground biomass	Derived from biomass equations or fixed carbon conversion factors	Mg ha ⁻¹
Diameter distribution	DB	Diversity or variability of tree diameters in the stand	Shannon index, Gini index, or mean quadratic diameter	–
Tree species composition	TS	Diversity of species in the stand	Species richness, Shannon index, Simpson index	–
Slenderness coefficient	SLC	Relation between tree height and diameter, indicator of stability	Mean height/mean diameter or height/diameter at tree-level	–
Deadwood	DW	Volume of deadwood	Direct measurement or inventory-based estimates	m ³ ha ⁻¹
Regeneration	RN	Amount and quality of natural regeneration	Density counts, diversity indices, or regeneration-specific indices	–
Forest damage	FD	Tree health and vitality	Crown condition, defoliation, damage classes	–
Increment-Felling	IF	Balance between increment and harvest	Ratio between harvested volume and increment	–
Roundwood	RW	Ratio of industrial to fuelwood production	Based on assortments in harvested volume	–

Calculation of the composite CSF index

Once the CSV files are uploaded, users can calculate the smartness values for each dataset by clicking the “Calculate Smartness” button. The application produces results at two complementary levels (Figure 8):

- **Plot level:** a table displays the CSF smartness values calculated for each individual plot, including both uploaded indicator values (i.e. CS, GS, DW, RN, SP, DB, FD, IF, SLC, RW) and the computed scores for overall smartness (csf), as well as for the three dimensions - mitigation (csf_mit), adaptation (csf_adp), and socio-economic dimension (csf_sd);;
- **Stand level:** summary boxes show the mean values of these CSF components (Smartness, Mitigation, Adaptation, and Socio-Economic dimension) aggregated across all plots belonging to the same forest stand.

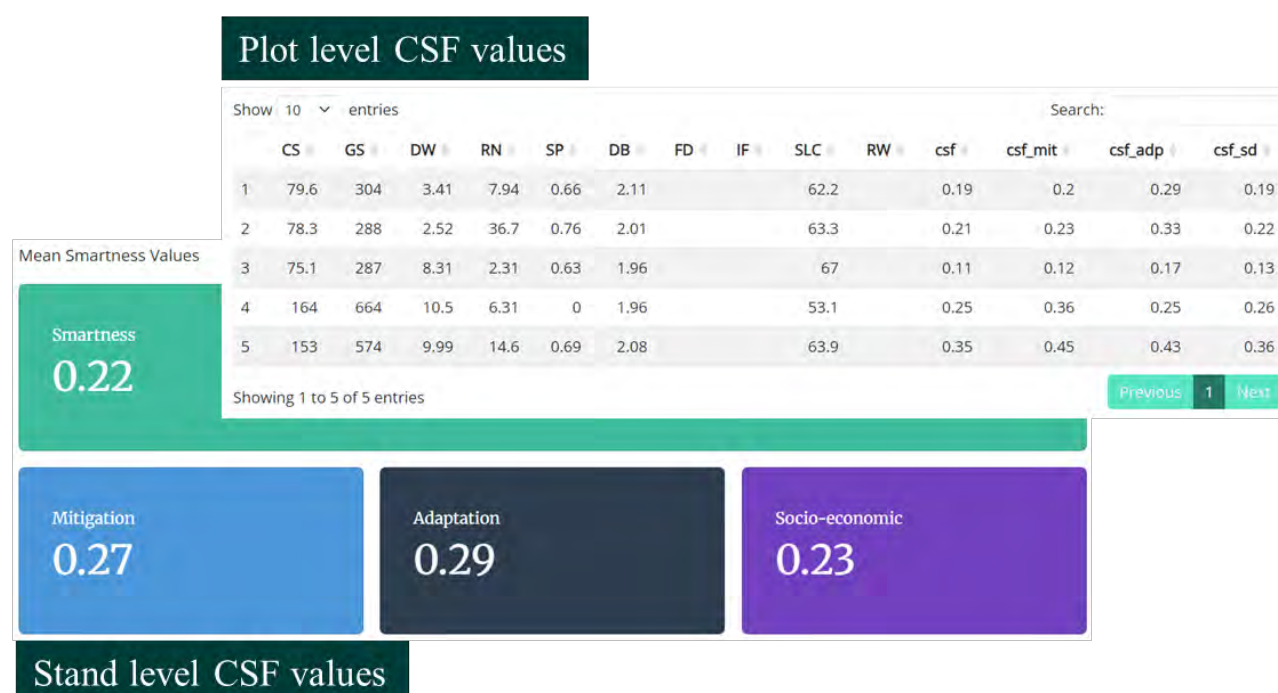


FIGURE 8 EXAMPLE OF OUTPUT GENERATED BY THE APPLICATION WITH THE VALUE OF CSF AT PLOT AND STAND LEVEL.

The calculation relies on a **weighted hierarchical aggregation model** (Alfieri et al., 2024), which is structured as follows (Figure 9):

- **Level 1:** Composite Smartness Index (I_{CSF}), representing the overall smartness of the forest stand;
- **Level 2:** The three main pillars of CSF — Mitigation, Adaptation, and Socio-economic dimension;
- **Level 3:** Individual indicators.

The methodology is based on the following steps:

1. **Normalization of indicators** — For each dataset, indicator values are normalized across the plots within the stand using min–max scaling, in order to make them comparable on a common 0–1 scale, where 0 represents the least desirable case and 1 the most desirable one.
2. **Weighting of criteria and indicators** — Normalized indicators are assigned weights derived from expert judgment and structured according to the Analytic Hierarchy Process (AHP). This ensures that the relative importance of mitigation, adaptation, and socio-economic aspects is explicitly reflected in the calculation.
3. **Hierarchical aggregation** — Weighted indicator values are aggregated first within each pillar, and then across pillars, to obtain the composite index (I_{CSF}).

This entire procedure has been automated within the CSF-AT app: once users click “Calculate Smartness”, the normalization, weighting, and hierarchical aggregation are performed in sequence, and the resulting indices are displayed interactively at both plot and stand level.

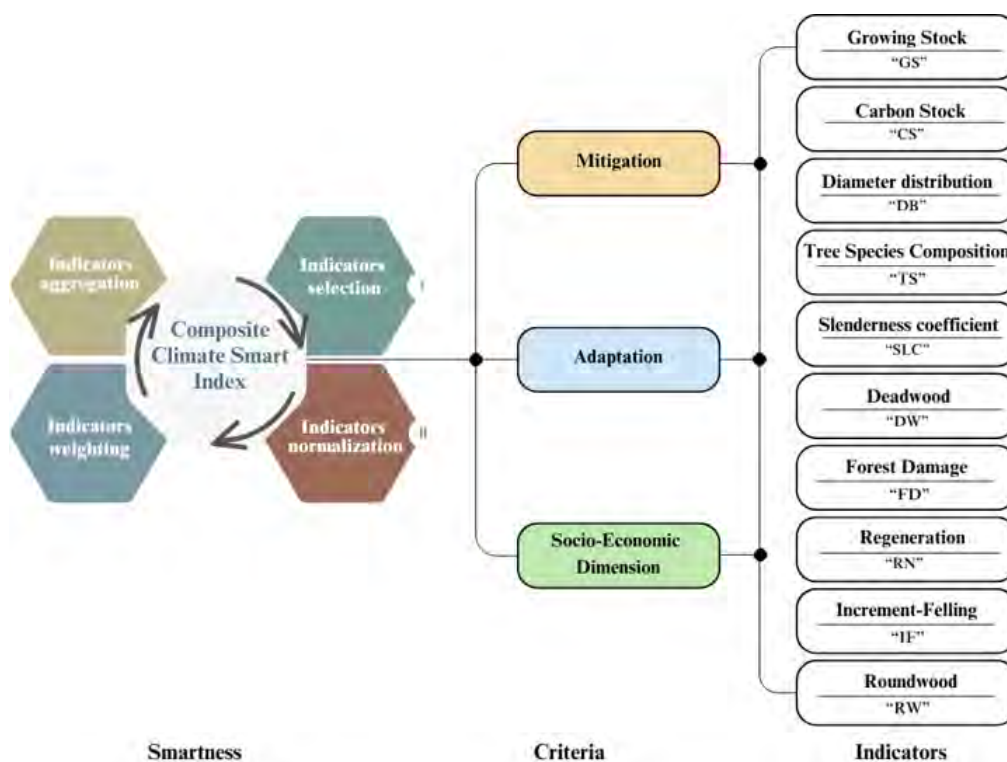


FIGURE 9 FRAMEWORK TO ASSESS CLIMATE-SMART FORESTRY. ON THE LEFT, THE STEPWISE PROCEDURE USED TO CALCULATE THE COMPOSITE INDEX; ON THE RIGHT, THE HIERARCHICAL STRUCTURE OF THE COMPOSITE INDEX.

2.3.2. CSF-AT app data output

Once the smartness values have been calculated, the application provides different options for exporting and visualizing the results. These functionalities are designed to allow both a detailed analysis of individual datasets and a synthetic comparison across multiple cases.

Data export

Users can export the results by clicking the “*Download all*” button. The application generates a single .zip archive containing two CSV files for each uploaded dataset:

- **Plot-level results** (e.g., `dataset_1_results.csv`), which report the CSF index calculated for each individual plot.
- **Stand-level values** (e.g., `dataset_1_means.csv`), which summarize the average smartness values per dataset, providing an aggregated overview.

This structure allows users to maintain both the fine-scale information (plot resolution) and the aggregated perspective (stand level), making it possible to conduct additional analyses or comparisons outside the application.

Graphic comparison

The application includes a dedicated section called “*Graphic comparison*” where the results can be explored interactively through a ternary plot.

- Visualization of CSF dimensions: each dataset is represented as a point within the ternary diagram, which illustrates the relative contribution of the three CSF dimensions (mitigation, adaptation, and socio-economic functions).
- Interactive functions: users can zoom in on specific areas of the graph to better distinguish the positioning of the datasets. By positioning the cursor over a point, a tooltip displays both the relative values of the three criteria and the absolute value of the composite smartness index.
- Interpretation support: a short explanatory text is provided alongside the graph to guide the user in understanding the meaning of the plot and the positioning of the datasets.
- Export option: the ternary plot can be saved in .png format, allowing users to include the graphical results in reports, presentations, or further analyses.

Through this combination of tabular and graphical outputs, the application facilitates both **data availability** (via CSV files) and an **immediate interpretation of trade-offs** between the three CSF dimensions such as the ternary visualization, as reported in the “Results” of the case study presented in section 2.5.2.

2.4. Connecting the CSF-Ind web-app to CSF-AT web-app

The CSF-Ind and CSF-AT web-apps are connected, since the CSF-Ind web-app serves as a preparatory tool for the CSF assessment carried out in CSF-AT (Figure 10). In particular, CSF-Ind allows the calculation of all the mandatory indicators required by CSF-AT, i.e. Growing stock (GS), Carbon stock (CS), Diameter distribution (DB), Tree species composition (TS), and Slenderness coefficient (SLC), as well as one optional indicator, i.e. Deadwood (DW). These indicators can be directly used as input variables in the CSF-AT app, where they contribute to the assessment of the overall smartness score and also its three dimensions (mitigation, adaptation, and social dimension). This linkage, illustrated in Figure 10, shows how the variable output generated by CSF-Ind can be imported as data input in CSF-AT. Therefore, the integration between these two tools provides a streamlined process, supporting users - without coding experience and CSF expertise - in producing a comprehensive evaluation of CSF management, from indicator calculation to overall smartness evaluation.

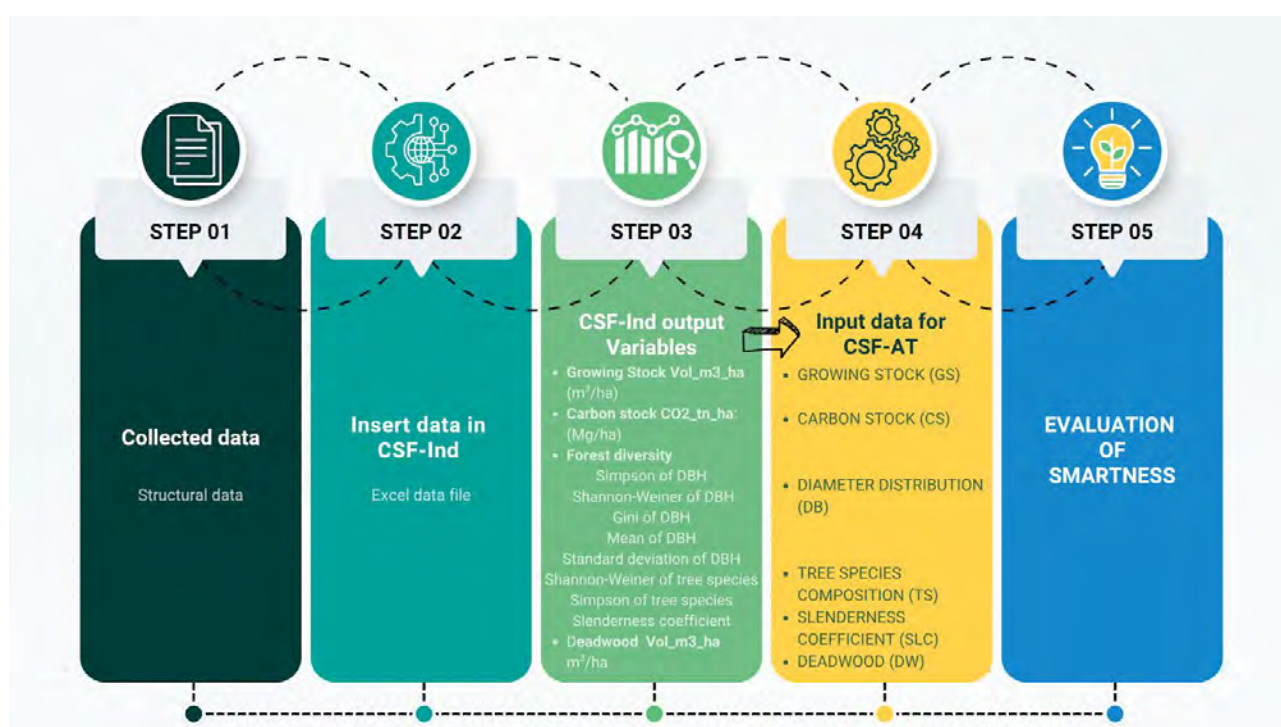


FIGURE 10 WORKFLOW OF CSF-IND AND CSF-AT WEB-APPS (ABOVE). BELOW ARE REPORTED VARIABLES GENERATED PER EACH INDICATOR BY THE CSF-IND WEB-APP THAT CAN BE USED IN THE CSF-AT WEB-APP.

2.5. Case study demonstrations

2.5.1. Practical application of the CSF-Ind: Bosco Pennataro case study

Background

CSF indicators have an essential role in climate smart forest management. However, there are difficulties in computing them. The most common indicators that are currently used are: growing stock, carbon stock, deadwood and forest diversity indices (as reported above, see 2.1). This section aims to show the overall functionality of the CSF-Ind web-app, how to input data and how to obtain output. The web-app was tested to compute the CSF indicators in an Italian FORWARDS demo site.

Method

CSF-Ind web-app was used to compute the CSF indicators for the Bosco Pennataro forest (Molise, Italy), which is within the Italian demo site of the FORWARDS Project. The study area, Bosco Pennataro (Figure 11), is located in the Molise region (Central Italy), in the Municipality of Vastogirardi (Isernia), within a Site of Community Importance “Bosco Monte di Mezzo - Monte Miglio - Pennataro - Monte Capraro - Monte Cavallerizzo”. The mean altitude of this area is 930 m a.s.l., with a mean annual temperature of 8.5 °C, and a mean annual precipitation of approximately 1012 mm (Santopuoli et al., 2019). 50 plots of 529 m² (23 m x 23 m) were established in the area, using the one-per-stratum stratified sampling scheme (Barabesi et al., 2012). Forest inventory variables, including deadwood, were collected and provided as input data to the CSF-Ind web-app.

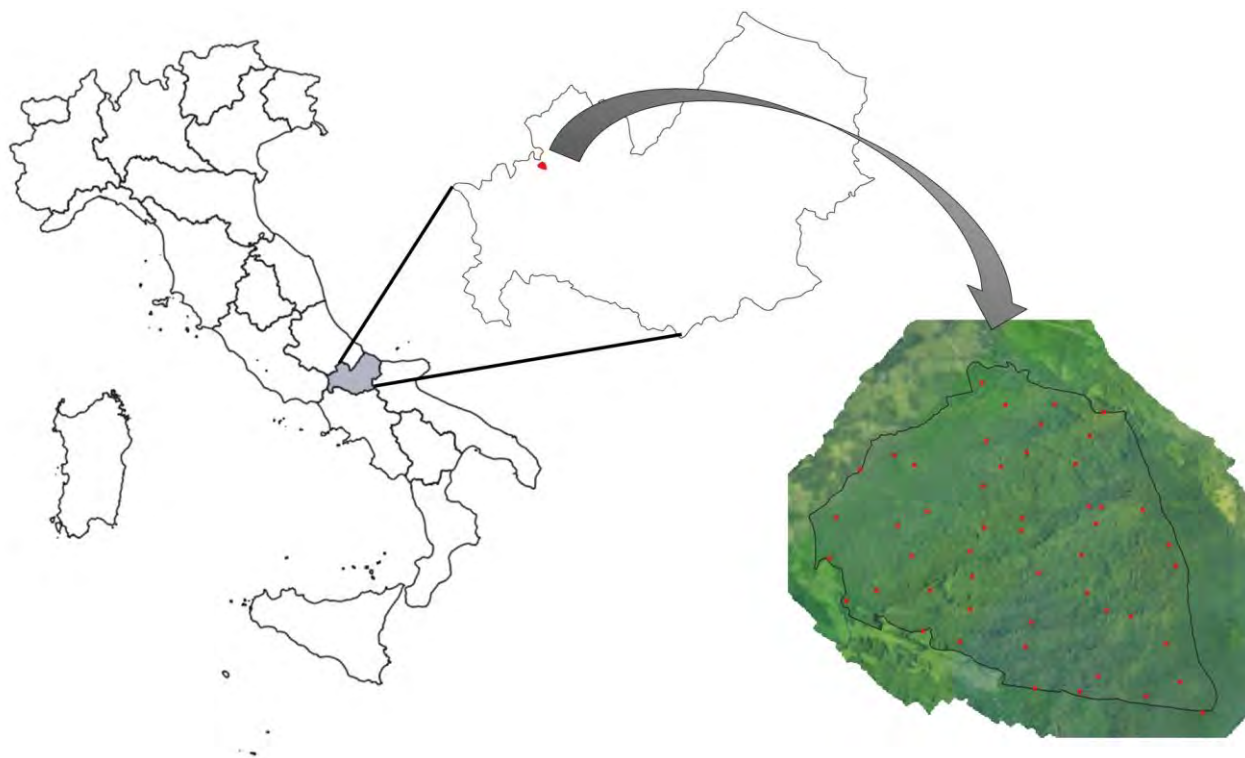


FIGURE 11 LOCATION OF BOSCO PENNATARO IN MOLISE REGION. THE RED DOTS INDICATE ALL THE PLOTS ESTABLISHED IN THE STUDY AREA WHICH WERE USED FOR THE CASE STUDY.

Results

The CSF-Ind web-app computed the growing stock indicator, finding that in the study area the mean volume is $386.5 \text{ m}^3 \text{ ha}^{-1}$ (Figure 12, extracted from the report as PDF file), with a minimum of 141.53 m^3 and maximum of $645.99 \text{ m}^3 \text{ ha}^{-1}$. In our case, we found that more than 50% of total plots have as dominant tree species *Q. cerris*, followed by *F. sylvatica*, other broadleaves, and only one plot is dominated by *Acer* species (Figure 12).

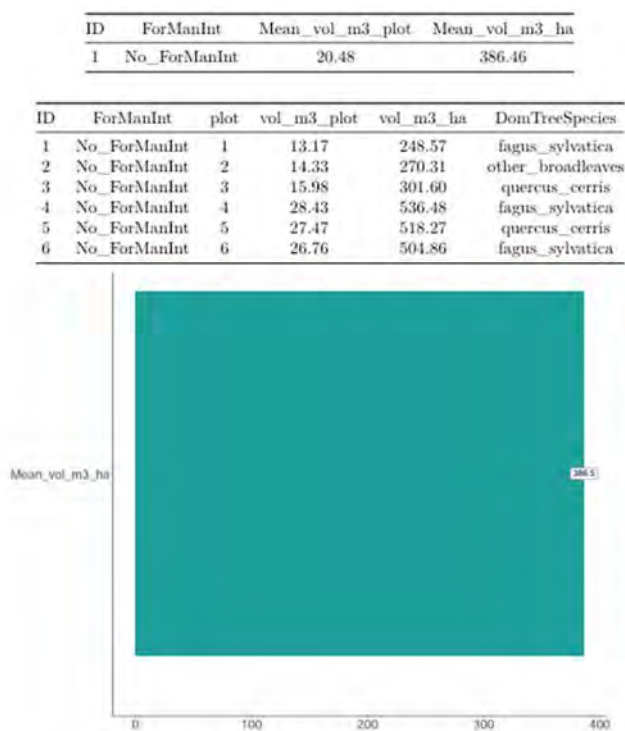
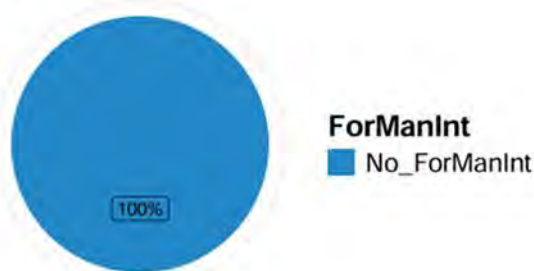


Figure 2: Bar graph/Boxplot of indicator calculated.

Pie-chart of ForManInt



Pie-chart of DomTreeSpecies

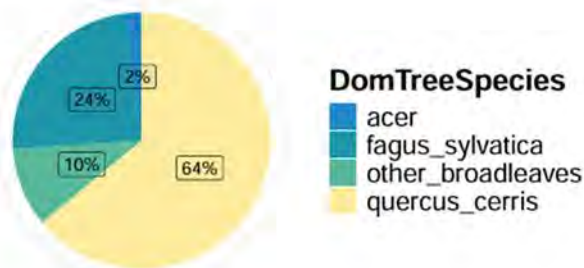


FIGURE 12 MAIN OUTPUT FROM CSF-IND. TABLES REPORT THE VOLUME PER MANAGEMENT INTESITY AND FOR DOMINANT TREE SPECIES, WHILE PIE CHARTS SHOW THE PERCENTAGE OF MANAGEMENT INTENSITY CATEGORIES AND THE TREE DOMINANT SPECIES.

Thereafter the output data were downloaded as a .csv file, at plot level. The table obtained as output shows: the forest management intensity, the name of the plot, the volume per plot and per hectares as well as the dominant tree species (Figure 13):

- **ID:** is the row number
- **ForManInt:** if the forest management intensity information is available, if not “No_ForManInt” is present in the column
- **Plot:** is the ID of plot
- **Vol_m3_plot:** volume in each plot in m³
- **Vol_m3_ha:** volume per hectare (m³ ha⁻¹)
- **DomTreeSpecies:** tree species with the highest % of basal area within the plot.

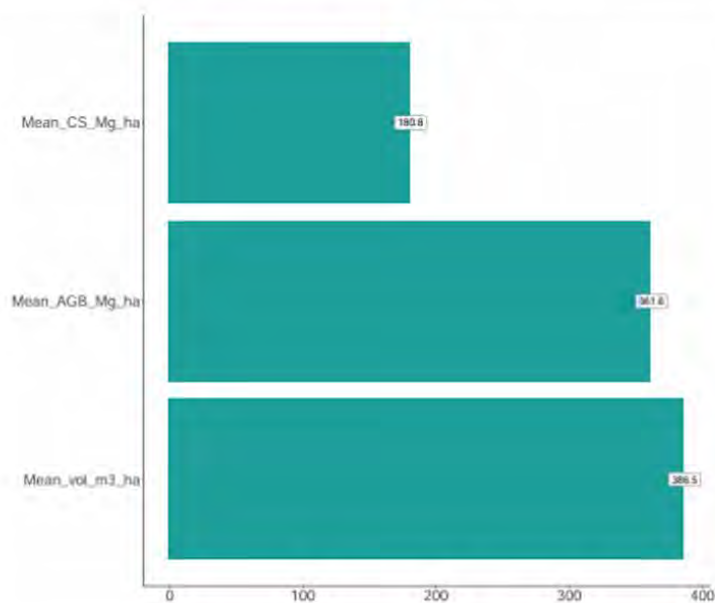
	A	B	C	D	E	F
1	ID	ForManInt	plot	vol_m3_plot	vol_m3_ha	DomTreeSpecies
2	1	No_ForManInt	1	13.17	248.57	fagus_sylvatica
3	2	No_ForManInt	2	14.33	270.31	other_broadleaves
4	3	No_ForManInt	3	15.98	301.6	quercus_cerris
5	4	No_ForManInt	4	28.43	536.48	fagus_sylvatica
6	5	No_ForManInt	5	27.47	518.27	quercus_cerris
7	6	No_ForManInt	6	26.76	504.86	fagus_sylvatica
8	7	No_ForManInt	7	22.01	415.38	quercus_cerris
9	8	No_ForManInt	8	18.9	356.52	quercus_cerris
10	9	No_ForManInt	9	17.24	325.3	quercus_cerris
11	10	No_ForManInt	10	11.65	219.85	quercus_cerris

FIGURE 13 EXAMPLE OF THE OUTPUT FOR THE INDICATOR “GROWING STOCK”.

The dataset was used to calculate the carbon stock. The mean carbon stock reported in our study area is 180.8 Mg ha⁻¹ (Figure 14). The carbon stocked by the *Acer* plot amounted to 120.07 Mg ha⁻¹, while the average of carbon stock in the beech-dominated (*Fagus sylvatica*) plot is 181.81 Mg ha⁻¹. Considering only the plot with Turkey oak (*Quercus cerris*) dominant trees, the carbon stock is 194.77 Mg ha⁻¹, while the other broadleaf plot had a carbon stock of 100.97 Mg ha⁻¹.

ID	ForManInt	TotalPlots	Mean_vol_m3_ha	Mean AGB_Mg_ha	Mean_CS_Mg_ha
1	No_ForManInt	50	386.46	361.57	180.78

ID	ForManInt	plot	DomTreeSpecies	Mean_vol_m3_ha	Mean AGB_Mg_ha
1	No_ForManInt	1	fagus_sylvatica	248.57	206.21
2	No_ForManInt	2	other_broadleaves	270.31	210.60
3	No_ForManInt	3	quercus_cerris	301.60	301.75
4	No_ForManInt	4	fagus_sylvatica	536.48	445.06
5	No_ForManInt	5	quercus_cerris	518.27	518.53
6	No_ForManInt	6	fagus_sylvatica	504.86	418.83



Pie-chart of ForManInt



Pie-chart of DomTreeSpecies

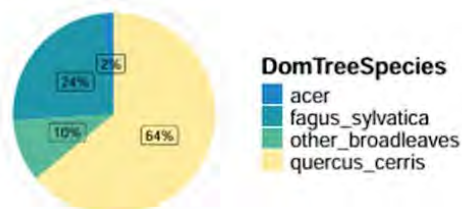
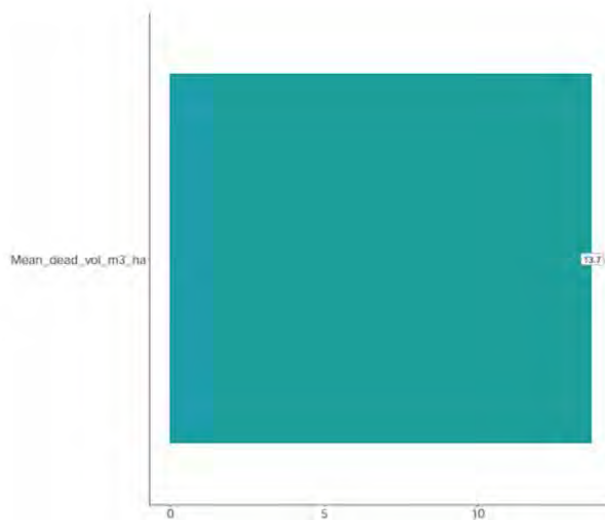


FIGURE 14 TABLES AND BAR-CHART EXTRACTED FROM THE CSF-IND PDF REPORT FOR THE INDICATOR “CARBON STOCK”.

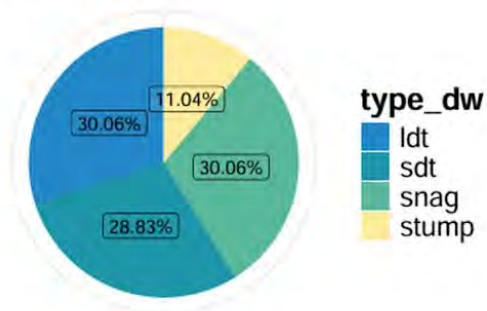
The study area amounts to $13.72 \text{ m}^3 \text{ ha}^{-1}$ average of deadwood (Figure 15). As can be seen in the output at plot-level, in plot 1 the deadwood components present are “ldt” (lying dead wood), and “snag”, where “ldt” is the predominant deadwood component representing $1.34 \text{ m}^3 \text{ ha}^{-1}$. However, in plot 2 the deadwood components are: “ldt”, “sdt” (standing dead wood) and “snag”. In this case, the “sdt” component is the predominant with $4.98 \text{ m}^3 \text{ ha}^{-1}$, followed by the “ldt” with $4.68 \text{ m}^3 \text{ ha}^{-1}$, and lastly the “snag” with a value of $1.66 \text{ m}^3 \text{ ha}^{-1}$. Looking at the deadwood components in our study area, the deadwood is represented mainly by “ldt” (Figure 15), followed by “sdt” (31.90%), “snag” (26.52%), and stump (7.59%).

ID	ForManInt	Mean_dead_vol_m3_plot	Mean_dead_vol_m3_ha
1	No_ForManInt	0.73	13.72

ID	plot	type_dw	ForManInt	dead_vol_m3_plot	dead_vol_m3_ha
1	1	ldt	No_ForManInt	0.07	1.34
2	1	snag	No_ForManInt	0.02	0.32
3	2	ldt	No_ForManInt	0.25	4.68
4	2	sdt	No_ForManInt	0.26	4.98
5	2	snag	No_ForManInt	0.09	1.66
6	3	ldt	No_ForManInt	0.09	1.78



Pie-chart of type_dw



Pie-chart of ForManInt

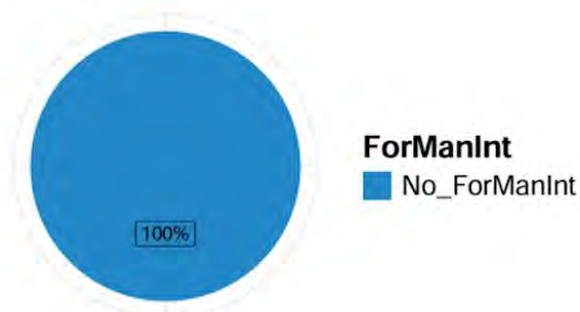


FIGURE 15 OUTPUT FOR INDICATOR “DEADWOOD”. TABLES SHOW DEADWOOD VOLUME, WHILE PIE CHARTS SHOW THE PERCENTAGE OF DEADWOOD COMPONENTS IN THE STUDY AREA, AND THE MANAGEMENT INTENSITY.

Using forest diversity indicators, the results showed that in the study area the number of species vary from 2 species to 6 with an average of 5 (Figure 16), the Simpson index ranged between 0.23 and 0.8 with an average of 0.7, while the Shannon index ranged from 0.4 to 1.68 with a mean value of 1.3. About forest structural data, in the study area the mean of DBH ranged between 9.93 and 26.89 cm, while the mean of tree height had a range between 8.24 and 23.1 m. Regarding the DBH, the indices extracted from CSF-Ind app presented the following ranges:

- Simpson index: 0.92 - 0.99, mean value of 0.97;
- Shannon index: 2.78 - 5.13, mean value of 3.8;
- Gini index: 0.24 - 0.53, mean value of 0.4.

The slenderness coefficient ranged between 0.64 and 1.06, with a mean of 0.84, while the Shannon index of tree height presents a range between 2.98 and 5.22.

ID	ForManInt	TotalPlots	Mean_dbh	Mean_th	Mean_SLC
1	No_ForManInt	50	16.36	13.56	0.84

ID	ForManInt	plot	Mean_dbh	Mean_th	Mean_SLC
1	No_ForManInt	1	20.47	18.52	0.90
2	No_ForManInt	2	12.77	9.81	0.77
3	No_ForManInt	3	13.50	12.22	0.91
4	No_ForManInt	4	26.89	17.14	0.64
5	No_ForManInt	5	16.06	16.64	1.04
6	No_ForManInt	6	22.51	16.87	0.75

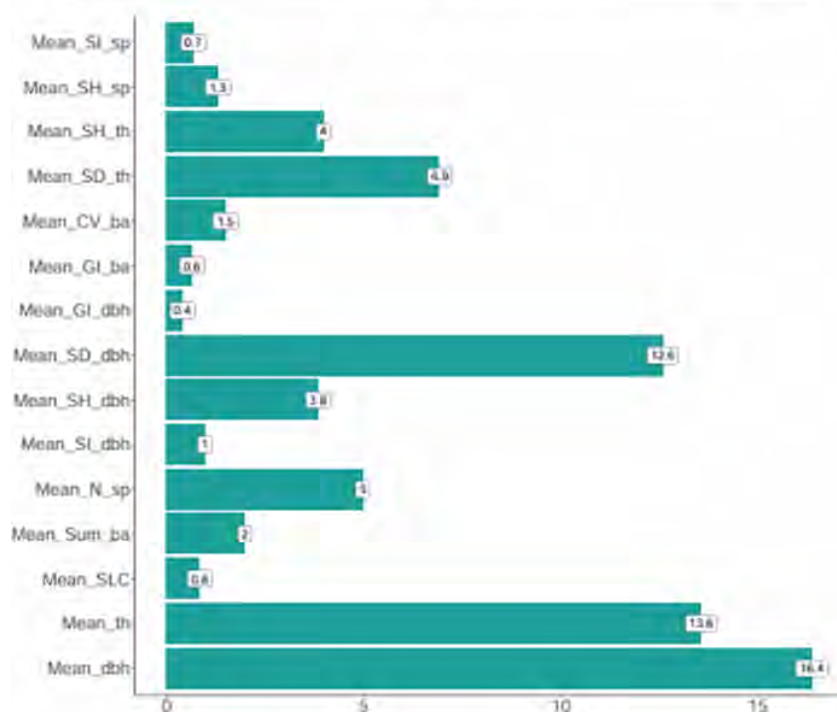


FIGURE 16 TABLE AND BAR-CHART EXTRACTED FROM THE PDF REPORT FOR FOREST DIVERSITY. THE TABLE SHOWS ONLY THE FIRST 6 COLUMNS OF THE OUTPUT BUT ALL INDICES COMPUTED BY THE WEB-APP ARE REPORTED IN THE .CSV FILES.

Conclusion

The CSF-Ind is a web-app useful to estimate the CSF indicators. Through a user-friendly interface, it is possible to compute the growing stock and the carbon stock, indicators that forest managers generally use to describe the forest. The possibility of extracting a summarised table (.csv), or a report (.pdf), could be useful for forest managers to describe the forest in different periods, or also under different forest management systems. CSF-Ind is especially useful in the calculation of forest structure indices that are used less by forest practitioners, i.e., Shannon, Simpson and Gini index for tree species composition diameter/tree height distribution. Additionally, the computed indicators can be used as input data to the CSF-AT web-app, which is especially useful for forest diversity indices which are not easy and intuitive to compute. Although the current version of CSF-Ind web-app was developed and tested with Italian allometric equations, it can be improved through the introduction of equations for other EU countries allowing to calculate indicators accurately at the European level.

2.5.2. Practical application of the CSF-AT: the Collemeluccio case study

Background

Climate change is increasingly affecting forest ecosystems, threatening their structure, functioning, and the ecosystem services they provide (Allen et al., 2010; Lindner et al., 2010). Mediterranean mountain regions, in particular, face amplified pressures due to drought, pests, and fire (Peñuelas and Sardans, 2021). At the same time, forests are expected to contribute to carbon sequestration while sustaining socio-economic benefits, creating a dual challenge for forest management (Cooper and MacFarlane, 2023; Tognetti et al., 2022; Nabuurs et al., 2018).

Climate-Smart Forestry (CSF) offers a basis to integrate climate change adaptation, mitigation, and social dimension in forest management (Bowditch et al., 2020). While conceptual frameworks and national-scale applications exist (Mathys et al., 2021), practical tools for evaluating forest smartness at the stand level are limited. The CSF Assessment Tool (CSF-AT) provides a replicable, user-friendly method to quantify forest smartness based on a weighted set of indicators and criteria, allowing integrated evaluation across the three CSF pillars.

This section shows and tests the CSF-AT web-app to the Collemeluccio forest (Molise, Italy), where two different management systems coexist: a long-term unmanaged area and an adjacent area managed through single-tree selection system aimed to promote regeneration and stand continuity of silver fir (*Abies alba* Mill.). The selection of the Collemeluccio forest as a case study provides an ideal “natural laboratory” to explore how forest management affects CSF performance. The case study area also lies within the Italian demo site of the FORWARDS project.

Additionally, data from silver fir forests in four other Italian regions were included to provide a broader comparative context.

Specifically, this practical application aims to:

1. Assess differences in climate-smart performance between managed and unmanaged areas.
2. Compare Collemeluccio results with other silver fir stands located in three Italian regions, in order to capture regional variability and provide a broader frame of reference.
3. Evaluate the usability of the CSF-AT app as a decision-support tool, by testing its capacity to translate field data into meaningful indicators and composite indices that guide forest managers decisions.

Method

Study area: Collemeluccio forest (Molise)

The Collemeluccio forest, located in the “Alto Molise” area of the Central Apennines (41°43′5.88″N, 14°21′3.96″E), represents one of the most important silver fir stands in central-southern Italy. The forest covers a total area of 450 ha, of which 363 ha belong to the “Collemeluccio-Montedimezzo Alto Molise” UNESCO Man and Biosphere (MaB) Reserve and are included in the European Union's Natura 2000 network (site code IT7222215). Silver fir dominates the forest, representing one of the southernmost natural populations of the species in Italy and a relict from the last post-glacial period (Antonucci et al., 2019). Frequently, fir is mixed with *Quercus cerris* L., *Fagus sylvatica* L., *Carpinus betulus* L., and *Ilex aquifolium* L., conferring high ecological value and making Collemeluccio one of the few areas in Italy where these species coexist within the same stand (Abbate, 1990). Historically, the forest has been subject to management interventions aimed at preserving silver fir and supporting its natural regeneration (Santopuoli et al., 2016).

Two experimental sites were analysed for testing the CSF-AT web-app. The first is located within the core area of the MaB Reserve, where harvesting activities ceased over 50 years ago and the forest has since been left to natural dynamics. The second lies outside the Reserve boundary and is regularly managed through a single-tree selection system, with the most recent harvest occurring in 2020–2021 (Figure 17). According to the regional forest type map (Garfi and Marchetti, 2011), both sites are classified as “Mesophilous Turkey oak forests with Silver fir”, described as uneven-aged high forest and corresponding to the habitat type “9510 – Southern Apennine silver fir forests” (European Commission, 1992).

Field data were collected during summer 2021. The sampling design consisted of a 1-ha north-oriented square plot, within which five circular subplots (radius 13 m) were established to measure forest structural parameters (Figure 17). Measurements included diameter at breast height (DBH, 1.3 m) for all trees with DBH ≥ 7.5 cm, total height, vitality status (alive or dead), volume of lying deadwood (diameter ≥ 10 cm, length ≥ 1 m), and of stumps (height ≥ 10 cm) (Marchetti and Lombardi, 2006). Within each circular plot, tree regeneration was assessed by recording species and height in a 26 m² transect (1 × 26 m) passing through the plot centre and oriented north–south.

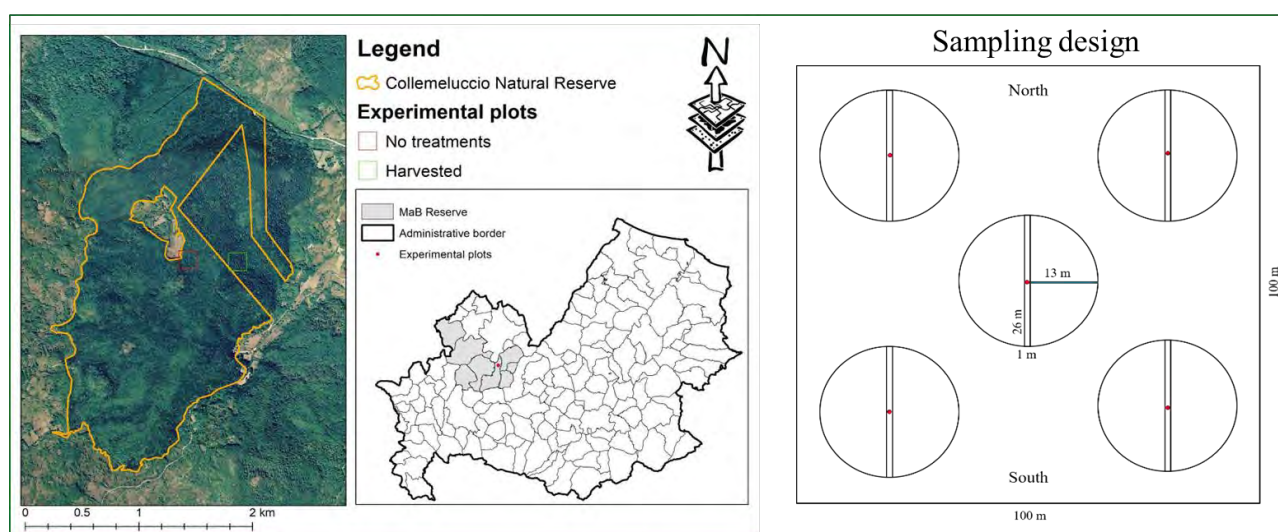


FIGURE 17 STUDY AREA. THE MAP ON THE LEFT DISPLAYS THE BOUNDARIES OF THE COLLEMELUCCIO FOREST AND THE TWO FIELD PLOTS. THE MAP IN THE CENTER SHOWS THE LOCATION OF THE STUDY AREA WITH ADMINISTRATIVE BOUNDARIES OF THE MUNICIPALITIES IN THE MOLISE REGION. THE MAB RESERVE IS HIGHLIGHTED IN LIGHT GRAY, WHILE THE RED POINT INDICATES THE STUDY AREA. THE BOX ON THE RIGHT SHOWS THE SAMPLING DESIGN OF THE FIELD SURVEY.

Comparative datasets from other Italian regions

To extend the analysis beyond Collemeluccio Forest, additional datasets were compiled using data from the Italian National Forest Inventory (INFC), which provides a systematic and representative overview of forest conditions across the country (Gasparini et al., 2022). The INFC sampling design is based on multiple circular plots, randomly located within a 1 km × 1 km systematic grid, and distributed across forest stands (Fattorini et al., 2006). Forests are classified according to a hierarchical system, with categories defined by dominant species or species groups in terms of coverage (Pignatti, 2003).

For this study, only plots belonging to silver fir forest types were selected in four regions of central and southern Italy—Molise, Abruzzo, Basilicata, and Calabria—where the species occurs naturally. This selection captures part of the ecological variability of the Apennine range.

Several plots were available in each region, allowing the calculation of mean indicator values representative of the silver fir stands. The selected data include both quantitative and qualitative attributes of the forest, such as growing stock, species composition, regeneration, deadwood, and carbon stock, ensuring consistency with the local dataset collected in Collemeluccio.

Preparation of input datasets and application of CSF-AT

Five datasets were prepared for analysis with the CSF-AT: two from Collemeluccio (managed and unmanaged) and four from the INFC corresponding to silver fir forests in Molise, Basilicata, Abruzzo, and Calabria regions. For Collemeluccio, seven field-based indicators were used: growing stock (GS), carbon stock (CS), slenderness coefficient (SLC), diameter distribution (DB), deadwood (DW), tree species composition (SP), and regeneration (RN). For the INFC datasets, the indicators included carbon stock (CS), forest damage (FD), growing stock (GS), tree species composition (SP), diameter distribution (DB), deadwood (DW), regeneration (RN), and increment and felling (IF).

All datasets were organized into CSV files formatted for direct use in the CSF-AT, ensuring consistency across local (field-based) and regional (inventory-based) sources.

Once uploaded, the CSF-AT calculated the composite CSF smartness index for each dataset. The calculation followed a hierarchical procedure:

1. Normalization: indicator values were rescaled using min–max scaling to ensure comparability, with values ranging from 0 (least desirable) to 1 (most desirable).
2. Weighted aggregation:
 - a. normalized indicators X_{norm} were multiplied by corresponding weights $W_{n,k}$ representing their importance within each CSF criterion k (i.e., mitigation, $k = 1$; adaptation, $k = 2$; socio-economic dimension, $k = 3$) resulting in a weighted indicator score $X_{n,k}$:

$$X_{n,k} = x_{norm} * w_{n,k} \quad (1)$$

- b. Criterion-specific sub-indices: weighted indicators belonging to the same criterion were summed to obtain sub-indices $I_{s,k}$:

$$I_{s,k} = \sum X_{n,k} \quad (2)$$

- c. Overall CSF index: the three sub-indices were combined using predefined criterion weights W_k to produce the overall smartness index I_{CSF} :

$$I_{CSF} = \sum I_{s,k} * W_k \quad (3)$$

The CSF-AT calculated plot-level and stand-level values of the composite indices for both Collemeluccio and the regional datasets.

Results were visualized through the interactive ternary plot, which shows the relative contributions of mitigation, adaptation, and social dimension for each dataset. The plot allows zooming to examine individual points in detail and provides a brief guide on how to interpret the graph. All results, including the ternary plot, can be exported as .csv or also as image, such as .png file for further analysis.

CSF-AT outputs

The CSF-AT provides multiple ways to explore and export the calculated smartness values. Results for each dataset are displayed at two levels: plot-level indices and stand-level averages for both Collemeluccio and the regional datasets.

An interactive ternary plot visualizes the relative contributions of the three CSF dimensions—mitigation, adaptation, and social dimension—allowing comparison across all datasets. Users can zoom in to examine individual points in detail and read the absolute value of the composite CSF index.

All outputs, including the ternary plot, can be downloaded in CSV and PNG formats, facilitating further analysis, reporting, or integration with other tools.

Results

The CSF-AT allowed the calculation of the composite smartness index and its three sub-indices—mitigation, adaptation, and social dimension—for the Collemeluccio experimental sites and the regional datasets (Molise, Calabria, Basilicata, and Abruzzo).

For the Collemeluccio forest, the managed site showed slightly higher smartness values compared to the unmanaged site (Table 3). The overall CSF index was 0.222 for the managed site and 0.214 for the unmanaged site. Differences were also observed in the three dimensions: mitigation scored 0.272 and 0.258, adaptation 0.294 and 0.284, and the social dimension 0.232 and 0.224 for managed and unmanaged sites, respectively. These results indicate a modest increase in smartness under active management, particularly in adaptation and mitigation.

At the regional scale, the smartness index varied across silver fir forests in central and southern Italy. Abruzzo exhibited the highest overall smartness (0.325), followed by Molise (0.300), Basilicata (0.290), and Calabria (0.261). The sub-indices reflected similar trends: mitigation ranged from 0.305 (Calabria) to 0.400 (Abruzzo), adaptation from 0.343 (Calabria) to 0.405 (Abruzzo), and the socio-economic dimension from 0.293 (Calabria) to 0.355 (Abruzzo).

TABLE 3 VALUES OF CSF COMPOSITE INDEX AND ITS THREE MAIN PILLARS (MITIGATION, ADAPTATION AND SOCIAL DIMENSION) FOR COLLEMELUCCIO EXPERIMENTAL SITES AND REGIONAL SILVER FIR FORESTS.

Experiment/Region	Smartness	Mitigation	Adaptation	Social dimension
Managed	0.222	0.272	0.294	0.232
Unmanaged	0.214	0.258	0.284	0.224
Molise	0.300	0.350	0.400	0.320
Calabria	0.261	0.305	0.343	0.293
Basilicata	0.290	0.345	0.370	0.305
Abruzzo	0.325	0.400	0.405	0.355

The ternary plot generated by the CSF-AT provides a visual representation of these results, showing the relative contributions of mitigation, adaptation, and social dimension for each dataset (Figure 18). Points are clustered close to the centre of the diagram, indicating generally balanced profiles across the three dimensions. Quantitatively, adaptation contributes the largest share in all cases (around 35–37% of the three-pillar total), mitigation is the second largest (around 32–34%), and the social dimension is the smallest (around 29–31%).

Local differences are small but informative: the managed and unmanaged Collemeluccio plots overlap closely, reflecting very similar relative profiles (the managed plot shows a marginally higher share of mitigation and adaptation). At the regional level, Abruzzo stands out with the highest overall smartness (see Table 3) and relatively large shares of mitigation and adaptation, while Calabria displays a somewhat higher

relative contribution of the socio-economic dimension compared with the other regions. Molise and Basilicata occupy intermediate positions.

The ternary diagram is therefore useful for a rapid visual comparison of how each case balances the three CSF pillars; for absolute sub-index values and composite scores, refer to Table 3.

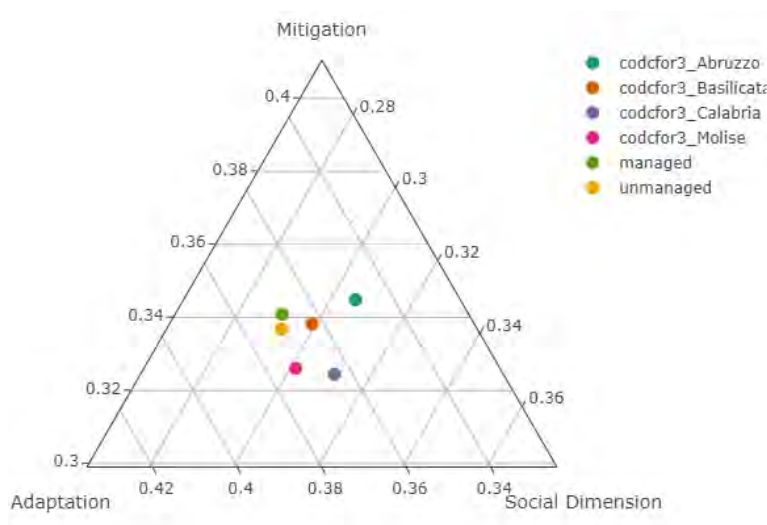


FIGURE 18 TERNARY PLOT SHOWING THE RELATIVE CONTRIBUTIONS OF MITIGATION, ADAPTATION, AND SOCIAL DIMENSION TO THE CSF INDEX FOR THE COLLEMELUCCIO EXPERIMENTAL SITES (MANAGED AND UNMANAGED) AND FOR REGIONAL SILVER FIR FORESTS (ABRUZZO, BASILICATA, CALABRIA).

Conclusions

The practical application of the CSF-AT web-app to the Collemeluccio case study and to regional silver-fir datasets (Molise, Basilicata, Abruzzo, Calabria) provided information for both site-level management and regional planning. The following points synthesize operational implications, methodological considerations, and recommended next steps.

Operational implications

- For local managers: the differences observed suggest that targeted management interventions (e.g., targeted selection or thinning) can marginally improve mitigation and adaptation components without compromising the social dimension. However, it is important to plan interventions with long time horizons and repeated monitoring to evaluate delayed effects.
- Regional context and background conditions tend to drive larger differences in climate-smart performance than single local contrasts; therefore, policy instruments and incentives should take into account local conditions rather than be applied uniformly.

Methodological strengths and future developments

Strengths

- The CSF-AT translates field and inventory data into an integrated composite index and sub-indices.
- It provides intuitive graphical outputs (ternary plot) and exportable data (CSV/PNG) suitable for reporting and further analysis.

Future developments

The tool could be further improved in several ways:

- Broader set of indicators: especially on the social dimension side, with the inclusion of aspects such as timber value, accessibility, or other ecosystem services. The web-app could also allow users to integrate their own indicators.
- Improved data management and performance: enabling the upload of multiple datasets at once, facilitating time series analysis, and optimising calculations for larger datasets with progress feedback.
- Weight sensitivity: adding a tool to explore how different weight settings affect results.
- Enhanced reporting: generating a final report that combines results from multiple datasets, with tables and figures available in different formats for easier use.

2.6. References

- Abbate, G.** (1990). *Le foreste della riserva MAB "Collemeluccio-Montedimezzo" (Molise, Italia meridionale)*.
- Alfieri, D., Caicoya, A. T., Tognetti, R., & Santopuoli, G.** (2025). Is climate-smart forestry affected by thinning intensity and tree species composition? A case study in long-term experimental plots of central Europe. *Journal of Environmental Management*, 394, 127603. <https://doi.org/10.1016/j.jenvman.2025.127603>
- Alfieri, D., Tognetti, R., & Santopuoli, G.** (2024). Exploring climate-smart forestry in Mediterranean forests through an innovative composite climate-smart index. *Journal of Environmental Management*, 368(July), 122002. <https://doi.org/10.1016/j.jenvman.2024.122002>
- Antonucci, S., Rossi, S., Lombardi, F., Marchetti, M., & Tognetti, R.** (2019). Influence of climatic factors on silver fir xylogenesis along the Italian Peninsula. *IAWA Journal*, 40(2), 259--S3.
- Barabesi, L., Franceschi, S., & Marcheselli, M.** (2012). Properties of design-based estimation under stratified spatial sampling with application to canopy coverage estimation. *The Annals of Applied Statistics*. 2012;6: 210–228.DOI: 10.1214/11-AOAS509
- Baumer, B., & Udwin, D.** (2015). R markdown. *Wiley Interdisciplinary Reviews: Computational Statistics*, 7(3), 167-177.
- Bowditch, E., Santopuoli, G., Binder, F., Del Rio, M., La Porta, N., Kluvankova, T., Lesinski, J., Motta, R., Pach, M., Panzacchi, P., & others.** (2020). What is Climate-Smart Forestry? A definition from a multinational collaborative process focused on mountain regions of Europe. *Ecosystem Services*, 43, 101113.
- Cooper, L., & MacFarlane, D.** (2023). Climate-Smart Forestry: Promise and risks for forests, society, and climate. *PLOS Climate*, 2(6), 1–26. <https://doi.org/10.1371/journal.pclm.0000212>
- Duncker, P. S., Barreiro, S. M., Hengeveld, G. M., Lind, T., Mason, W. L., Ambrozy, S., & Spiecker, H.** (2012). Classification of forest management approaches: a new conceptual framework and its applicability to European forestry. *Ecology and Society*, 17(4).
- European Commission**, 1992. Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. Off. J. Eur. Communities L206: 7-50.
- Garfi, V., Marchetti, M., & others.** (2011). *Tipi forestali e preforestali della regione Molise*. Edizioni dell'Orso Srl.
- Gasparini, P., Di Cosmo, L., Floris, A., & De Laurentis, D.** (2022). *Italian National Forest Inventory—Methods and Results of the Third Survey: Inventario Nazionale delle Foreste e dei Serbatoi Forestali di Carbonio—Metodi e Risultati della Terza Indagine*. Springer Nature. <https://doi.org/10.1007/978-3-030-98678-0>
- Lindner, M., Maroschek, M., Netherer, S., Kremer, A., Barbati, A., Garcia-Gonzalo, J., Seidl, R., Delzon, S., Corona, P., Kolström, M., Lexer, M. J., & Marchetti, M.** (2010). Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems. *Forest Ecology and Management*, 259(4), 698–709. <https://doi.org/https://doi.org/10.1016/j.foreco.2009.09.023>
- Lombardi, F., Lasserre, B., Tognetti, R., & Marchetti, M.** (2008). Deadwood in relation to stand management and forest type in Central Apennines (Molise, Italy). *Ecosystems*, 11(6), 882-894. <https://doi.org/10.1007/s10021-008-9167-7>

- Marchetti, M., & Lombardi, F.** (2006). Quali-quantitative analysis of deadwood in unmanaged stand: " Bosco Pennataro", Alto Molise. *L'Italia Forestale e Montana*, 61(4), 275–302.
- Mathys, A. S., Bottero, A., Stadelmann, G., Thürig, E., Ferretti, M., & Temperli, C.** (2021). Presenting a climate-smart forestry evaluation framework based on national forest inventories. *Ecological Indicators*, 133, 108459. <https://doi.org/10.1016/j.ecolind.2021.108459>
- Nabuurs, G.-J., Verkerk, P. J., Schelhaas, M.-J., González Olabarria, J. R., Trasobares, A., & Cienciala, E.** (2018). *Climate-Smart Forestry: mitigation impacts in three European regions* (Vol. 6). EFI Helsinki, Finland.
- Peñuelas, J., & Sardans, J.** (2021). Global Change and Forest Disturbances in the Mediterranean Basin: Breakthroughs, Knowledge Gaps, and Recommendations. *Forests*, 12(5). <https://doi.org/10.3390/f12050603>
- Pignatti, S.** (2003). Guida alla classificazione della vegetazione forestale. Inventario Nazionale delle Foreste e dei Serbatoi Forestali di Carbonio. *Istituto Sperimentale per l'Assestamento Forestale e per l'Alpicoltura*.
- Puletti, N., Canullo, R., Mattioli, W., Gawryś, R., Corona, P., & Czerepko, J.** (2019). A dataset of forest volume deadwood estimates for Europe. *Annals of Forest Science*, 76(3), 68.
- R Core Team.** (2024). *R: A Language and Environment for Statistical Computing*. <https://www.r-project.org/>
- Santopuoli, G., Lasserre, B., Di Martino, P., & Marchetti, M.** (2016). Dynamics of the silver fir (*Abies alba* Mill.) natural regeneration in a mixed forest in the Central Apennine. *Plant Biosystems-An International Journal Dealing with All Aspects of Plant Biology*, 150(2), 217–226.
- Santopuoli G, di Cristofaro M, Kraus D, Schuck A, Lasserre B, Marchetti M** (2019). Biodiversity conservation and wood production in a Natura 2000 Mediterranean forest. A trade-off evaluation focused on the occurrence of microhabitats. *iForest* 12: 76-84. – doi: 10.3832/ifor2617-011
- Santopuoli, G., Temperli, C., Alberdi, I., Barbeito, I., Bosela, M., Bottero, A., Klopčič, M., Lesinski, J., Panzacchi, P., & Tognetti, R.** (2021). Pan-European sustainable forest management indicators for assessing Climate-Smart Forestry in Europe. *Canadian Journal of Forest Research*, 51(12), 1741–1750.
- SoEF.** (2020). State of Forests in Europe. https://www.wwfkenya.org/keep_kenya_breathing_/state_of_forest_in_kenya/
- Tabacchi, G., Di Cosmo, L. & Gasparini, P.** Aboveground tree volume and phytomass prediction equations for forest species in Italy. *Eur J Forest Res* 130, 911–934 (2011). <https://doi.org/10.1007/s10342-011-0481-9>
- Tognetti, R., Smith, M., & Panzacchi, P.** (2022). *Climate-smart forestry in mountain regions*. Springer Nature. <https://doi.org/10.1007/978-3-030-80767-2>
- Wijewardana, D.** (2008). Criteria and indicators for sustainable forest management: The road travelled and the way ahead. *Ecological Indicators*, 8(2), 115–122. <https://doi.org/10.1016/j.ecolind.2006.11.003>
- Winston, C., Joe, C., JJ, A., Carson, S., Barret, S., Yihui, X., Jeff, A., Jonathan, M., Alan, D., & Barbara, B.** (2012). *shiny: Web Application Framework for R*. The R Foundation. <https://doi.org/10.32614/cran.package.shiny>

3. Data assimilation routines for spatially explicit data on forest structure and composition to enable forward-looking projections

3.1. Introduction

3.1.1. Motivation

This chapter focuses on a data assimilation routine that was developed to allow for the use of European-wide, spatially explicit data on forest structure and composition in forward-looking modelling. We focused on the specific case of EFISCEN-Space in this chapter, but the developed approach can be useful for other models and purposes as well. EFISCEN-Space is a geographically explicit model of forest development in Europe at the forest stand level (Schelhaas et al., 2022B). The primary way to initialize EFISCEN-Space is with tree-level national forest inventory (NFI) data obtained from EU-countries, but it can run on any inventory data source that provides data at tree-level, such as district-level inventories or forest management planning data. A requirement is that the forest inventory data contains measurements of individual trees which, for EFISCEN-Space, are translated into a per-species diameter distribution of number of trees per hectare (Figure 19).

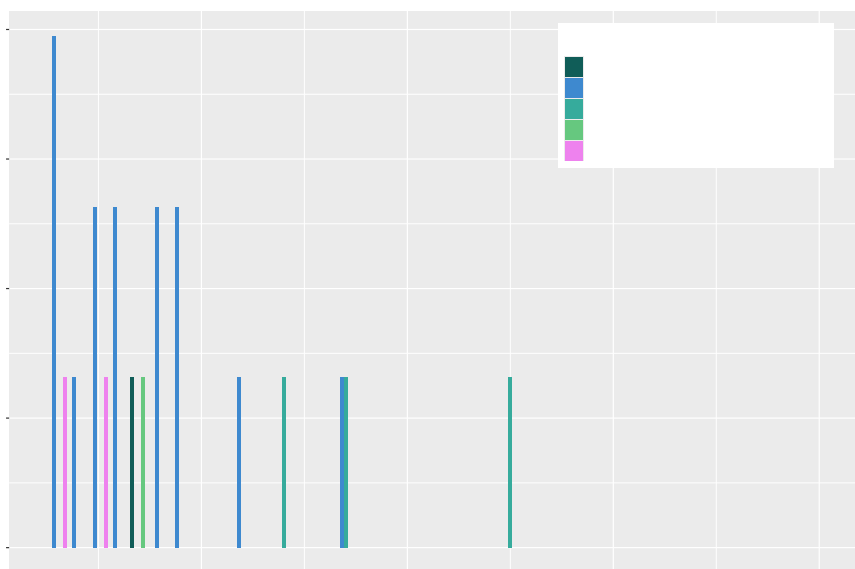


FIGURE 19 FOREST INVENTORY PLOTS ARE PRESENTED TO EFISCEN-SPACE AS PER-SPECIES DIAMETER DISTRIBUTIONS OF THE STEM NUMBER PER HECTARE. THE EXAMPLE SHOWS THE TREEWISE DISTRIBUTION OF AN NFI PLOT IN THE NETHERLANDS.

All plots within an inventory are representative of a certain area, depending on the sampling strategy (e.g., regular grid sampling, random sampling, stratified sampling). The spatial density of NFI sample plots varies between countries, and sometimes between regions within a country. These densities are chosen in such a way that NFIs can provide reliable information at regional to national scale, with a sufficient number of sample plots for the underlying strata. When zooming in to smaller spatial units, the uncertainty of the

estimates becomes larger as the number of sample plots decreases. The same principles apply to EFISCEN-Space, as it essentially projects the development of the NFI sample plots into the future. Reliable projections at sub-national or local scale are therefore only possible if the number of sample plots for the region under consideration can be increased. Since it is not feasible to carry out new inventories for every regional study, a method to artificially increase the sampling density would be needed. Another problem in making projections with EFISCEN-Space for Europe is that not all countries have NFI data or similar, or they do not share such data. Many remote sensing products exist that provide forest-related information at the European scale (e.g., Pucher et al., 2022; Miettinen et al., 2024; De Keersmaecker et al., 2025), but none of them provide the (tree-level) detail as needed to initialise EFISCEN-Space. However, these remote-sensing products could potentially be combined with the tree-level inventory data from other countries to create “synthetic” NFI plots in the country of interest. In this chapter we therefore explore the possibilities of creating synthetic NFI datasets to 1) increase the sampling densities at sub-national scale where national NFI data are available and 2) cover countries where no NFI data are available.

To handle these types of limitations, this chapter describes a methodology to generate sets of so-called “synthetic” tree-level forest inventory plots, based on whichever data sources are available to be used (Aussenac et al., 2023). This chapter explores the choices and requirements of the input data, the data preparation and processing, as well as possible validation options. In addition, the performance of the proposed methodology is evaluated through several case studies. It is important to note that, even though we focus on the specific case of EFISCEN-Space in this chapter, the developed approach can be useful for other models and purposes as well.

3.1.2. Synthetic data concept

Synthetic plots are here defined as plots with synthesised tree-level forest information which are copied from an existing inventory plot and placed at locations where only aggregated plot-level information is available. To create a synthetic dataset, two different types of input data are required: *tree-level* and *plot-level* data. The plot-level data contain aggregated information on the state of the forest at a particular location. This could be a summarised inventory plot or a grid cell from a remote sensing interpretation (e.g., a map of basal area). The plot-level data must be from *within* the area of interest, with an exact location. The tree-level data must contain the same summarised information, so a match can be made between the data sets, based on similarity (Figure 20). On top of that, these tree-level data contain tree-wise diameter and species information which can be copied to the synthetic plot. The tree-level data can be from *outside* the area of interest and do not need exact location information.

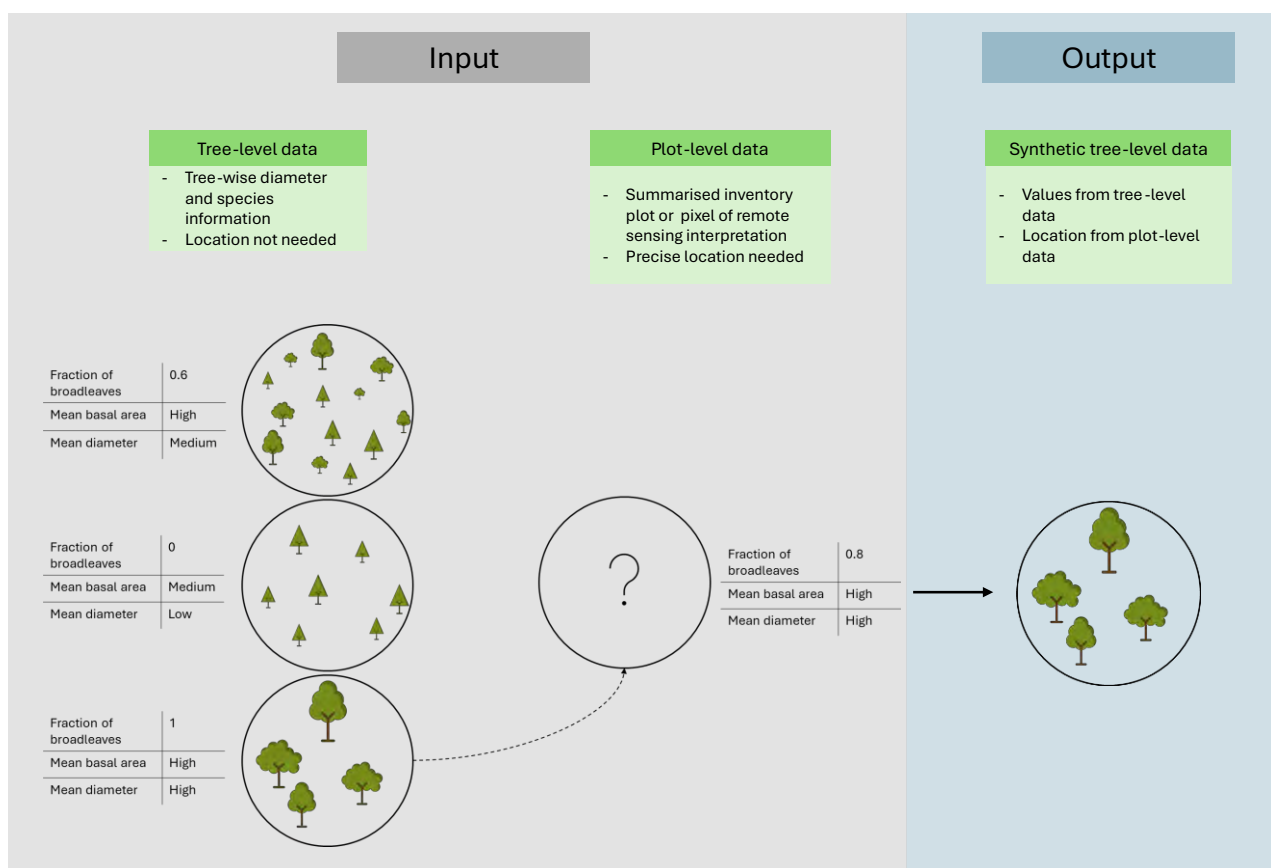


FIGURE 20 CONCEPTUAL DIAGRAM SHOWING HOW A SYNTHETIC INVENTORY PLOT IS CREATED. PLOT-LEVEL DATA (E.G. SUMMARISED INVENTORY DATA OR PIXEL FROM A REMOTE SENSING-BASED MAP), WHICH ON ITS OWN CANNOT BE USED TO INITIALISE EFISCEN-SPACE IS MATCHED TO TREE-LEVEL DATA FROM A TREE-WISE FOREST INVENTORY (E.G. NFI OR MANAGEMENT INVENTORY). A MATCH IS MADE BASED ON SIMILARITY BETWEEN SUMMARY STATISTICS OF A SELECTION OF PLOT CHARACTERISTICS.

3.1.3. Overall approach

3.1.3.1. Case studies

The availability and quality of both tree- and plot-level data can vary between countries. Hence, the choice of input data needs to be evaluated on a case-by-case basis. The aim of the case studies is to explore what kind of results can be expected from different combinations of data sources in different biogeographical settings. This could then support decisions on an approach in real applications. We selected three countries with differing forests, climate and topography: Finland, The Netherlands, and Italy (Table 4).

TABLE 4: CASE STUDY COUNTRIES AND THE RESPECTIVE REASONS FOR INCLUSION.

Case study country	Reason for inclusion
Netherlands	Heterogeneous forests, intermediate number of tree species, very little relief, temperate and oceanic climate
Finland	Homogeneous forests, few tree species, little relief, cold climate
Italy	Heterogeneous forests, very high number of tree species, great biogeographic heterogeneity, temperate and arid climate

Table 5 show different sources of plot- and tree-level data that can be used. The data sources are arranged in a matrix that explains which combinations have been tested in the case studies that will be presented in this chapter.

TABLE 5 MATRIX OF DATA SOURCES USED IN THIS STUDY TO TEST AND CREATE SETS OF SYNTHETIC TREE-LEVEL INFORMATION, WITH CASE STUDIES AT THEIR INTERSECTIONS. THE FIRST COLUMN SHOWS SOURCES OF PLOT-LEVEL DATA, THE FIRST ROW SOURCES OF TREE-LEVEL DATA. THE NUMBER IN PARENTHESES REFERS TO THE RESPECTIVE EXPERIMENT WITHIN A CASE STUDY. IF MULTIPLE DATA SOURCES WERE USED WITHIN THE SAME EXPERIMENT, THE NUMBER CAN OCCUR MORE THAN ONCE IN THIS TABLE (E.G., FINLAND (FI-1)).

Plot-level data		Tree-level data	
		from within the area of interest	from outside the area of interest
	from inventory data	Finland (FI-1) Netherlands (NL-1)	Finland (FI-3) Netherlands (NL-4)
	from national remote-sensing products	Finland (FI-1) Netherlands (NL-2) Italy (IT-2, IT-3)	Finland (FI-3)
	from European remote-sensing products	Finland (FI-2) Italy (IT-1, IT-3) Netherlands (NL-3)	

The following questions will be addressed:

1. Which inputs are essential in the procedure of creating synthetic plots?
2. How well can we do with available national data sets?
3. How well can we do with European maps?
4. Can we use tree-level data from neighbouring countries?
5. Does the inclusion of abiotic variables improve the results?
6. How do the synthetic data sets influence the results of EFISCEN-Space, compared to model runs with original inventory data?

3.1.3.2. Input variables

Because the main two inputs to EFISCEN-Space relate to the diameter distribution and species composition, the main focus here is to optimise the assignment of basal area, (quadratic) mean diameter, stem density (i.e., stems per hectare) and dominant tree species or species group. Other variables, for example, volume, biomass or height, can be used as well, if they provide additional information on the plot structure. In those cases where national datasets are not available or not sufficient, we used available European maps that were developed within the Horizon-Europe projects PathFinder and ForestPaths. The PathFinder project developed European forest maps of volume, above-ground biomass and deciduous-coniferous proportion, with a resolution of 10 x 10m for the year 2020. The maps were created using a combination of Sentinel 2 mosaic, Copernicus layers, and National Forest Inventory datasets. More detailed information in Miettinen et al. (2024). The ForestPaths project developed a European forest map of tree genus distribution, with a resolution of 10 x 10m. The map includes information about the following tree genus: Larix, Picea, Pinus, Fagus, Quercus, other needleleaf, and other broadleaf (De Keersmaecker et al., 2025).

3.1.3.3. Data preparation

After the selection of available and relevant forest variables, the preparation of the input data is important to ensure compatibility of the input datasets. It is likely that the selected variables are measured in different units and scales. Therefore, they should be scaled to ensure that they all have the same weight in the distance measure that is used to find the most similar plots. If the variables follow a normal distribution they can be scaled between 0 and 1. In case of large outliers or non-normal distribution, quantiles can be applied to improve the results of the distance analysis. Here, the percentile bounds were chosen visually but for future improvement a statistical method for percentile selection could be implemented. The levels of categorical variables should be the same in the tree-level and plot-level data.

Data from remote-sensing layers are extracted based on a pre-defined grid, e.g., NFI plot pattern or EEA grid, unless the original resolution of the remote-sensing layers is kept. Once the locations for synthetic plots have been chosen, they are overlaid on the map layers and the underlying data are extracted. Data extraction from remote-sensing layers has been tested in different ways: i) synthetic plot as point location and value extraction of grid cell that point falls into, ii) synthetic plot as polygon and value extraction as mean of grid cells that polygon falls into, iii) synthetic plot as polygon and value extraction as weighted mean of grid cells that polygon falls into. The different options showed no significant difference in the results, however using weighted means increased the processing time considerably. Therefore, i) and ii) are the recommended options.

3.1.3.4. Nearest distance analysis

Each synthetic plot now has variable values that were extracted from the plot-level data. The distance to each point in the tree-level dataset is calculated by a pairwise comparison of the variables, using Gower's distance. The tree-level plot that is most similar according to this metric, is placed at the location of the synthetic plot. The Gower's distance is a coefficient that measures the similarity between pairs of variables. Variables can be dichotomous, qualitative and quantitative. If pairs of variables are considered different, a score of 0 is assigned; if pairs of variables are considered to have some degree of similarity, a positive score up to 1 is assigned, where 1 means that the pair of variables is equal. In the case of qualitative variables, the score is 0 if the two variables differ, and the score is 1 if they are the same. In the case of quantitative variables, the score is 1 if the two variables are the same, the score is 0 if the two variables are at opposite ends of their range, and the score is between 0 and 1 for intermediate values (Gower, 1971). We chose the Gower's distance, because we have both quantitative (dbh, height, etc.) and qualitative (dominant tree species) variables and the Gower's distance allows the distance calculation based on both types of data. We compute the Gower's distance with the R package 'gower', version 1.0.1 (van der Loo, 2022), using the 'gower_topn' function.

3.1.3.5. Evaluation & Validation

The results of each experiment are evaluated by calculating a number of metrics based on the available validation data (independent data where possible) and the synthetic value of each variable. These metrics include RMSE, RMSE% (RMSE expressed as percentage of the mean), and bias for numeric variables (Equations 1,2, and 3). For categorical variables overall accuracy and Cohen's Kappa – a metric that corrects by agreement occurring by chance – were used (Equations 4 and 5). In addition, we provide visual evaluation of the results by plotting original against synthetic data in scatterplots and density plots, and categorical variables, i.e. species or dominant species distribution, in pie charts. However, pie charts, for example, have shown to display good performance while accuracy values show a different picture, therefore it is advisable to prioritize the metrics results over the visual evaluation. In some cases (e.g. Finnish experiment FI-1 and FI-2) it is possible to compare the original and the synthetic data with official regional data to evaluate the accuracy of the created datasets compared to recent measurements.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad 1$$

$$RMSE\% = \frac{RMSE}{\mu} \times 100 \quad 2$$

$$Bias = \frac{\sum (y_i - \hat{y}_i)}{n} \quad 3$$

$$Overall\ Accuracy = \frac{Number\ of\ correctly\ classified\ plots}{Total\ number\ of\ classified\ plots} \quad 4$$

$$\text{Cohen's Kappa} = \frac{\text{Observed Agreement} - \text{Expected Agreement}}{1 - \text{Expected Agreement}}$$

5

To validate the developed approach for creating synthetic inventory data we test its performance with EFISCEN-Space. For this, we run test scenarios with the original dataset and the respective synthetic datasets in EFISCEN-Space and compare the simulation results. Scenarios are run for the case study country Italy (see case study 3 on Page 95).

The following software was used for the case studies: R (version > 4.3.3); R-Studio (version > 09-2024), QGIS (version > 3.36), and ArcGIS Pro (version > 3.4.0).

3.2. Case study demonstrations

3.2.1. Case study 1: Netherlands

Background

Approximately 11% (370 000 ha) of the Netherlands is covered by forest. 7 cycles of NFIs have been carried out since 1938. According to the latest NFI (NFI-7; Schelhaas et al., 2022A), describing the state of the forest between 2017-2022, more than half of the forest contained a mix of species while about 19% had only one species. About half of the surveyed forest plots were eligible for subsidy intended for forests with a production function, 36% were eligible for subsidy intended for forests with a clear natural component. In addition to the diverse appearance of the forests, the availability of NFI data with exact plot locations, species and structure maps, and an independent forest inventory data set make the Netherlands an interesting case study. The synthetic plots are generated for two provinces (NUTS-2): Limburg and Noord-Brabant.

The procedure described above for generating synthetic forest inventory data is tested and validated in the following experiments:

Experiment 1: Which inputs are essential in the procedure of creating synthetic plots (given perfect data)?

Experiment 2: How well can we do with available national datasets?

Experiment 3: How well can we do with available pan-European datasets?

Experiment 4: Can we use tree-level data from another region?

Available input data and preprocessing

Tree-level inventory data: Dutch NFI-7 data

The Dutch NFI-7 reports measurements of individual trees in 3197 plots from an unaligned systematic sample (Figure 21), of which 1413 are repeated measurements of NFI-6. At plot level, among other variables dominant tree species (>60% of the basal area), basal area, stem density, dominant height, and volume living and dead (standing and lying) wood are reported, together with estimates of crown and understory cover, vitality and rejuvenation. The density of the NFI is about one plot per 100 hectares.

Plots with missing location information were removed, as well as plots with the label “not visited/deforested”. Apart from plot name and coordinates the variables basal area, stem density, volume, above ground biomass, dominant height and dominant tree species were selected, because these were also available as map layers. Additionally, quadratic mean diameter (qmd) was calculated from basal area and stem density using:

$$\text{quadratic mean diameter [cm]} = \text{sqrt} \left(\frac{\text{basal area [m}^2 \text{ h}^{-1}]}{0.00007854 * \text{stem density [h}^{-1}]} \right)$$

6

The variables ‘group’ and ‘type’ were added. In ‘group’ the dominant tree species were grouped in six classes that the European genus map (introduced below) contains: Larix sp., Pinus sp., Fagus sp., Quercus sp., other conifer species and other broadleaf species. In ‘type’ the dominant tree species were grouped in broadleaf and conifer classes. Subsequently, the numeric layers were scaled using quantiles. The lowest observed value

was set to 0 while the value that was set to 1 was the 0.98/0.99/1 percentile, depending on the distribution of the variable (e.g., 0.98 was used on data with a very long tail). Per value the following percentile bounds were used: basal area (0; 0.98), stem density (0; 0.98), quadratic mean diameter (0; 0.99), volume (0; 0.99), above ground biomass (0; 0.98), dominant height (0; 1). Here, the percentile bounds were chosen visually but for future improvement a statistical method for percentile selection could be implemented.

Plot-level inventory data: Dutch SyHI data

SyHI (systeem voor houtmeetkundige inventarisatie or Woodstock; *system for wood measurement inventory* in English) is an inventory performed on request of forest owners for (parts of) their property. For this project, only SyHI data collected on state forest service land, within the study area of the provinces Noord Brabant and Limburg, were used (Probos, unpublished; Figure 21). State forest service areas make up about one quarter of the total Dutch forest area. Because the inventory is performed at the property level, the spatial density of the inventory is, locally, much higher than that of the NFI (1 plot per 1-3 ha versus 1 plot per 100 ha in the NFI) but it does not cover the whole country evenly. Apart from plot name and coordinates, basal area, stem density, volume, and dominant tree species were selected. Quadratic mean diameter was calculated in the same way as for the NFI data (Equation 6). Within that area, 2600 plots inventoried between 2017 and 2022, were selected so they would temporally correspond to the structure maps. The measurements were scaled using the same lower and upper bounds that were used to scale the NFI-7 variables. The variables 'group' and 'type' were added based on the dominant species, as was done for the NFI-7 data.

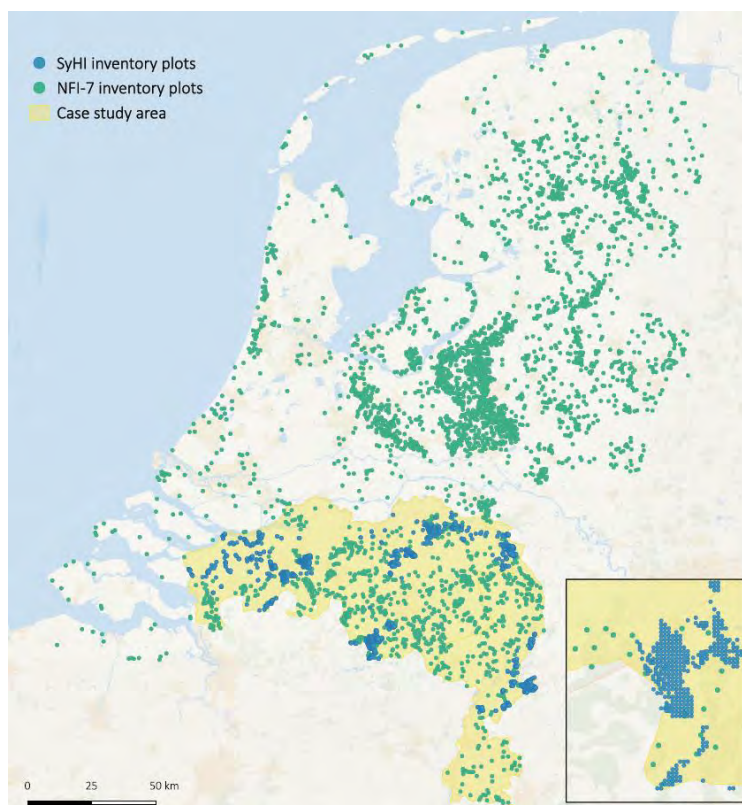


FIGURE 21 MAP OF THE NETHERLANDS, WITH THE CASE-STUDY AREA HIGHLIGHTED IN YELLOW. TWO SOURCES OF FIELD-INVENTORY DATA THAT WERE USED ARE ALSO DISPLAYED.

National structure maps

Structure maps with a 10 x 10m spatial resolution were created within this project, using NFI-7 data as dependent variables and airborne LiDAR-derived data layers as predictors in random forest (RF) models (Francini, unpublished A). The maps covered the provinces Limburg and Noord-Brabant in the south of the Netherlands. These included map layers of basal area, quadratic mean diameter, stem density, dominant height, volume and volume increment. The map values were scaled using the lower and upper bounds that were used to scale the NFI-7 and SyHI data. Note that these data are not independent of the NFI-7 data.

European genus map (ForestPaths)

See section 3.1.3.2

European structure maps (PathFinder)

See section 3.1.3.2

Soil map

The soil map that is used within the Dutch LULUCF was used in *experiment 1*. This raster map with a resolution of 25 x 25m dates from 2014 and contains 12 soil classes (Lesschen et al., 2012).

German NFI-3 data

Germany has performed four national forest inventories (Bundeswaldinventur; BWI) since 1986. The data from the third NFI (BWI 2012) were used as the tree-level data in *experiment 4* (Riedel et al., 2017). The variables basal area, qmd, stem density, and volume were used.

Area of interest boundary

The boundaries of the study area corresponded to the two Dutch provinces Limburg and Noord-Brabant. The outlines were retrieved from gadm.org using the R package *rspatial/geodata* (Hijmans et al., 2024; v0.6-2).

Forest mask

All input rasters were masked to the forest area of the Dutch LULUCF 2021 land use map, which is also used to define forest area in the NFI (Kramer and Los, 2022).

Comparison of data sources

The SyHI and NFI data are similarly distributed, while the national structure maps have a much narrower distribution (Figure 22A). The SyHI data have a slightly lower mean for basal area, quadratic mean diameter and volume than the NFI and national maps. The European structure maps (Miettinen et al., 2024) have a much wider distribution of the variables volume and AGB, showing a considerably higher mean than the other data sets (Figure 22B).

The species distribution of the NFI and SyHI data are quite similar, although the NFI data has slightly more beech, larch and other conifers while the SyHI data has more oak and pine (Figure 22C). The European genus map (de Keersmaecker et al., 2025) has much less oak than NFI and SyHI, but more “other broadleaf”.

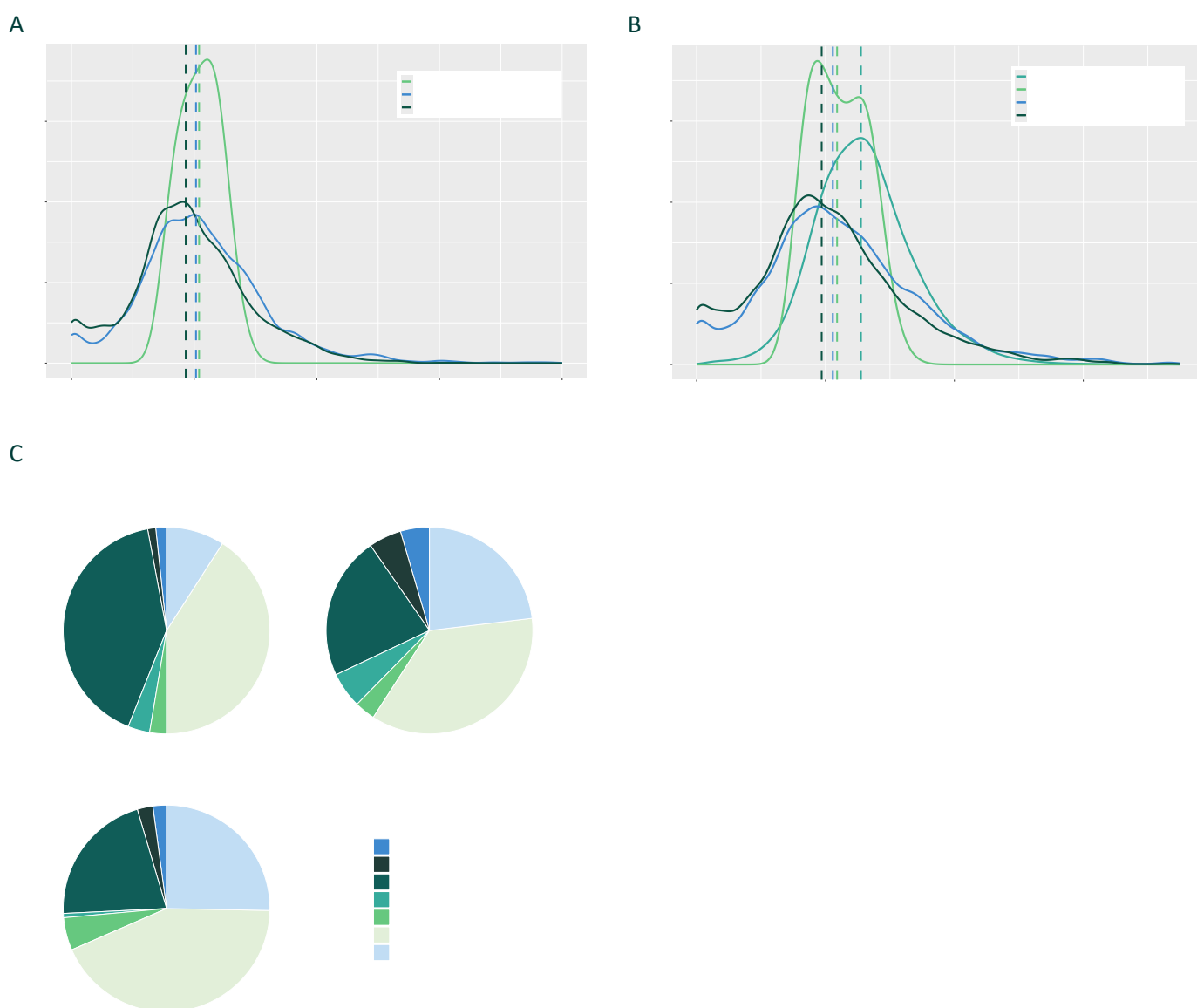


FIGURE 22 COMPARISON OF THE DATA SETS USED FOR THE DUTCH CASE STUDY. A) BASAL AREA, B) STEM VOLUME, C) DOMINANT SPECIES GROUP. X-AXES HAST BEEN TRUNCATED FOR THE VISUALISATION.

Experiment NL-1: Which inputs are essential (given perfect datasets)?

Description

The goal of experiment NL-1 is to test how well plot-level inventory data can be approximated using the nearest distance method, given perfect plot-level data and local tree-level inventory data. NFI-7 plots are placed at locations of SyHI plots, based on the similarity in their summary statistics.

Data

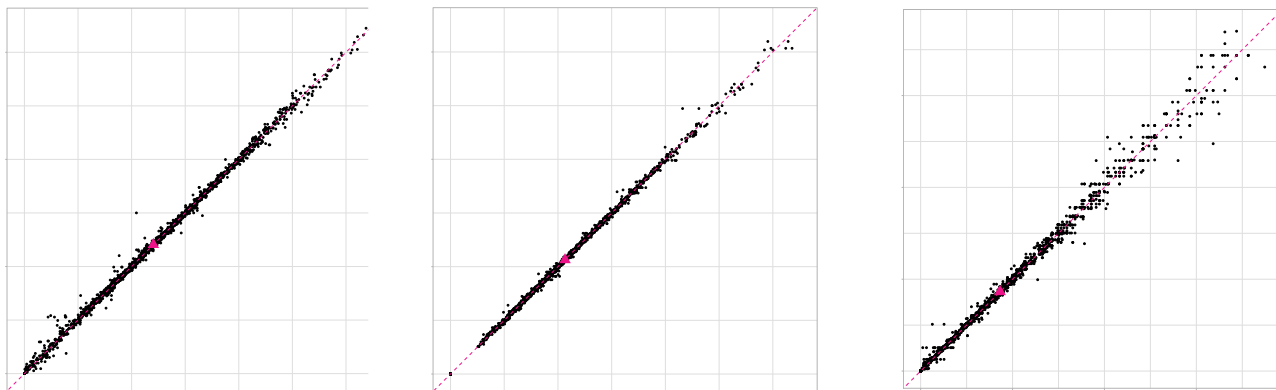
- **NFI-7:** tree-level inventory data
- **SyHI:** plot level inventory and validation data
- **Soil:** raster data extracted to the SyHI inventory points

Results

The synthetic data set has a near-perfect match for diameter and species related variables (Figure 23A). Using two variables out of basal area, qmd, and stem density gives a good result (%RMSE < 10) for the third one because of the exact relationship between them. Adding any third variable adds very little new information to the synthetic data.

In the absence of basal area data, volume is a good alternative constraining variable. Using volume decreases the RMSE% of the synthetic basal area from ~60% to ~16%, compared to ~3% when really using basal area. Furthermore, using volume instead of basal area decreases the RMSE% for synthetic stem density from ~80% to ~26%, compared to ~10% when really using basal area (Figure 23B).

A



B

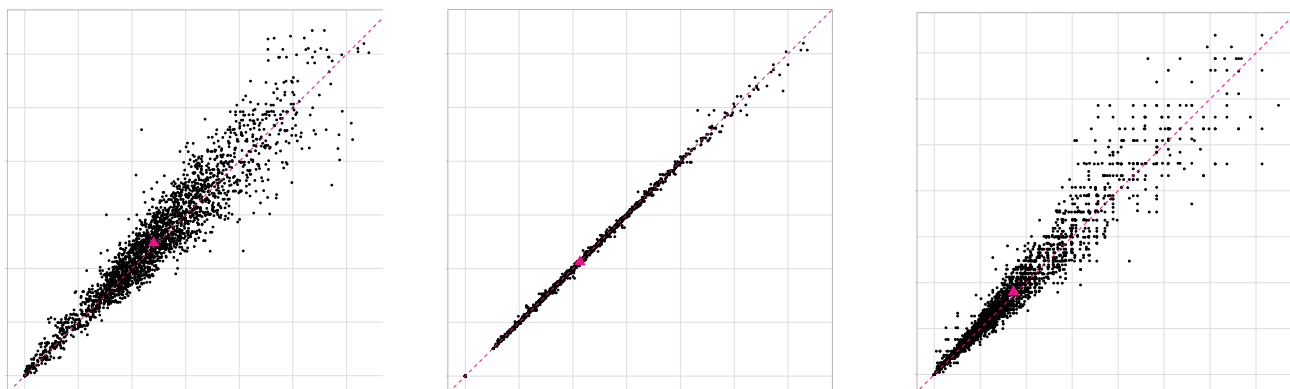


FIGURE 23 SYNTHETIC PLOTS DERIVED FROM TREE-LEVEL NFI-7 DATA PLOTTED AGAINST THE ORIGINAL SYHI DATA FOR BASAL AREA, QMD, STEM DENSITY. A) THE SYNTHETIC DATA WERE ALLOCATED USING BASAL AREA AND QMD; B) USING VOLUME AND QMD.

Using broadleaf-conifer data to create synthetic data improves the broadleaf or conifer match from an overall accuracy of 0.5 and Kappa score of 0.1 to an overall accuracy of 1 and Kappa score of 1 without compromising on the accuracy of the synthetic diameter-related variables. The synthetic dominant species group is also improved, from overall accuracy of 0.3 to 0.6 and Kappa score 0.1 to 0.4. Using the dominant species group further improves the overall accuracy of synthetic dominant species group to 1 with a Kappa score of 0.9 (Figure 24).

Finally, using soil type instead of species groups to create synthetic data is only slightly better than not using any species information. Using soil data brings the overall accuracy of broadleaf or conifer from 0.5 to 0.6 and dominant species group from 0.3 to 0.4. However, it makes the accuracy of the synthetic structure variables slightly worse. Overall, it is not a very useful addition.

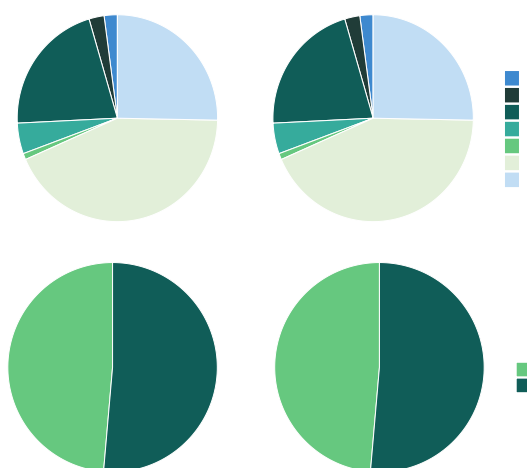


FIGURE 24 SPECIES DISTRIBUTION OF THE SYNTHETIC DATA DERIVED FROM NFI-7 DATA COMPARED TO THE ORIGINAL SYHI PLOTS. THE SYNTHETIC DATA WERE ALLOCATED USING THE DOMINANT SPECIES GROUP.

Conclusions and takeaway for the following experiments

Only two variables out of basal area, stem density and quadratic mean diameter are needed to create a synthetic data set when using high quality data. Furthermore, volume, which is more often available in maps, can be used as a good alternative for basal area. Broadleaf-conifer or species group information is needed for a good estimate of species distribution. Soil information does not offer a good alternative.

Experiment NL-2: Given the best data we have, how well can we do?

Description

The goal of experiment NL-2 is to use the takeaways from experiment NL-1 to test the synthetic data method in a setting where perfect data is not available. The best combinations of variables from experiment NL-1 are tested again (i.e., two out of three diameter variables and dominant species groups). Structure maps that were made specifically for the Netherlands (see National structure maps) are used in combinations with a European genus map (see 3.1.3.2). Data is extracted from these maps at each SyHI plot location. NFI-7 plots are matched to these map extracts based on similarity. The SyHI data are subsequently used to validate the performance. Because the variable “height” is available in both NFI and structure map data, this variable is experimentally added to the matching process.

Data

- **NFI-7:** tree-level inventory data
- **National structure maps** (basal area, quadratic mean diameter, height, volume): plot-level inventory data
- **European genus map** (groups, conifer/broadleaf): plot-level inventory data
- **SyHI:** validation data

Results

Using basal area, quadratic mean diameter and dominant species group as constraining variables, the resulting synthetic plots show an RMSE% of 45, 40, and 82 for basal area, quadratic mean diameter, and stem density, respectively (Figure 25). Swapping the basal area map for a volume map makes a very marginal difference. The distribution of structure variables in the synthetic plots is much narrower than in the SyHI data, and the medians are too high (e.g., basal area has a median of 23 and 26 m²/ha in the SyHI and synthetic data, respectively; Figure 26A). The cause of this is probably that the national structure maps, which the synthetic data are based on, are already much more narrowly distributed and median values differ from the medians of the SyHI data (Figure 22A).

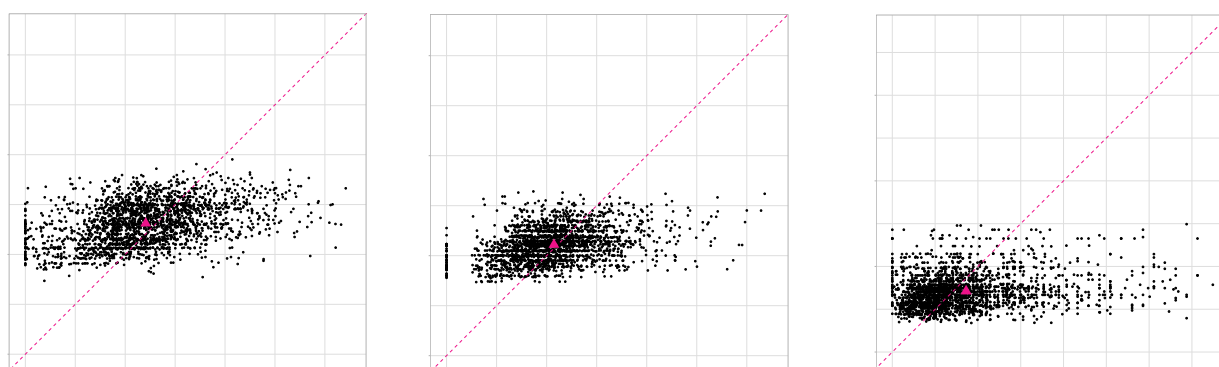


FIGURE 25 SYNTHETIC PLOTS DERIVED FROM TREE-LEVEL NFI-7 DATA AND NATIONAL STRUCTURE MAPS, USING THE OPTIMAL SETTINGS FROM EXPERIMENT 1; PLOTTED AGAINST THE ORIGINAL SYHI DATA FOR BASAL AREA, QMD, AND STEM DENSITY.

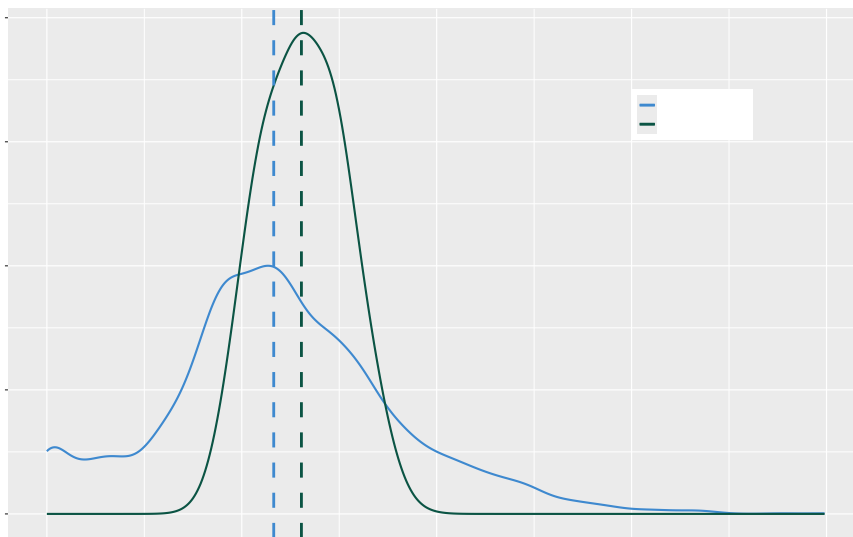


FIGURE 26 DENSITY CURVE OF BASAL AREA FOR SYNTHETIC AND REFERENCE DATA. SYNTHETIC DATA ARE BASED ON NFI-7 DATA AND NATIONAL STRUCTURE MAPS. THE DISTRIBUTION OF THE SYNTHETIC DATA IS NARROWER AND THE MEDIAN IS HIGHER.

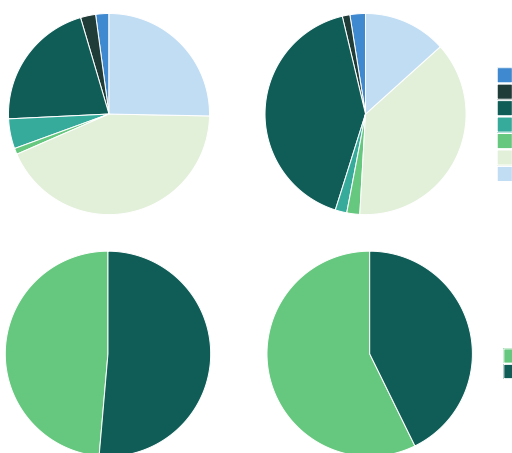


FIGURE 27 SPECIES DISTRIBUTION OF THE SYNTHETIC DATA COMPARED TO THE ORIGINAL SYHI PLOTS, WHEN USING THE SPECIES GROUPS FROM THE EUROPEAN GENUS MAP COMBINED WITH NFI-7 DATA TO CREATE SYNTHETIC PLOTS.

Using dominant species group information leads to an overall accuracy of 0.6 and 0.8 and Kappa scores of 0.5 and 0.7 on dominant species group and broadleaf/conifer distinction in the synthetic data, respectively. The proportion of oak and other broadleaf trees in the synthetic data is almost twice as high as in the validation data, while the proportion of conifer trees is underestimated (Figure 27). Finally, adding height as a constraining variable does not improve the synthetic data set in any way.

Conclusions

Using the best combination of variables from experiment NL-1 in combination with national structure maps and a European species map leads to high inaccuracies in the structure variables of the synthetic plots. The accuracy of the dominant tree species group is satisfactory, although the proportion of broadleaf trees is overestimated. The dominant height does not improve the synthetic data either. Considering that exactly the same combination of variables as in *experiment 1* was used, we conclude that the results can never be better than the quality of the maps that are used as aggregated, plot-level data. This experiment shows that using a map product with narrower value distributions than in the original data leads to a synthetic data set where the values are also much narrower.

Experiment NL-3: In the absence of national data sets, how well can we do with pan-European products?

Description

The goal of experiment NL-3 is to test if currently available European map products can be used to create synthetic data sets when national, spatially explicit, plot-level data is not available. For this purpose, European-wide forest structure maps from the PathFinder project and the tree genus map from the ForestPaths project were used (see 3.1.3.2).

Data

- **NFI-7:** tree-level inventory data
- **European structure maps** (volume, AGB, percentage of conifer AGB): plot-level inventory data
- **European genus map** (groups, conifer/broadleaf): plot-level inventory data
- **SyHI:** validation data

Results

Using volume or AGB as constraining variable, the resulting synthetic plots show an RMSE% of 61% and 55% for basal area and qmd respectively, and more than 100% for stem density. The AGB map performs slightly better than the volume map (Figure 28).

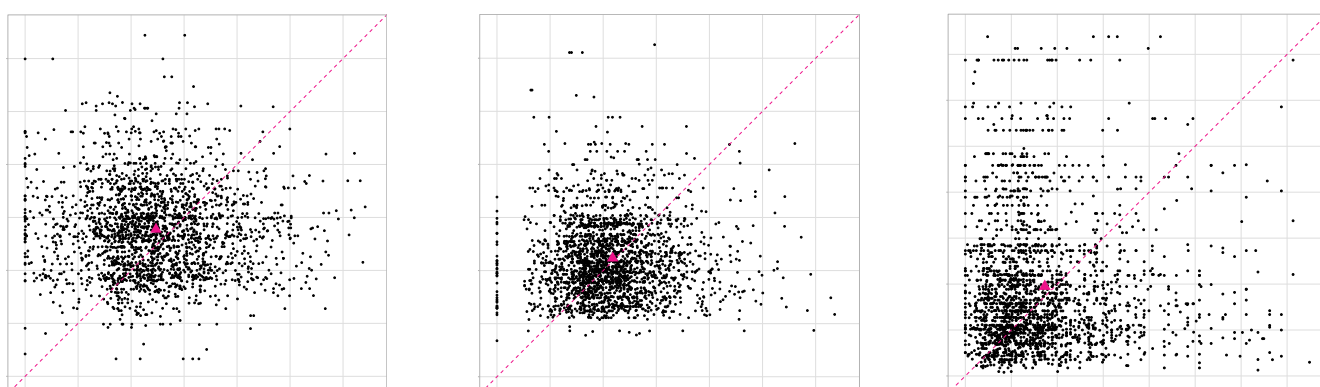


FIGURE 28 SYNTHETIC DATA DERIVED FROM TREE-LEVEL NFI-7 DATA AND EUROPEAN STRUCTURE MAPS, USING AGB AND THE DOMINANT SPECIES GROUP TO ALLOCATE PLOTS; PLOTTED AGAINST THE ORIGINAL SYHI DATA FOR BASAL AREA, QMD, AND STEM DENSITY.

The ForestPaths pan-European dominant species map leads to an overall accuracy of 0.6 and 0.8 and Kappa scores of 0.5 and 0.7 on dominant species group and broadleaf/conifer distinction, respectively. The PathFinder map of percent conifer biomass leads to an overall accuracy of 0.7 and Kappa score of 0.5 in conifer/broadleaf distinction, as well as an overall accuracy of 0.4 and Kappa score of 0.2 for the dominant species group.

Conclusions

The current PathFinder European structure maps that we used in this study do not allow for creating a synthetic dataset with representative diameter distribution for the study area. The conifer biomass map is decent for determining if the dominant species is broadleaf or conifer but better results are obtained when a dedicated genus map is used.

Experiment NL-4: Can we use data from neighbouring regions?

Description

The goal of experiment NL-4 is to test if tree-level data from a neighbouring country can be used to create a synthetic data set. In the previous experiments, we concluded that neither the national nor the European structure maps deliver the desired results, even with national tree-level data. Therefore, in this experiment, we limit ourselves to “perfect data” for the plot-level data, as was done in experiment NL-1. The source of tree-level data will be the German NFI-3.

Data

- **German NFI-3:** tree-level inventory data
- **SyHI:** plot-level and validation data

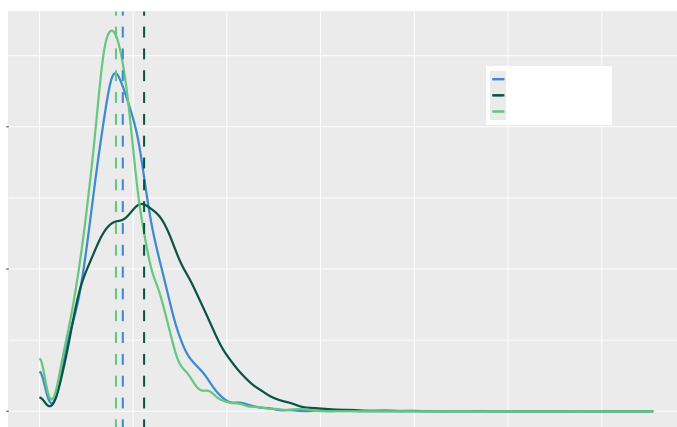


FIGURE 29 QMD DISTRIBUTION OF THE DIFFERENT INVENTORY SOURCES. VERTICAL DASHED LINES INDICATE THE MEDIAN VALUES.

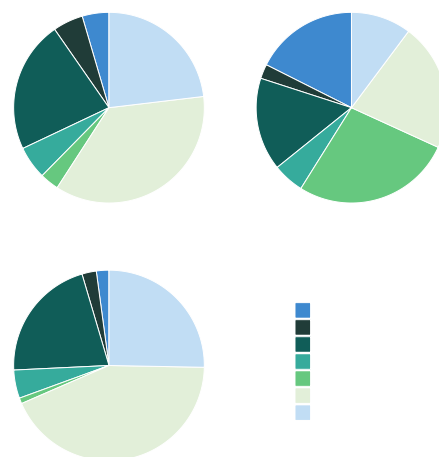


FIGURE 30 DOMINANT SPECIES GROUP DISTRIBUTION OF THE DIFFERENT INVENTORY SOURCES.

The German data have a different structure from both Dutch data sets. Basal area and qmd have a wider distribution with higher medians, while stem density has a lower median (e.g. Figure 29). The distribution of dominant species groups is quite different with much more spruce and beech in the German NFI data, and less oak (Figure 30).

Results

Comparing the synthetic data made with German NFI data, using basal area, qmd, and/or stem density variables to allocate the synthetic plots, with the results from experiment NL-1 (Figure 23) shows very similar results; meaning a nearly perfect synthetic dataset is made. However, the results of using volume instead of basal area leads to %RMSE values of 25% and 82% for basal area and stem density (Figure 31), compared to 16% and 25% in *experiment 1* (Figure 23B). This is probably due to the relationship of basal area to volume being different in the Netherlands than in Germany (greater volume at the same basal area, due to generally taller trees in Germany).

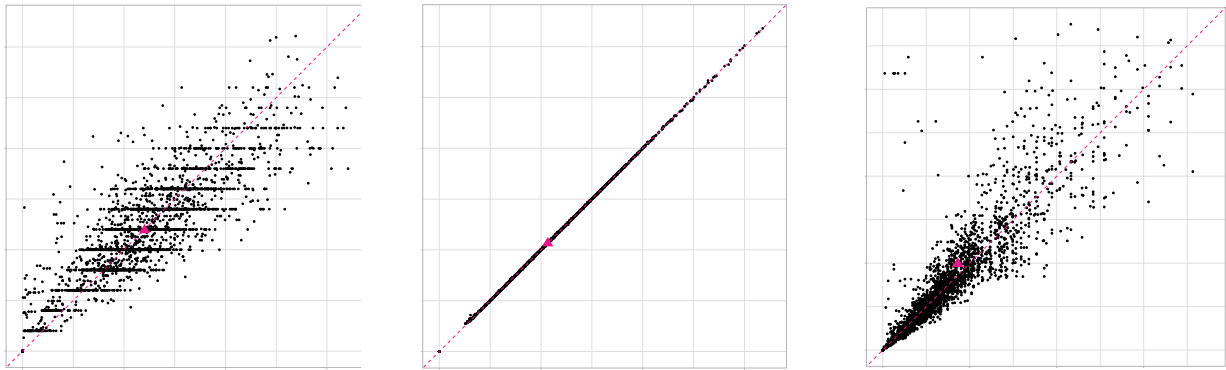


FIGURE 31 SYNTHETIC PLOTS DERIVED FROM GERMAN NFI-3 TREE-LEVEL DATA USING VOLUME AND QMD AS CONSTRAINING VARIABLES; PLOTTED AGAINST THE ORIGINAL SYHI DATA FOR BASAL AREA, QMD, AND STEM DENSITY.

Conclusions

Data from outside the Netherlands can be used to create good structure maps if basal area is used. Volume cannot be used, because the relationship of basal area to volume is different in both countries.

Main findings & conclusions of case study 1

Only two variables out of basal area, stem density, and quadratic mean diameter are needed to create a synthetic data set when using high-quality, plot-level data. Furthermore, volume, which is more often available in maps, can be used as a good alternative for basal area. Broadleaf-conifer or genus (or another grouping) information is needed for a good estimate of species distribution. Soil information does not offer a good alternative to this. These conclusions are also valid when the tree-level data is from a different region; although volume may not be a good alternative if the volume-to-basal-area ratios in this region are generally different.

With the national structure maps that were available within this project, the synthetic method using the best combination of variables from experiment NL-1 performs poorly on creating a synthetic data set that represents the actual distribution of structure variables. The dominant height does not improve the match either.

Current European numeric structure maps that we used in this study do not allow for creating a synthetic dataset with a representative diameter distribution for the study area. The conifer biomass map helps to determine if the dominant species is broadleaf or conifer, but better results are obtained when a dedicated species (or genus) map is used. Miettinen et al., (2025) explain that their map products are mainly meant for auxiliary data in model-assisted estimations and are suitable as broader overview maps, but are less suitable for local interpretation. This probably explains why the synthetic data obtained when using these maps do not represent the local forest conditions well. The European genus map that was used ensured quite a good allocation of species in the synthetic data, although it overestimated the proportion of broadleaf trees.

3.2.2. Case study 2: Finland

Background

In Finland, 86% of the area is forestry land, that is land that is not used for agriculture or infrastructure. Of the productive forest in Finland (around 75% of the total forestry land), 65% of forests are dominated by pine, 25% by spruce, and 10% by broadleaves. Mixed deciduous and coniferous forests with at least 10% of mixed trees can be found on 45% of the forest land, and 14% of forests have at least 25% of mixed trees. The growing stock in Finland varies significantly between regions with a mean volume of growing stock of 122 cubic meters per hectare (Korhonen et al. 2024).

National Forest Inventories (NFI) in Finland have been conducted since the 1920s, initially by the Finnish Forest Research Centre (Metla), now by the Natural Resource Institute Finland (LUKE). The inventory is based on statistical sampling of a regular field plot network across entire Finland and is nowadays repeated every 5 years (every 10 years until the end of 20th century). The latest available measurement cycle is the 12th NFI cycle from 2014 to 2018. Currently, data on sample plot and forest stand level are publicly available, but they do not include the tree-level measurements (Korhonen, 2023).

In addition to the NFI data, LUKE prepares multi-source inventory data, based on the field data and by using satellite images, digital map data and other georeferenced data. Various thematic forest maps are produced as georeferenced raster layers with a spatial resolution of 16 x 16 m, including mean diameter at breast height (dbh) and height of stands, stand basal area, canopy cover of trees, and canopy cover of broad-leaved trees (Natural Resources Institute Finland, 2021).

The Finnish Forest Centre (in Finnish: Metsäkeskus) is a state-funded organization that provides forest resource information based on sample plot measurements as well as laser scanning from aircraft and aerial photography. In 2020, the second remote sensing-based inventory started and will finish in 2025. In contrast to the NFI by LUKE, the Finnish Forest Centre focuses on collecting information on private forests while still providing basic inventory information from all forests (Finnish Forest Centre, 2025).

In addition to national datasets in Finland, several European or global maps are available, although often with lower resolution and/or accuracy than national datasets (see section 3.1.3.2).

The variety of available datasets make Finland a suitable and interesting case study. The data provided by the Finnish Forest Centre, hereinafter referred to as 'Metsäkeskus' data, is the only dataset containing tree-level information. Additionally, we added an experiment that uses the Swedish NFI data as tree-level information.

We conducted the following experiments to test the procedure on generating synthetic forest data for Finland:

Experiment 1: How well can we do with the available national datasets?

Experiment 2: How well do European maps perform, compared to national datasets?

Experiment 3: Can we use tree-level data from neighbouring countries with similar conditions?

For each of these experiments, we evaluated the general applicability of the developed procedure as well as the usability of the different datasets. Additionally, we compared the resulting synthetic datasets with official regional growing stock statistics from LUKE, where possible (National Resources Institute Finland, 2024).

Input data and preprocessing

National Forest Inventory (NFI12)

The Finnish NFI12 was conducted as a systematic cluster sampling, as in previous inventories, with 60% permanent plots and 40% temporary plots (Figure 32). In total, almost 65,000 plots were measured; measurements are done at three different levels: tree, sample plot and forest stand, of which only data on sample plot and forest stand level are publicly available. These data contain the following plot level information: mean dbh, mean height, age, basal area, number of stems, dominant tree species, and more. For the experiment, the following variables were extracted for all NFI plots: dominant tree species, mean dbh, mean height, stem number, and basal area.

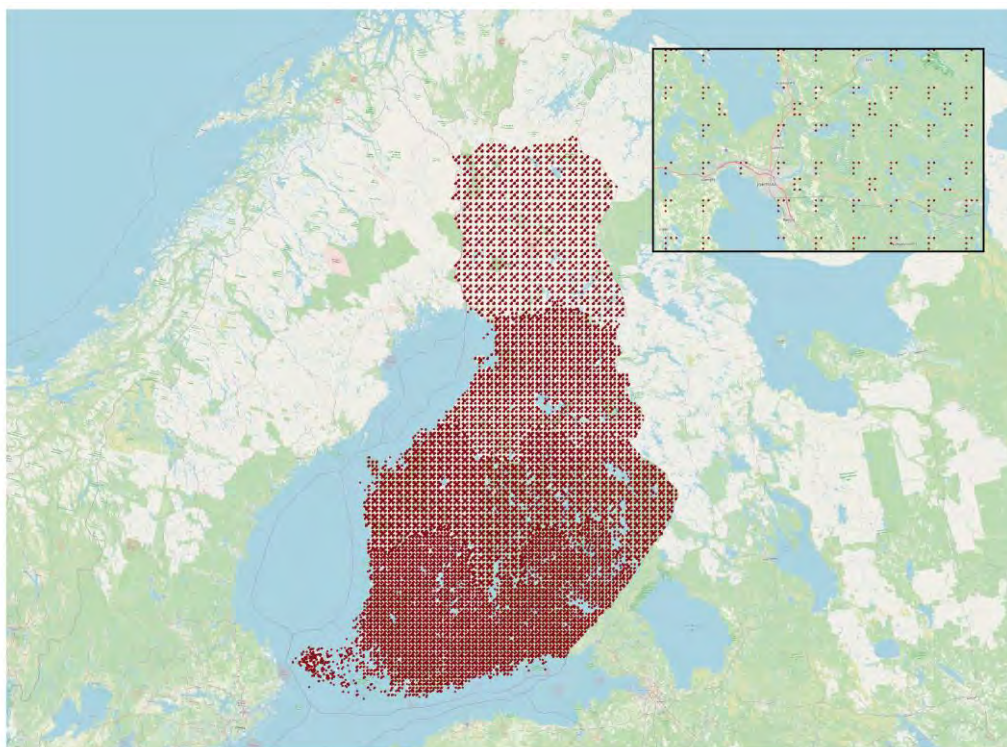


FIGURE 32 DISTRIBUTION OF FINNISH NFI12 PLOTS ACROSS FINLAND (CREATED IN QGIS).

Multi-source forest inventory (MS-NFI)

The available Finnish MS-NFI data correspond to the years 2017 and 2019 and includes field data from the last three NFI cycles (2012-2019) as well as satellite images for the years 2017 and 2019. Satellite images include data from Sentinel-2A/B overflights and Landsat 8 overflights. A non-parametric k-NN estimation was used to generate forest data for each pixel across Finland. The available raster layers have a resolution of 16 x 16m.

For the experiment, the following raster layers were used: dbh, basal area, height, and volume of birch, broadleaves, spruce and pine. The information of volume for different tree species/groups was used to identify the dominant tree species of a cell/synthetic plot, as no other data on species composition are available from the MS-NFI data.

Finnish Forest Centre (Metsäkeskus)

The tree-level data from the Finnish Forest Centre are a mix of field-measured data and data estimated based on field and remote-sensing data. In each sample plot, all trees are counted and the respective species noted, but only for a subset of trees, additional forest structure variables are measured, e.g. dbh, height. The forest structure variables for the remaining trees are estimated using the measured data and remote-sensing data. Until early 2024, forest resource data from the newest measuring cycle were available for approximately half of Finland, with more data becoming available in the following years. For the case study, available data until early 2024 were used (Figure 33).

For the experiment, the following variables were extracted for all sample plots and respective trees: dominant tree species, tree species (for each tree), dbh (for each tree), basal area, stem number, and height (for each tree). Average values for dbh and height were calculated for each sample plot. Plots for which values for dbh or height is missing for more than 25% of the trees were excluded from the case study. The remaining missing information on dbh and height were estimated per plot, based on average values.

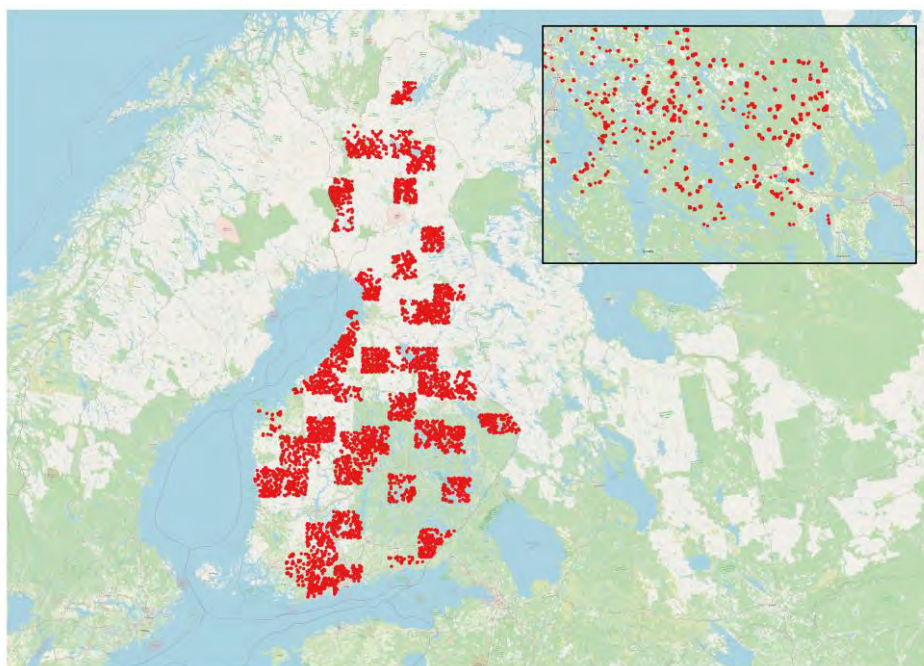


FIGURE 33 DISTRIBUTION OF AVAILABLE METSÄKESKUS PLOTS ACROSS FINLAND (CREATED IN QGIS).

European maps

For the experiment, we used the information on volume and deciduous-coniferous proportion from the European structure maps developed by Miettinen et al. (2024). We used the European genus map developed by De Keersmaecker et al. (2025) containing information about the following tree genus: *Larix*, *Picea*, *Pinus*, *Fagus*, *Quercus*, other needleleaf, and other broadleaf (see section 3.1.3.2).

Swedish NFI

Sweden has been conducting National Forest Inventories since 1923. Sample plots are circular and organized in clusters which vary in dimension across the country. The number of plots also changes for different parts of the country and plots are divided into permanent and temporary plots. Permanent plots are distributed in a systematic grid and are revisited every five years. Temporary plots are distributed following the LPM-method and are only visited once (Fridman et al., 2014).

Inventory data contain information on tree species, dbh, and tree density (Source: National Forest Inventory, SLU, 2025). For the experiments, we included data between 2015-2021 and used the basal area per species per plot to determine the dominant species. Due to missing information on tree height, the experiments do not include data combinations of Swedish NFI, as tree-level data, and the available European maps, as plot-level data, since volume estimations cannot be made reliably.

Finnish resource information (LUKE statistical data)

Based on the Finnish NFI data, LUKE produces information on growing stock volume on forest land by tree species (Pine, Spruce, Birch, and other broadleaves) and region. These data can be used to compare and validate the synthetic data outputs on a regional level (Figure 34).

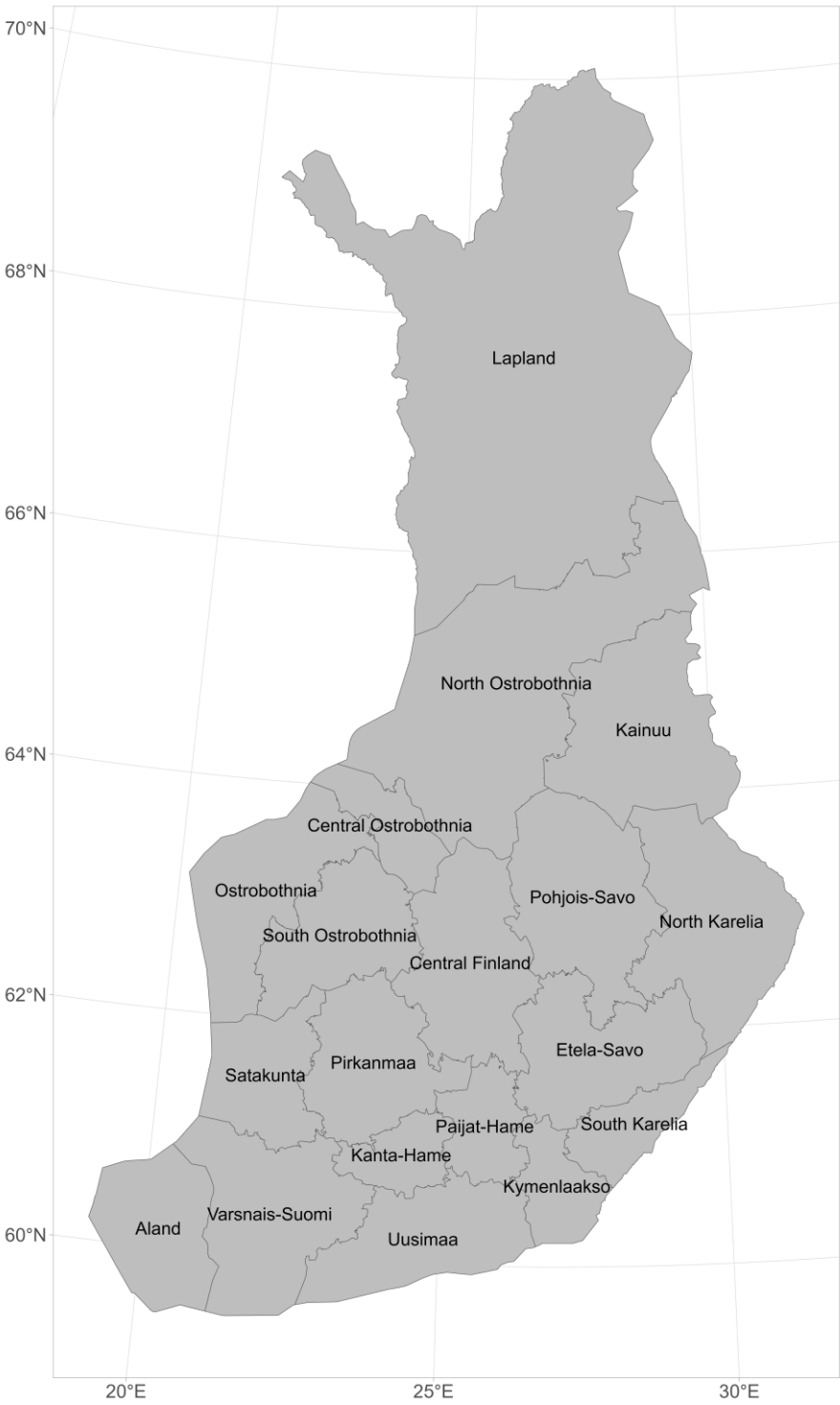


FIGURE 34 REGIONS IN FINLAND, AS USED FOR FOREST RESOURCE REPORTING (CREATED IN RSTUDIO).

Experiment FI-1: How well can we do with the available national datasets?

Methods

In experiment FI-1, we test the combination of Metsäkeskus data (tree-level data) with NFI12 data (plot-level data), and Metsäkeskus data (tree-level data) with MS-NFI data (plot-level data) to test the performance of available national datasets.

For the combination of Metsäkeskus data with NFI12 data, we use the plot locations of the NFI12 as grid for the synthetic dataset, i.e. for each NFI plot location the Metsäkeskus plot with the nearest n-dimensional distance was identified. We include the following variables: dbh, basal area, stem number and dominant tree species, following variable availability and conclusions drawn from NL-1 (Page 58).

For the combination of Metsäkeskus data with MS-NFI data, we use the plot location and respective area of all Metsäkeskus plots as grid for the synthetic data, i.e. for each Metsäkeskus plot location the MS-NFI values falling within the plot area are extracted and averaged to be used as plot-level dataset. This allows a direct comparison of the original tree-level data with the synthetic data in the evaluation step. Since the information on dominant species is limited to the four classes available from MS-NFI data, i.e. birch, other broadleaves, pine, and spruce, we adapt the classification for dominant species in the Metsäkeskus dataset accordingly. To determine the dominant species from the MS-NFI data, we extract the dominant species of all grid cells falling within a Metsäkeskus plot and use the one occurring the most often as dominant species. We include the following variables: dbh, basal area and dominant tree species, following variable availability.

Results

The results of synthetic basal area with NFI12 data show a %RMSE of 9.05 (Figure 35B), while results with MS-NFI data show a %RMSE of 44.91 (Figure 36B), suggesting that the NFI12 data perform noticeably better than the MS-NFI data. Similarly, synthetic mean dbh with NFI12 data results in a %RMSE of 4.76 (Figure 35A), while results with MS-NFI data show a %RMSE of 35.73 (Figure 36A). Value distributions of original and synthetic basal area and mean dbh show a similar picture, with results of NFI12 data fitting the original value distributions much better than results with MS-NFI data. Synthetic stem density shows similar %RMSE for results with either dataset, i.e. %RMSE = 52.4 with NFI12 (Figure 35C) and %RMSE = 59.42 with MS-NFI (Figure 36C). Value distributions show that results with either dataset tend to overestimate stem density, however results with NFI12 data are better following the original value distribution than results with MS-NFI data.

When comparing the assignment of dominant species to the synthetic plots, we can see that results with NFI12 data show a very high overall accuracy of 99.7% (Figure 35D), while results with MS-NFI data show a slightly lower accuracy of 77.3% (Figure 36D). The Cohen's Kappa values with NFI12 and MS-NFI data are 99.4% and 49.3%, respectively. The pie chart indicates that with MS-NFI data, the results are mainly over-representing *Pinus sylvestris* and *Betula sp.* while under-representing *Picea abies*.

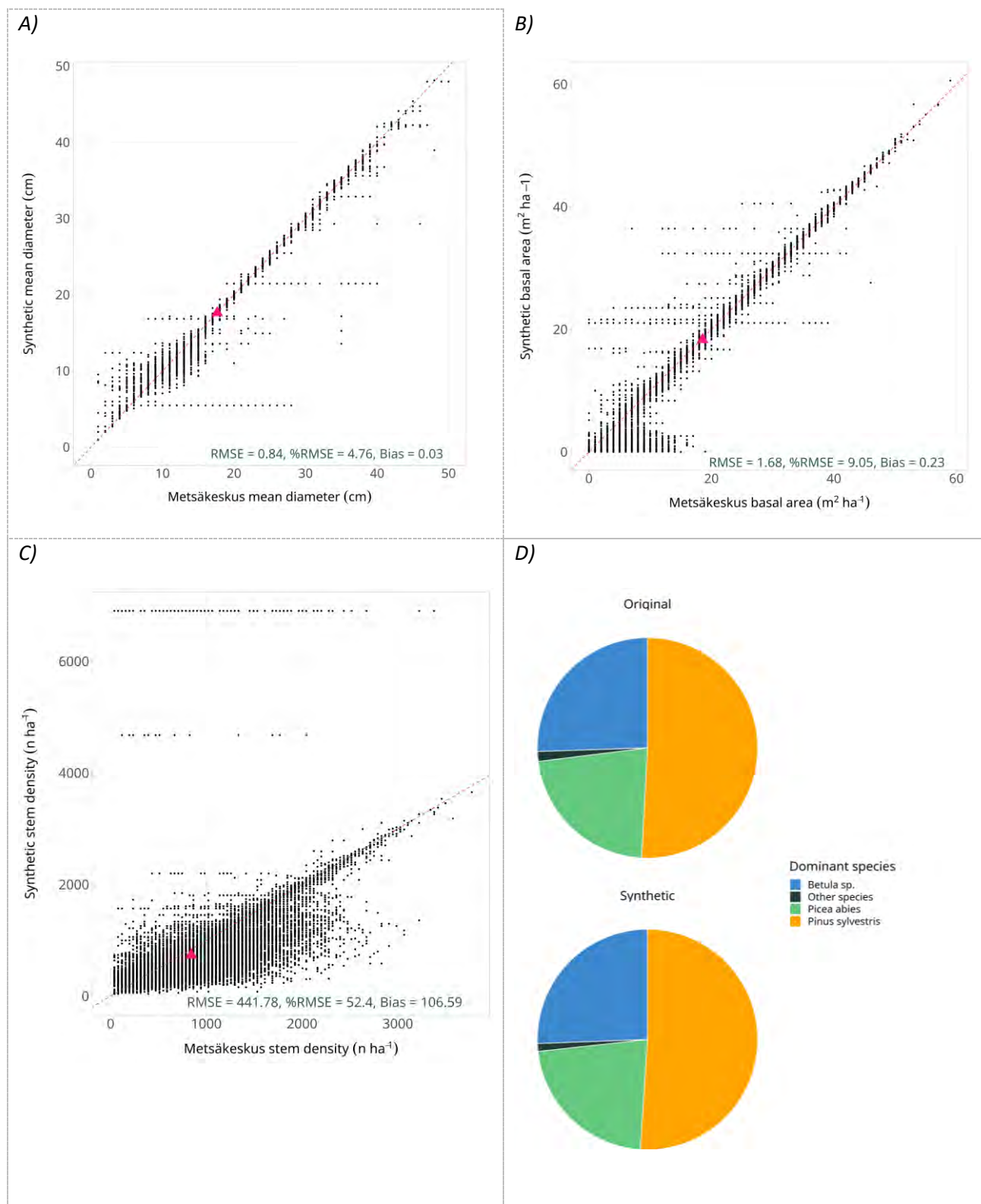


FIGURE 35: SCATTERPLOTS AND PIE CHART OF ORIGINAL AND SYNTHETIC DATA. ORIGINAL DATA = METSÄKESKUS DATA, SYNTHETIC DATA = METSÄKESKUS|NFI12 DATA. VARIABLES ARE A) MEAN DBH, B) BASAL AREA, C) STEM DENSITY AND D) DOMINANT SPECIES.

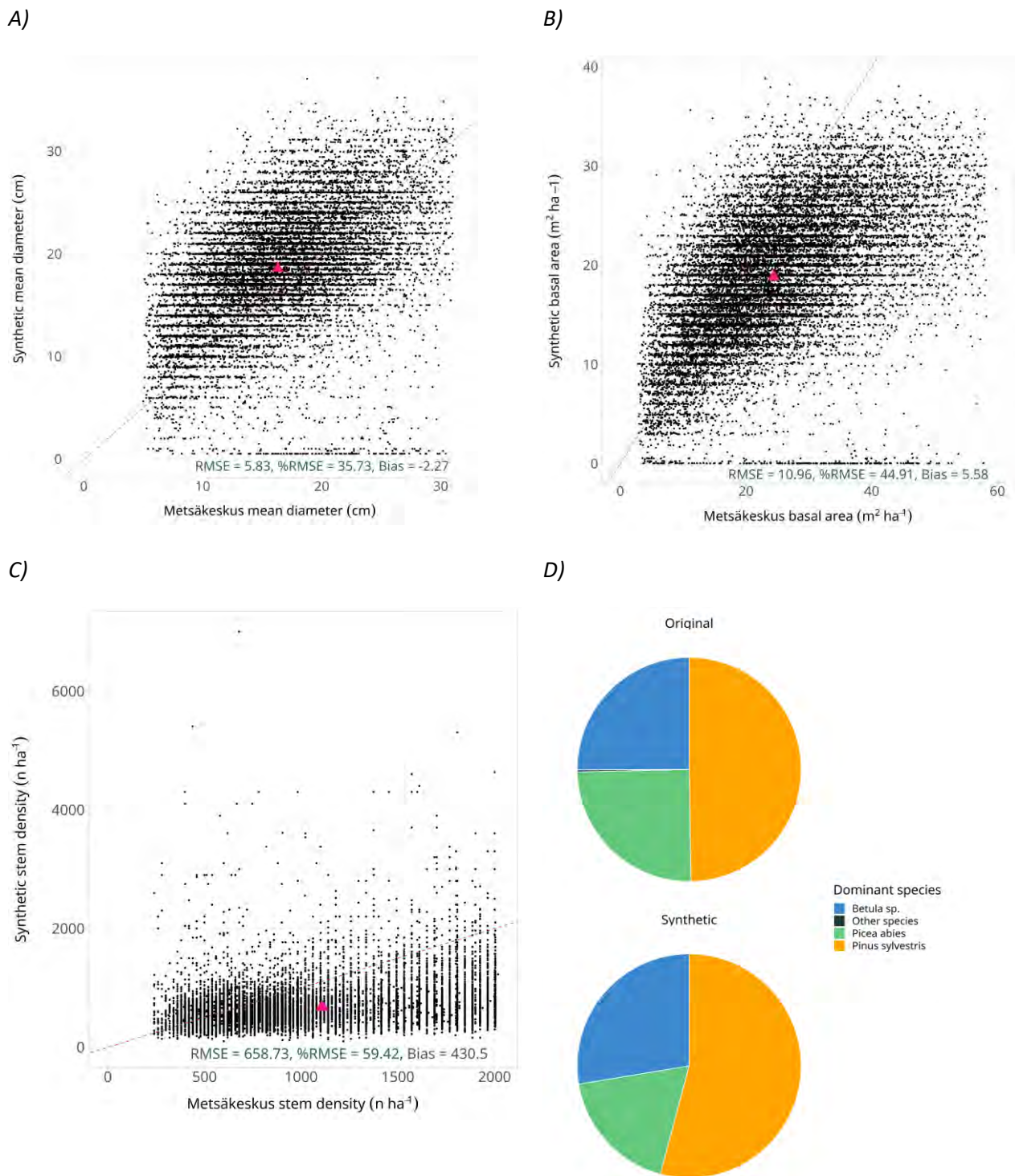
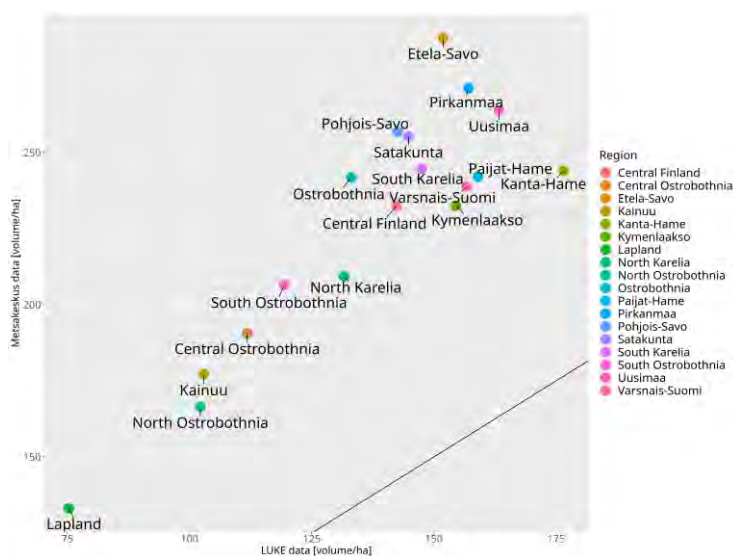


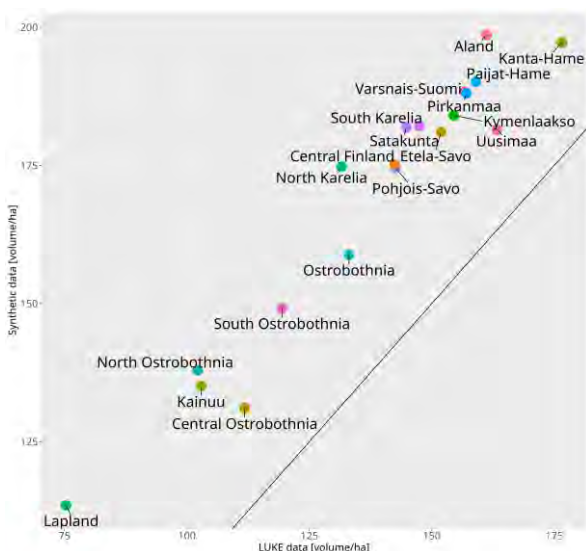
FIGURE 36: SCATTERPLOTS AND PIE CHART OF ORIGINAL AND SYNTHETIC DATA. ORIGINAL DATA = METSÄKESKUS DATA, SYNTHETIC DATA = METSÄKESKUS|MS-NFI DATA. VARIABLES ARE A) MEAN DBH, B) BASAL AREA, C) STEM DENSITY AND D) DOMINANT SPECIES.

The comparison of regional volume values from LUKE statistical data and regionally aggregated volume values from Metsäkeskus data shows that the original tree-level data from Metsäkeskus has a higher volume per ha growing stock than what is recorded in the data from LUKE (Figure 37A). Both synthetic datasets, with either NFI12 or MS-NFI data, show a better alignment with LUKE statistical data compared to the Metsäkeskus data (Figure 37B,C). This is to be expected, since the underlying maps used for the synthetic data come from the same or similar data sources as the regional growing stock data. However, for both synthetic datasets, the simulated growing stock is in every region higher than the reported values from LUKE, except for Etelä-Savo in Figure 37C.

A)



B)



C)

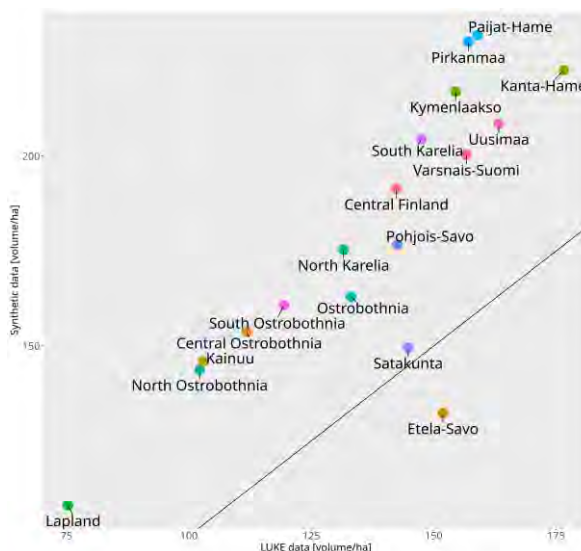


FIGURE 37: REGIONAL COMPARISON OF VOLUME, X-AXIS SHOWS THE OFFICIAL DATA FROM NATURAL RESOURCES INSTITUTE (LUKE). A) ORIGINAL METSÄKESKUS DATA, B) SYNTHETIC DATA WITH NFI12 AND C) SYNTHETIC DATA WITH MS-NFI.

Conclusion

As expected, the use of plot averages from NFI12 data as a basis for developing a synthetic dataset outperforms the use of grid cell averages derived from MS-NFI data for estimating structural variables. Stem density is the most challenging variable to align, even when it is included as an input (NFI12). In addition, NFI12 data provide higher accuracy in identifying the dominant tree species compared to MS-NFI, which may partly be due to the fact that in MS-NFI data the dominant species had to be inferred from volume data.

Experiment FI-2: How well do European maps perform compared to national datasets?

Methods

In experiment FI-2, we test the performance of available European maps, i.e. the European structure maps (volume, conifer proportion) (Miettinen et al., 2024), and the European genus map (de Keersmaecker et al., 2025). We use the plot location and respective area of all Metsäkeskus plots as grid for the synthetic data, i.e., for each Metsäkeskus plot location the values from the European maps falling within the plot area are extracted and averaged to be used as plot-level dataset. The Metsäkeskus dataset does not include volume as variable, therefore we use generic volume functions developed spruce, pine and birch trees in Finland, based on dbh and height, to estimate volume (Kangas et al., 2022). Since the information on dominant species is limited to the available genus classes, we adapt the classification for dominant species in the Metsäkeskus dataset accordingly. To determine the dominant genus from the European genus map by De Keersmaecker et al., we extract the present genus of all grid cells falling within a Metsäkeskus plot and use the one occurring the most often as dominant genus. We include the following variables: volume, conifer proportion and dominant tree genus.

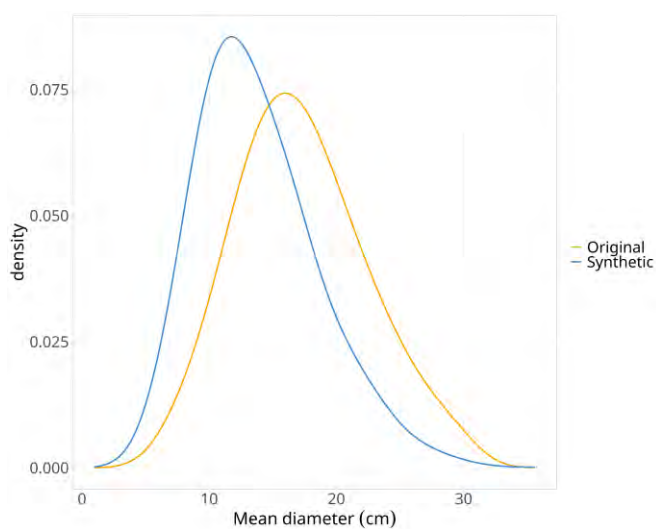
Results

For creating synthetic data, the use of the European forest structure map shows slightly better results for basal area than the MS-NFI dataset (%RMSE: European data = 37.75 (Figure 38B), MS-NFI = 44.91 (Figure 36B)) and very similar results for mean dbh (%RMSE: European data = 37.22 (Figure 38A), MS-NFI = 35.73 (Figure 36A)). The stem density shows a large error and a high variability for synthetic values (%RMSE 116.88), the value distribution indicates that lower original stem densities are under-represented and higher original stem densities are over-represented in the synthetic dataset (Figure 38C). Value distributions for basal area and mean dbh show an under-estimation of synthetic mean dbh compared to original mean dbh while basal area shows an acceptable overlay between original and synthetic data (Figure 38A,B).

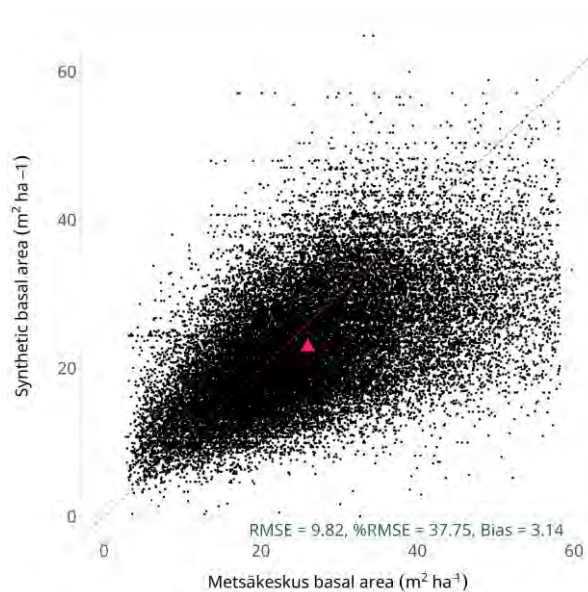
The synthetic dominant species result in an overall accuracy of 80.6% and a Cohen's Kappa value of 90.6%, indicating better results for creating synthetic data than with MS-NFI data, but lower than with NFI12 data (Figure 38D).

The aggregated regional volume values from synthetic data created with the European maps are similar to those from synthetic data with NFI12 and with MS-NFI. The simulated growing stock in the synthetic dataset is, however, still higher than the growing stock reported by LUKE statistical data (Figure 39).

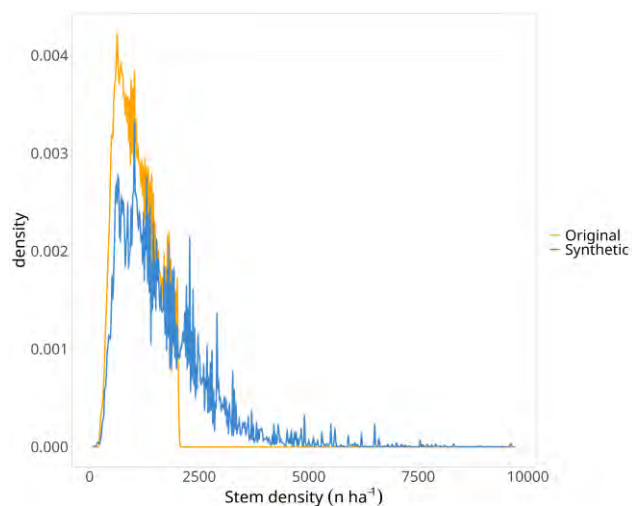
A)



B)



C)



D)

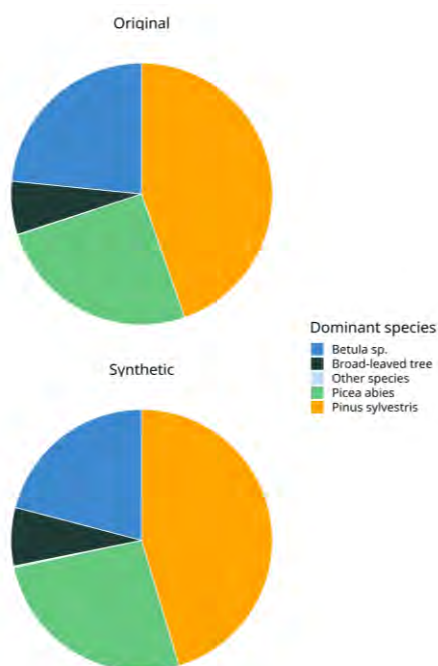


FIGURE 38: DENSITY PLOTS, SCATTER PLOT AND PIE CHART OF ORIGINAL AND SYNTHETIC DATA. ORIGINAL DATA = METSÄKESKUS, SYNTHETIC DATA = METSÄKESKUS|EUROPEAN MAPS. VARIABLES ARE A) MEAN DBH, B) BASAL AREA, C) STEM DENSITY AND D) DOMINANT SPECIES.

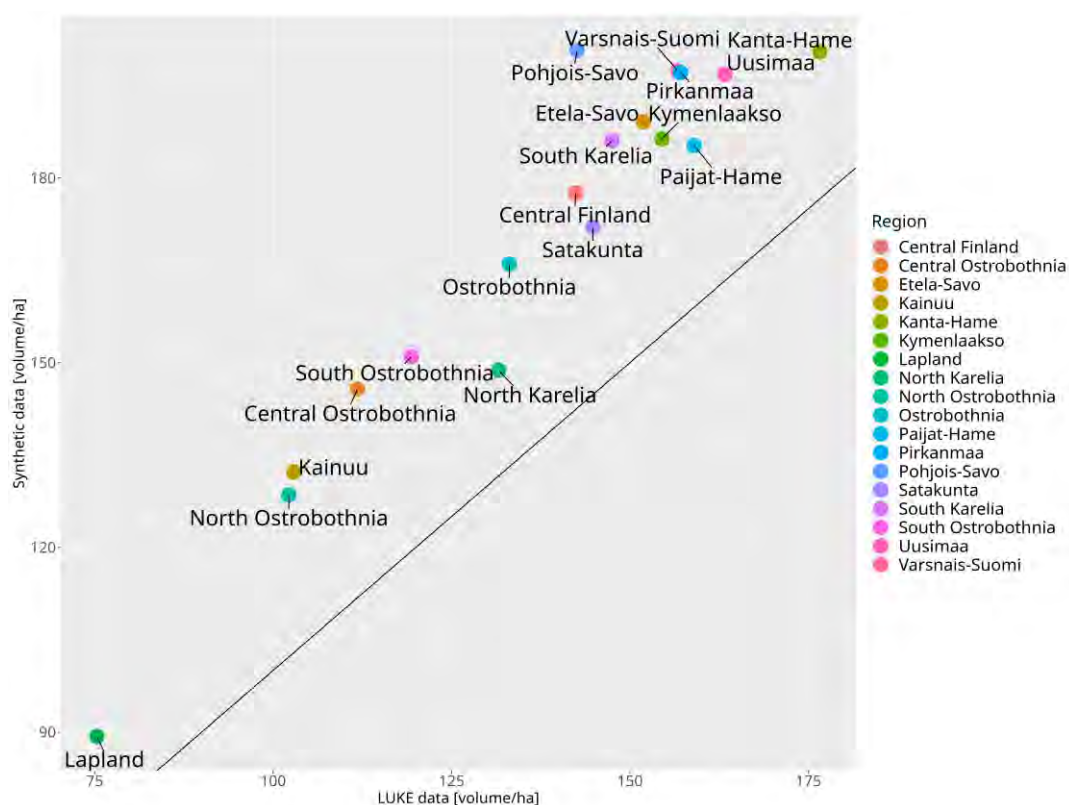


FIGURE 39: VOLUME COMPARISON - REGIONAL DATA AND SYNTHETIC DATA WITH EUROPEAN MAPS.

Conclusion

The use of European maps to create synthetic inventory datasets yields results that are comparable to the national MS-NFI data, but they provide a better match in dominant species allocation. This indicates that the use of MS-NFI in our approach of estimating dominant species from volume data may introduce bias. Overall, synthetic data derived from European maps could serve as a suitable alternative for creating synthetic datasets when tree-level information is unavailable.

Experiment FI-3: Can we use tree-level data from neighbouring countries with similar conditions?

Methods

In experiment FI-3, we test the usability of tree-level data from a neighbouring country with similar conditions by using National Forest Inventory data from Sweden. We test two combinations, the Swedish NFI data with NFI12 data, and Swedish NFI data with MS-NFI data.

For the combination of Swedish NFI data and NFI12 data, we use the plot locations of the NFI12 as grid for the synthetic data, i.e., for each NFI plot location we identify the Swedish NFI plot with the nearest n-dimensional distance. We include the following variables: dbh, basal area, stem number and dominant tree species.

For the combination of Swedish NFI data and MS-NFI data, we use the plot locations and respective areas of all Metsäkeskus plots as grid for the synthetic data, i.e. for each Metsäkeskus plot location we extract and average the MS-NFI values falling within the plot area to be used as plot-level dataset and match the Swedish NFI plot with the nearest n-dimensional distance as tree-level data. By using the Metsäkeskus plots as locations for the synthetic data, we can later compare results from experiment FI-3 with results from experiment FI-1. Since the information on dominant species is limited to the four classes available from MS-NFI data, i.e. birch, other broadleaves, pine, and spruce, we adapt the classification for dominant species in the Swedish NFI dataset accordingly. To determine the dominant species from the MS-NFI data, we extract the dominant species of all grid cells falling within a synthetic plot area and use the one occurring the most often as dominant species. We include the following variables: dbh, basal area and dominant tree species.

Results

NFI12: Comparing the results of using tree-level data from Metsäkeskus (Figure 35) for generating synthetic datasets vs. using tree-level data from the Swedish NFI (Figure 40) shows that experiments with the Swedish NFI give slightly better results for all variables (Basal area: %RMSE 7.87, mean dbh: %RMSE 1.19, stem density: %RMSE 49.24). Value distributions show a very close fit of the synthetic data with original values for all variables. The synthetic dominant species have a similar very high overall accuracy as when using Metsäkeskus data (Accuracy: 99.4%, Cohen's Kappa: 98.9%).

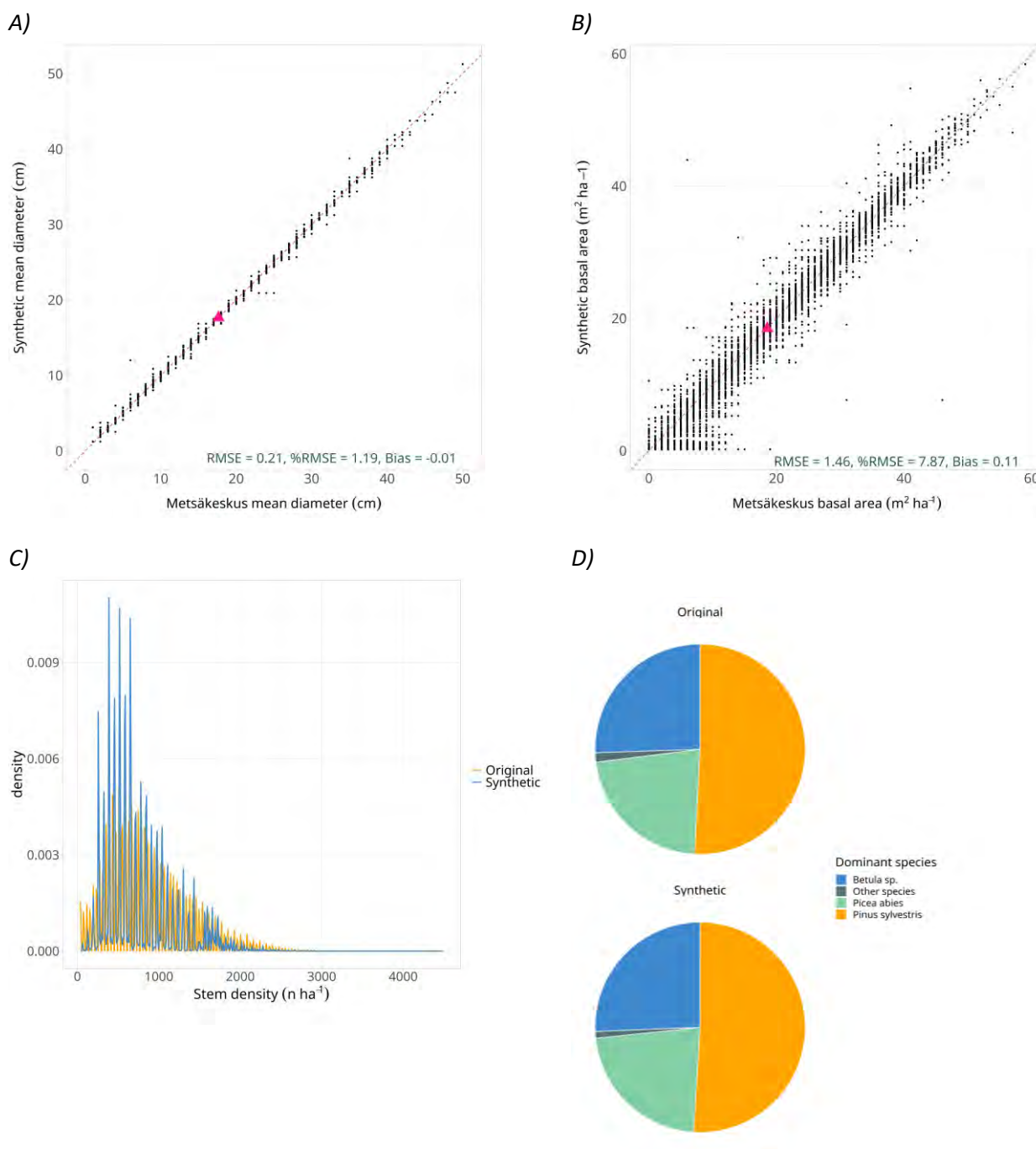


FIGURE 40: SCATTERPLOTS, DENSITY PLOT AND PIE CHART OF ORIGINAL AND SYNTHETIC DATA. ORIGINAL DATA = METSÄKESKUS DATA, SYNTHETIC DATA = SWE-NFI|NFI12. VARIABLES ARE A) MEAN DBH, B) BASAL AREA, C) STEM DENSITY AND D) DOMINANT SPECIES.

MS-NFI: Comparing the results of Metsäkeskus (Figure 36) and Swedish NFI (Figure 41) data gives very similar errors for synthetic basal area and synthetic mean dbh (Basal area: %RMSE 44.83, mean dbh: %RMSE 35.67), however synthetic stem density has a noticeable higher error when using Swedish NFI (%RMSE 71.27). Value distributions of Swedish NFI results show an under-representation of high values of basal area in the synthetic data, a general overestimation of mean dbh values, and an over-representation of low stem densities while

under-representing high stem densities compared to original values. The synthetic dominant species have a similar accuracy as when using Metsäkeskus data (Accuracy: 76.0%, Cohens' Kappa: 54.3%).

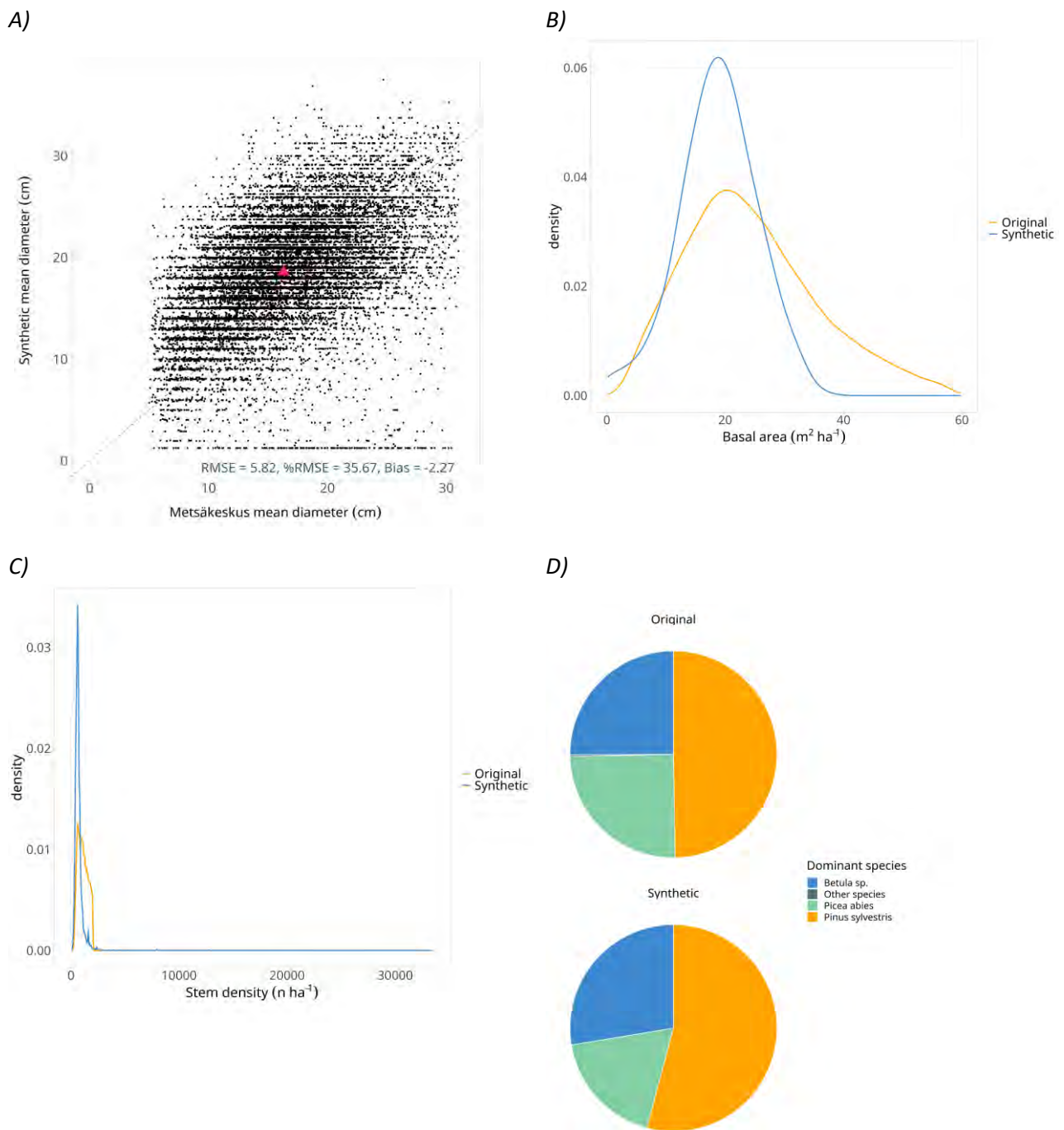


FIGURE 41: SCATTERPLOT, DENSITY PLOTS AND PIE CHART OF ORIGINAL DATA AND SYNTHETIC DATA. ORIGINAL DATA = METSÄKESKUS DATA, SYNTHETIC DATA = SWE-NFI|MS-NFI. VARIABLES ARE A) MEAN DBH, B) BASAL AREA, C) STEM DENSITY AND D) DOMINANT SPECIES.

Conclusion

Results are very similar to those of FI-1, when the same plot-level information was used. This shows that it is possible to use tree-level information from a different country or area as long as the forest structures and compositions are similar.

Main findings & Conclusions of case study 2

Overall, we see that results can only be as good as the input datasets, especially the datasets used as plot-level data. The comparison of using either Finnish or Swedish tree-level data showed that the source and location of the tree-level data used does not significantly change the results, if the tree-level data are similar to the forest structure in the area of interest. Plot-level data performed better when they are based on field-measurements than on remote-sensing information, however, also the European maps provided sufficient results, especially compared to the results of the Dutch case study. One factor that might influence the results, but is difficult to test, is the estimation of input variables that was necessary, since not all data sets contain the same information. For instance, it is possible that by estimating the volume of the tree-level data in experiment FI-2 with a generic volume model for Finland, we introduced inaccuracies into the data. In general, we can assume that creating realistic synthetic data sets in Finland is less challenging than in other European countries with more diverse structures and compositions and can be seen as a suitable method to initialize EFISCEN-Space.

3.2.3. Case study 3: Italy

Background

National Forest Inventories in Italy have been conducted since 1985, with the 4th cycle currently ongoing. In each cycle around 7,000 plots are measured, and information on tree species, dbh, height, deadwood, regeneration and soil conditions is collected. The inventories are run by CUFA and CREA, and data can be accessed via <https://www.inventarioforestale.org/>.

Based on the 3rd national inventory, the total wooded area is estimated to be more than 11 million hectares (=37% of the country area). Forest area in Italy is distributed across a wide range of altitude classes, with about two thirds of forest area located below 1000m and about one third above 1000m. The forest area in Italy consists mainly of pure broadleaved forests, i.e. more than 75% of the tree crown cover of species of one group. Pure conifer and mixed forests account for just over 10% of the total forest area. The inventories measure data of almost 180 tree species of which four species make up 50% of the wood in the forests: beech (*Fagus sylvatica*), spruce (*Picea abies*), chestnut (*Castanea sativa*) and Turkey oak (*Quercus cerris*). The two main silvicultural systems in Italy are coppices and high forests which cover just above 40% of the forest area, respectively. Even-aged stands can be found in about 75% of forests (De Laurentis et al., 2021; Gasparini et al., 2022).

Italy was chosen as one of the case studies to cover a southern European country with diverse forests, a different composition of tree species, a wide range of altitude and varying landscapes. We conducted the following experiments:

Experiment 1: How well can we perform using European maps?

Experiment 2: Can we improve results by using available national datasets?

Experiment 3: Does the inclusion of topographic layers improve the results?

Experiment 4: How do the synthetic data sets influence results of EFISCEN-Space runs, compared to the original data?

Italian INFC2015

Field surveys of the third inventory (INFC2015) started in 2017 and concluded in early 2020. Measurements on living trees were made at one point every 1276 ha of forest. Quantitative data on volume, biomass, deadwood and carbon stock were calculated as a combination of values recorded in the field and models developed during the inventory or in other equivalent studies (De Laurentis et al., 2021; Gasparini et al., 2022).

The dataset of the Italian INFC2015 we use in this case study contains almost 7000 plots from across Italy (Figure 42) and has information about individual tree volume, dbh and height as well as the respective tree species and the tree density.



FIGURE 42: PLOT LOCATIONS OF INFC2015 AS USED IN THE CASE STUDY.

National Structure Maps

Structure maps with a 10 x 10m spatial resolution were created within this project, using INFC2015 data as dependent variables and Landsat as predictors in random forest (RF) models (Francini, unpublished B). These included map layers of carbon stock and growing stock volume. Note that these maps are not independent of the INFC2015 data.

Experiment IT-1: How well can we perform using European maps?

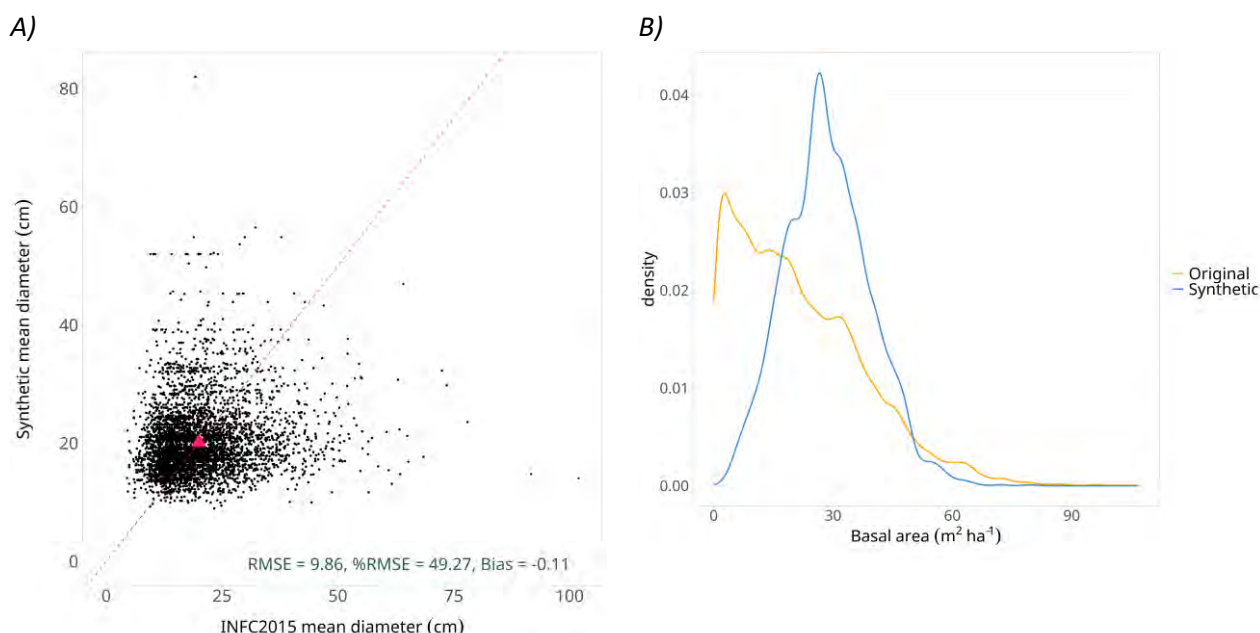
Methods

Since there are no additional official forest datasets available in Italy, besides the forest inventory data, we test the methodology with the European maps for forest structure by Miettinen et al. (2024) and tree genus by de Keersmaecker et al. (2025) in experiment IT-1 of this case study. We use the following variables: volume, main tree genus, and conifer proportion. For the combination of Italian NFI data and European maps, we use the original locations of the NFI plots as locations for our synthetic grid to allow for data comparison later on. Since the information on dominant species is limited to the available genus classes from the European genus map, we adapt the classification for dominant species in the Italian NFI dataset accordingly. To determine the dominant genus from the genus map, we extract the present genus of all grid cells falling within a NFI plot and use the one occurring the most often as dominant genus. The respective conifer proportion for NFI plots we calculate based on the volume.

Results

The results of original and synthetic basal area show an over-estimation of basal area in the synthetic dataset (%RMSE 87.69) (Figure 43B). Similarly, the dbh plot shows an over-estimation in the synthetic dataset, although indicating a better alignment than for basal area (%RMSE 49.31) (Figure 43A). The figure shows that the stem density in synthetic plots is also more often higher than lower compared to the original plots which aligns with the over-estimation of basal area (%RMSE 121.78) (Figure 43C).

The allocation of dominant tree species to the synthetic plots shows a low overall accuracy (12.4%) and low Cohen's Kappa value (4.3%). The pie charts indicate that there is a general over-representation of *Fagus sylvatica* as dominant tree species, while, e.g., *Quercus* species seem to be under-represented (Figure 43D).



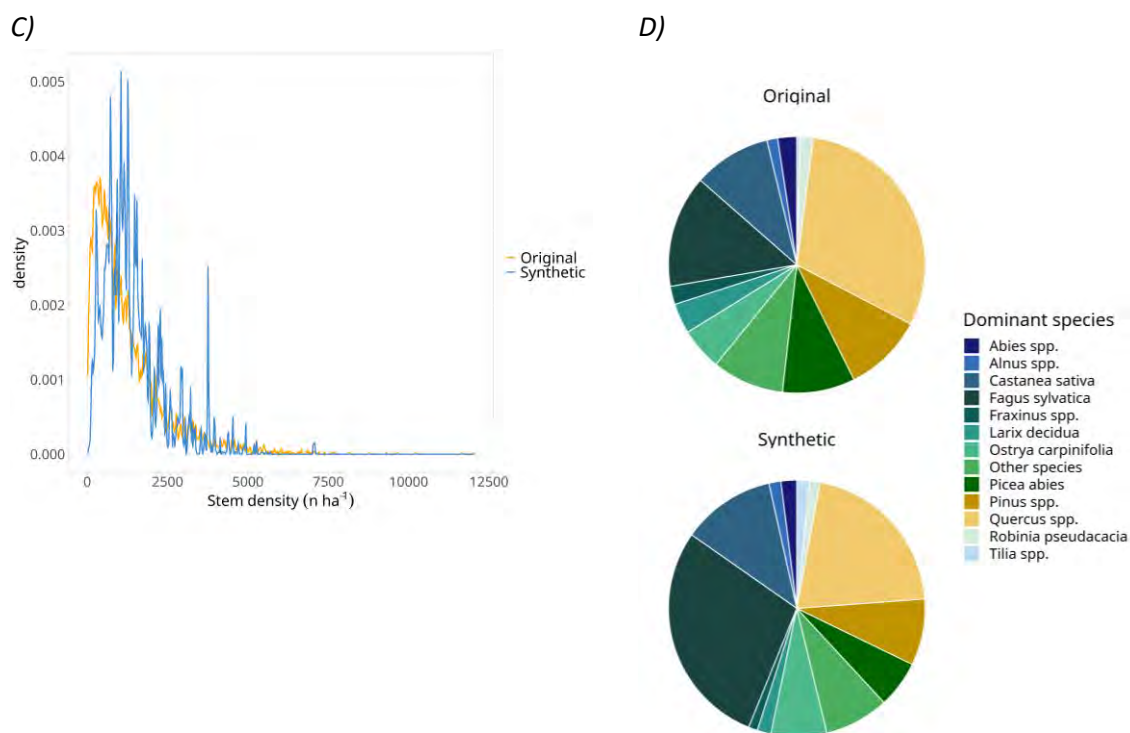


FIGURE 43: SCATTERPLOT, DENSITY PLOTS AND PIE CHART OF ORIGINAL AND SYNTHETIC DATA. ORIGINAL DATA = INFC2015, SYNTHETIC DATA = EUROPEAN MAPS. VARIABLES ARE A) MEAN DBH, B) BASAL AREA, C) STEM DENSITY AND D) DOMINANT SPECIES. DOMINANT SPECIES ARE SUMMARIZED IN 'OTHER SPECIES' IF NUMBER OF OCCURRENCES IS LOW.

Conclusion

Using European maps to create a synthetic dataset shows considerable differences to the original data, especially the alignment of dominant tree species is very low. Considering that Italy has a very diverse landscape, e.g., high elevational gradient, high biogeographic heterogeneity, the results seem promising, nevertheless. However, the resulting synthetic dataset is likely not yet sufficient for initializing EFISCEN-Space runs.

Experiment IT-2: Can we improve results by using available national datasets?

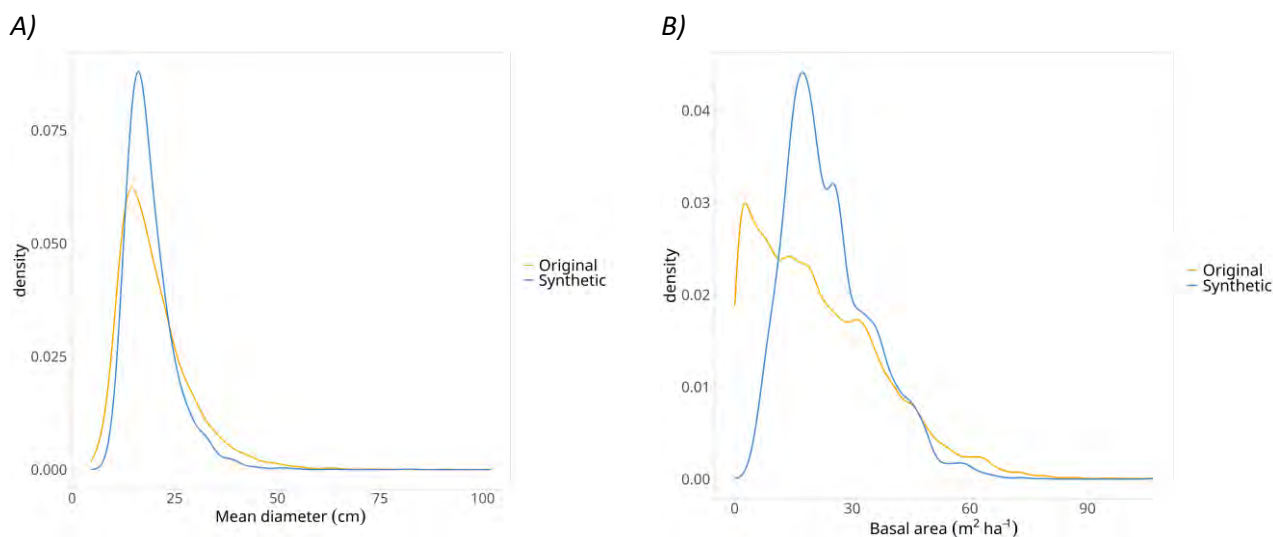
Methods

In experiment IT-2, we use a national forest map of volume in exchange of the European volume map from Miettinen et al. (2024) (see earlier descriptions). We again use the original locations of the Italian NFI plots as locations for our synthetic grid and adjust the classification of dominant species in the NFI dataset in accordance with the European genus map by de Keersmaecker et al. (2025) (see IT-1). The conifer proportion for the NFI plots we again calculate based on the volume.

Results

The national volume map improves the allocation of basal area, compared to experiment IT-1 (%RMSE 76.29), although an over-estimation of basal area in the synthetic dataset is still noticeable (Figure 44B). The distribution of synthetic dbh follows the distribution of original dbh very well, although still showing slight deviations, especially with increasing dbh (%RMSE 47.31) (Figure 44A). Results for stem density are similar to experiment IT-1 with synthetic plots more often having a higher than lower stem density compared to the original data (%RMSE 123.37) (Figure 44C).

The allocation of dominant tree species shows similar results as experiment IT-1, with a low overall accuracy (13.5%) and low Cohen's Kappa value (6.4%). Similar to experiment IT-1, there is a general over-representation of *Fagus sylvatica* and an under-representation of, e.g., *Quercus* species (Figure 44D).



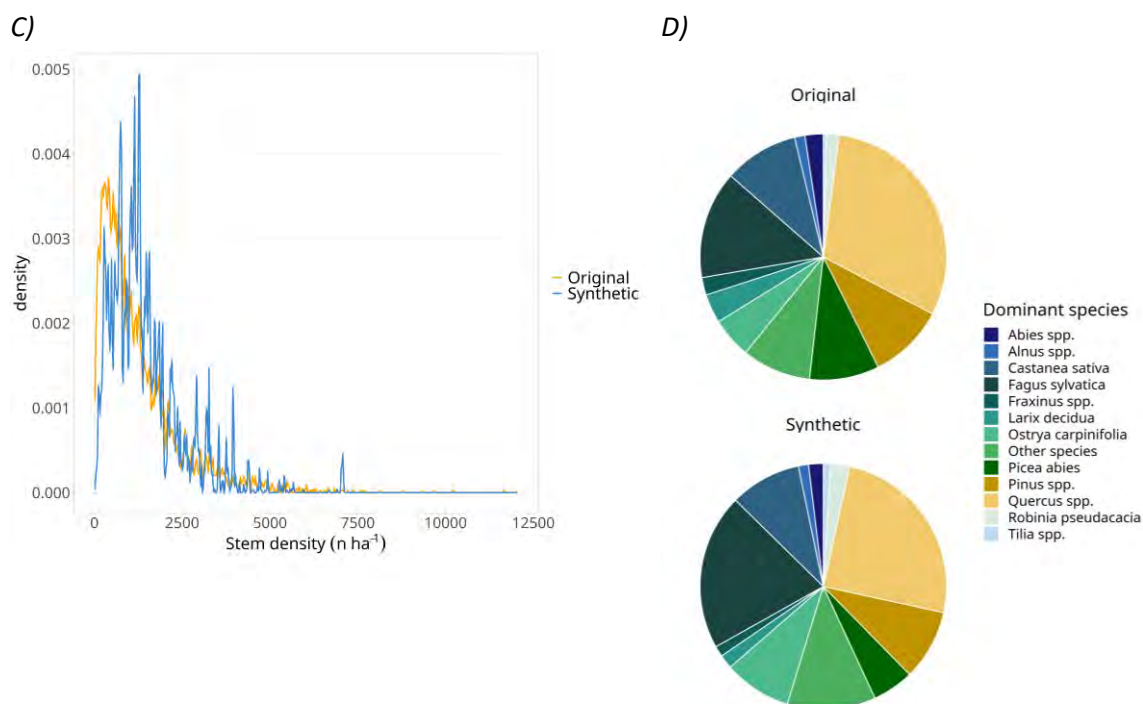


FIGURE 44: DENSITY PLOTS AND PIE CHART OF ORIGINAL AND SYNTHETIC DATA. ORIGINAL DATA = INFC2015, SYNTHETIC DATA = NATIONAL/EUROPEAN MAPS. VARIABLES ARE A) MEAN DBH, B) BASAL AREA, C) STEM DENSITY AND D) DOMINANT SPECIES. DOMINANT SPECIES ARE SUMMARIZED IN 'OTHER SPECIES' IF NUMBER OF OCCURRENCES IS LOW.

Conclusion

Exchanging the European volume map with a nationally derived volume map slightly improved the results, although still showing an overestimation of stem density and basal area. This raises the question whether both plot-level volume maps are generally overestimating plot volume and therefore result in too high values of basal area and stem density. Similarly to IT-1, the results are promising, but likely not accurate enough for EFISCEN-Space runs.

Experiment IT-3: Does the inclusion of topographic layers improve the results?

Methods

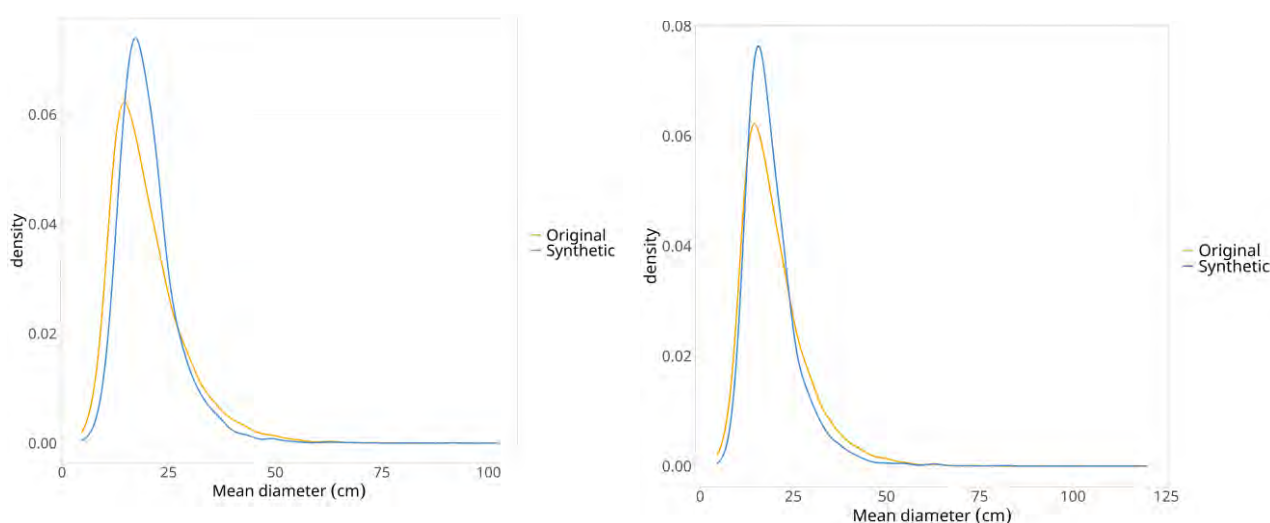
Because of the heterogeneous relief in Italy, experiment IT-3 tests the inclusion of additional layers, such as elevation, slope and aspect, to improve accuracy of the results. We use the elevation map from Global Multi-resolution Terrain Elevation Data (GMTED2010), which is also used in EFISCEN-Space, at a 250-meter spatial resolution and calculated the slope and aspect in RStudio with the function *terrain* of the *terra* package (v 1.7-65; Hijmans, 2025). We test different combinations of the additional layers to evaluate the impact of adding topographic layers on the final results.

Results

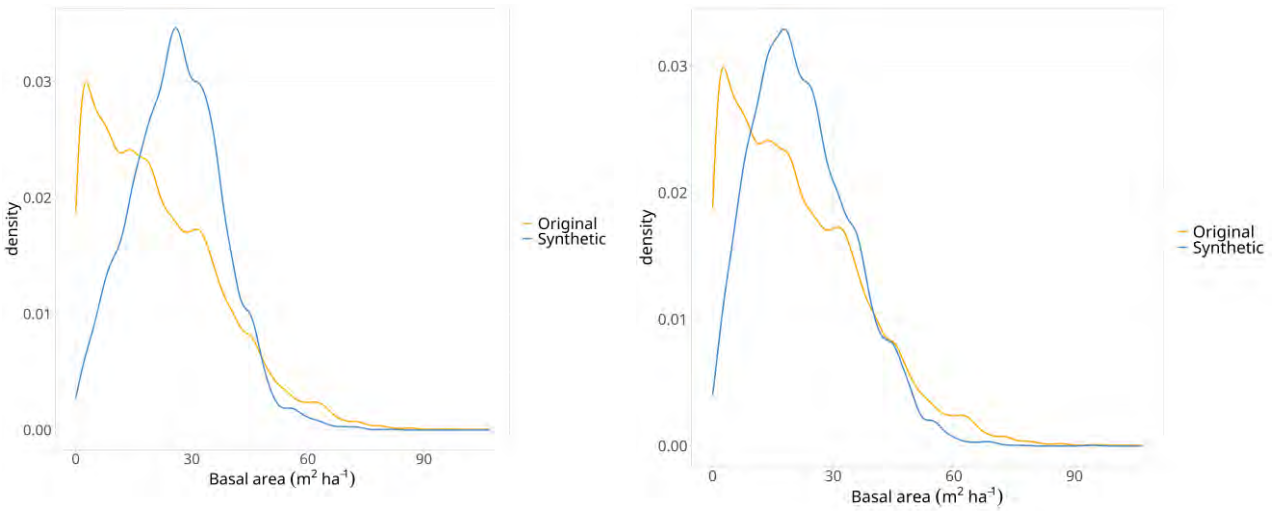
By using all three additional layers, we can achieve a slight improvement in the allocation of basal area while keeping all other input data the same (with data of experiment IT-1: %RMSE 82.45; with data of experiment IT-2: %RMSE 73.5) (Figure 45B). Using any of the topographic layers individually did not improve the results for basal area. Similarly, the synthetic dbh using all three topographic layers show a slightly better alignment (with data of experiment IT-1: %RMSE 46.11; with data of experiment IT-2: %RMSE 44.41) (Figure 45A). Although also for dbh only the combined usage of the topographic layers achieves an improvement. The same can be said for stem density (with data of experiment IT-1: %RMSE 109.78; with data of experiment IT-2: %RMSE 108.34) (Figure 45C).

We can see a considerable improvement for the allocation of dominant tree species when using all three topographic layers, increasing the accuracy to 34.6% and the Cohen's Kappa value to 28.8% with data of experiment IT-1, and accuracy to 39.4% and Cohens' Kappa to 34.2% with data of experiment IT-2 (Figure 45D). Again, the combination of all three topographic layers shows the best improvement of results, however using only elevation did also increase the accuracy of dominant species allocation (e.g., for data of experiment IT-1: accuracy = 18.5% and Cohen's Kappa = 11.2%).

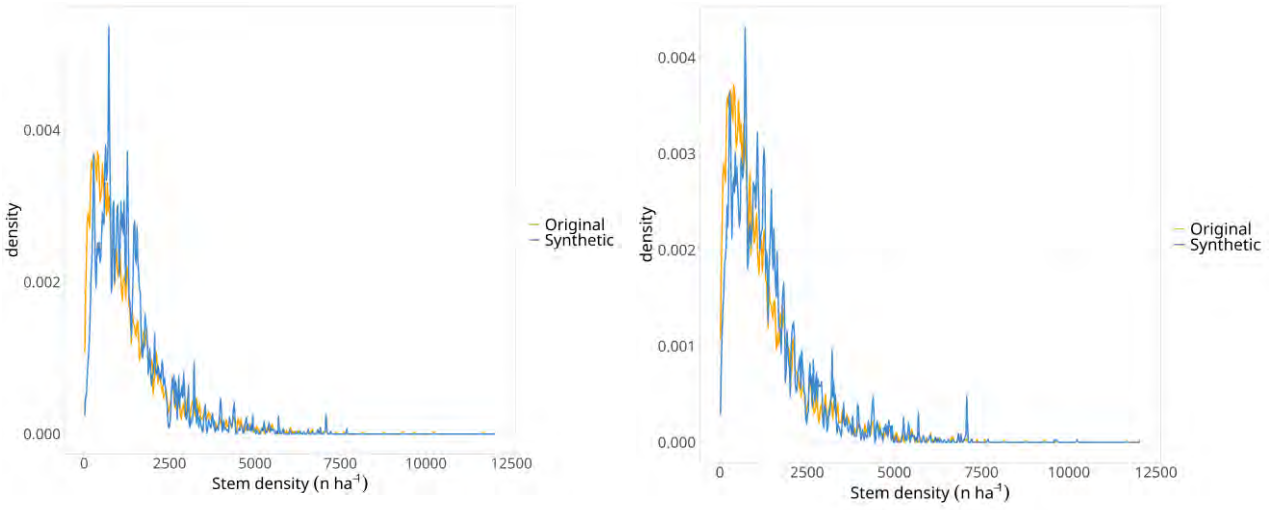
A)



B)



C)



D)

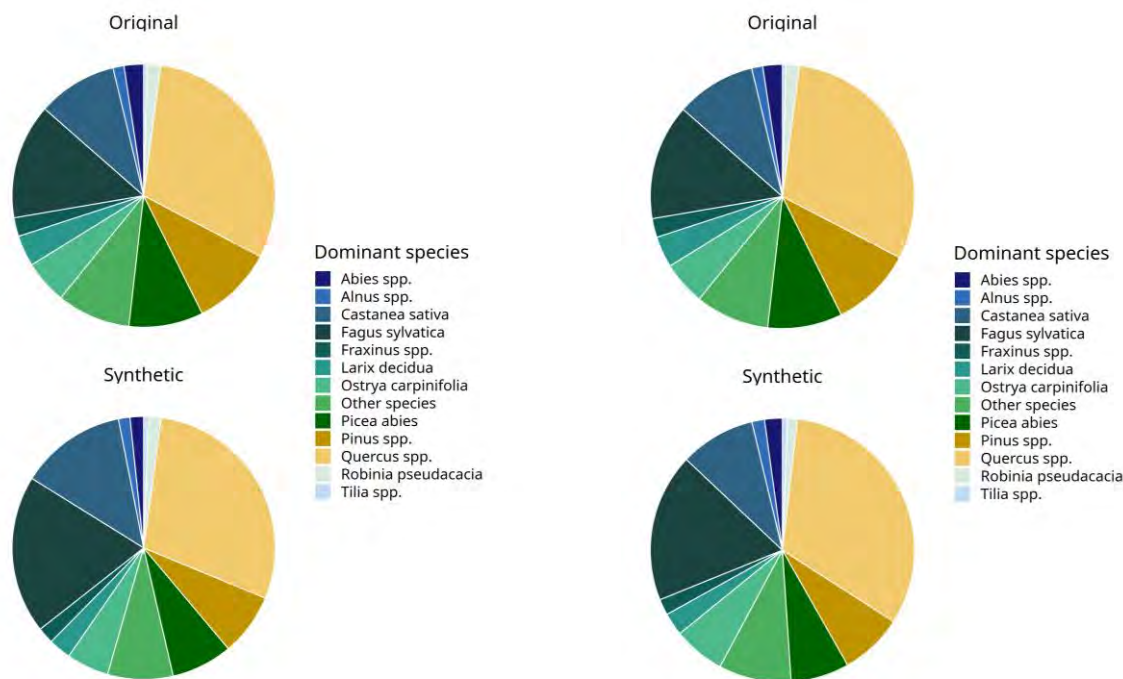


FIGURE 45: DENSITY PLOTS AND PIE CHART OF ORIGINAL AND SYNTHETIC DATA. ORIGINAL DATA = INFC2015, SYNTHETIC DATA: LEFT COLUMN = EUROPEAN MAPS + TOPOGRAPHIC, RIGHT COLUMN = NATIONAL/EUROPEAN MAPS + TOPOGRAPHIC. VARIABLES ARE A) MEAN DBH, B) BASAL AREA, C) STEM DENSITY AND D) DOMINANT SPECIES. DOMINANT SPECIES ARE SUMMARIZED IN 'OTHER SPECIES' IF NUMBER OF OCCURRENCES IS LOW.

Conclusion

Including topographic layers had a positive impact on both structural variables and dominant species composition. Especially elevation seems to improve results, indicating that plot structure and composition differ predominantly along the elevation gradient. This shows that adding abiotic layers to the input variables when creating synthetic datasets can have a positive impact on the results.

Experiment IT-4: How do the synthetic data sets influence results of EFISCEN-Space runs, compared to the original data?

Methods

To evaluate the impact of synthetic data sets on EFISCEN-Space outputs, we set up runs which simulate forest development based on observed harvest patterns. Harvest patterns are identified from information coming from repeated inventory measurements (Feliciano et al., 2025). In addition to harvest, EFISCEN-Space uses empirical equations to simulate growth and natural mortality (Schelhaas et al. 2018). The baseline run is conducted with the original NFI data, and additional runs are conducted individually with each generated synthetic dataset while all other settings, e.g. harvest pattern, growth and mortality equations, stay the same. Runs are conducted for the period 2015-2065.

Results

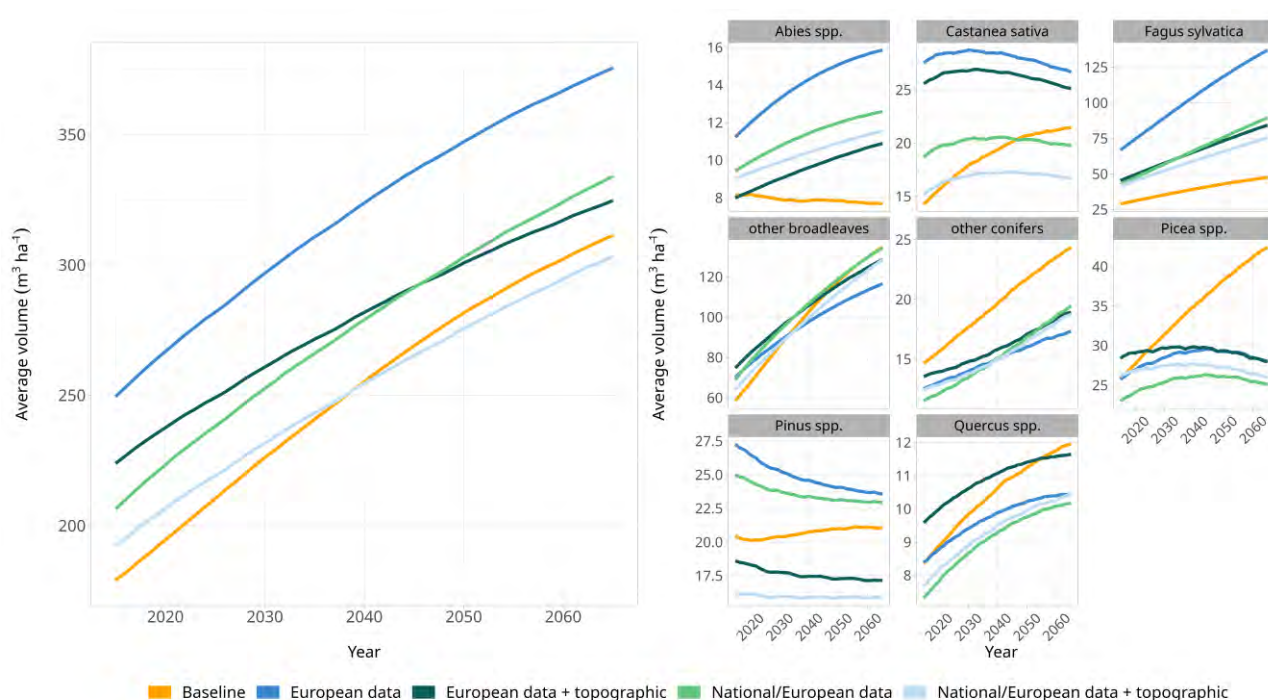


FIGURE 46: PROJECTED OVERALL AVERAGE VOLUME (LEFT) AND AVERAGE VOLUME DEVELOPMENT PER SPECIES GROUP (RIGHT). LINES SHOW RESULTS OF EFISCEN-SPACE RUNS WITH NFI DATA (BASELINE) AND RUNS WITH THE DIFFERENT SYNTHETIC DATASETS. THE GRAPHS TO THE RIGHT DEPICT THE CONTRIBUTION OF EVERY SPECIES TO THE NATIONAL AVERAGE VOLUME. DATASET DESCRIPTIONS IN THE LEGEND REFER TO THE SYNTHETIC DATASETS FROM PREVIOUS EXPERIMENTS AS FOLLOWS: EUROPEAN DATA = IT-1, EUROPEAN DATA + TOPOGRAPHIC = IT-3, NATIONAL/EUROPEAN DATA = IT-2, NATIONAL/EUROPEAN DATA + TOPOGRAPHIC = IT-3.

Compared to the baseline scenario, the EFISCEN-Space run with synthetic data created with only European maps overestimates average volume at the starting year and maintains the higher level throughout the entire period. The same can be seen for European maps combined with topographic layers, and the combined national/European maps, however here the overestimation is less prominent. Nevertheless, the graph shows that the run with national/European maps follows the baseline in a similar shape, compared to other runs. National/European maps combined with topographic layers results in the best outcome, i.e. the closest to the baseline (Figure 46).

The right side of the figure (Figure 46) shows the tree species development of average volume over time. Most broadleaves follow a similar pattern across the different runs, although we can see differences in the starting values (*Fagus sylvatica*, *Quercus* spp., other broadleaves). Volume values for *Castanea sativa* have a much longer and prominent increase for the baseline run while synthetic values show a decrease in volume after a short initial increase. Conifers show more obvious differences between baseline and synthetic data runs. *Abies* spp. stays constant in the baseline run, while the synthetic data runs show an increase. *Picea* spp. shows the most prominent difference, with baseline values showing a considerable increase and synthetic data values a decrease over time. Overall, considering the species distribution and development, it is however not clear which synthetic data run results in the most similar outcome compared to the baseline run.

Main findings & Conclusions for case study 3

Using the same approach with using plot-level dataset in Italy, as we did in Finland and the Netherlands, we can see that the results are less accurate than in Finland and the Netherlands, supporting the previous conclusion that higher structural diversity lowers the accuracy of synthetic data sets (IT-1). Especially the assignment of dominant tree species seems to be much more difficult in Italy, based on our experiments to create synthetic plots. The overestimation of basal area likely comes from the overestimation of stem density, which could not be used as one of the input variables. Since the dbh distribution of original data and synthetic data is very similar in all experiments, it is more likely that the higher basal area is a result of a higher stem density rather than of an increased dbh variance. Also using a national map as plot-level input that was partially created with the tree-level data did not considerably improve the results. This could indicate that both volume maps, European and national, are overestimating volume (as the only structural input variable used in the experiments). Adding topographic information improves the results, especially the accuracy of dominant species, but also of stem density and basal area. Because of Italy's diverse landscape and terrain, the topographic layers are supporting a better allocation of dominant species to their respective suitable terrain and therefore improving the accuracy. The improved structural allocation could be a result of the improved species allocation as well as a generally better allocation of plots along Italy's varied landscape.

EFISCEN-Space runs show what can already be seen in the previous results: average volume in synthetic datasets is overestimated compared to the baseline run with original NFI data, although the development of average volume follows a similar pattern between baseline and synthetic data runs. The results per species group show a few differences, already in the initial species distribution, but then also in the further development. For those species groups for which the development over time shows lower volume in the synthetic data runs than in the baseline run, we can argue that the species groups in the synthetic datasets have been placed in less suitable areas. In general, it is important to keep in mind that the most detailed species information that is available is on tree genus level and that, in addition, conifer proportion is used as variable which means that similar species could be easily wrongly allocated, as it might for example be the case for *Abies* spp. and *Picea* spp. Overall, we can conclude that the synthetic data approach for Italy looks promising, but additional information, especially for species allocation, could help to improve the results.

3.2.4. Discussion and Conclusions of the case-studies

Main findings

The purpose of this chapter was to develop a method for synthesizing spatially explicit tree-level forest inventory data, which could be used **i)** to increase the density of existing data or **ii)** to create data for regions where there is no data.

1. Which inputs are essential in the procedure of creating synthetic plots?

Case study NL-1 showed that the best variables to have in order to create a good synthetic data set are two out of: basal area, quadratic mean dbh, and stem density, as well as species (group) information. In the absence of basal area, volume can act as a good alternative, although this only works when the relationship between volume and basal area is consistent between the tree-level and plot-level data set (NL-1 vs NL-4). Case study FI-1 showed that using the mean plot dbh instead of the quadratic mean dbh still leads to reasonable results. Case IT-1, IT-2 and NL-3 confirmed that it is difficult to get the structural composition of the synthetic plots right, when only using one structure related variable.

2. How well can we do with available national data sets?

Using two different sources of field inventory data leads to quite good results (NL-1, NL-4 and FI-1). The drawback of this is that inventory data is more difficult to obtain than remote sensing products and this method can't be used for the aim of increasing the density of existing inventory data.

Using currently available remote sensing-based national maps (FI-1, FI-3, NL-2 and IT-2) leads to high errors in the structural variables, although the results for Finland and to some extent Italy are better than for the Netherlands. For all cases explored in this chapter it is questionable if the resulting synthetic data sets are good enough to create a good representation of the study region, however.

3. How well can we do with European maps?

Case studies NL-3, FI-2 and IT-1 tested the synthetic procedure with currently available European maps as plot-level data. In the Dutch case, the use of European structure maps for creating synthetic datasets resulted in noticeably lower accuracy than the national maps, and in the Italian case the results of national vs European maps were very similar. This probably has to do with map quality, on the one hand, and the available variables, on the other hand. Only in the Finnish case, the use of European structure maps yielded synthetic datasets of similar (even a little better) accuracy than the national maps. The most likely reason for this, is that the Finnish forest structure is generally less complex and variable than the Dutch and Italian, and therefore easier to predict. Another reason could be that Finnish data were used to inform the model to create these maps, while Dutch data were not (Miettinen et al., 2024). However, the fact that Italian data were also used for this speaks against this theory. What the Finnish case does show, is that the quality of a data set or map is probably more important than if it is national or European. The European genus map (de Keersmaecker et al., 2025) returned good results for Finland and the Netherlands, with 80% accuracy for the dominant species (FI-2) and 60% for the species group (NL-3). In Italy (IT-1), the results for the dominant species were less good, possibly due to a higher species diversity.

4. Can we use tree-level data from neighbouring countries?

Case study NL-4 showed that even if the borrowed data are differently distributed, perfect plot-level data will still lead to a near-perfect match of tree-level data. Similarly, comparing case FI-1 where Finnish tree-level data are used to case FI-3 where Swedish tree-level data are used, showed that the origin of the tree-level data does not change the results significantly. There are only two requirements to the tree-level data: **i)** the range of the structure variables must not be narrower than of the plot-level data and **ii)** the same species groups must occur. Furthermore, when substituting one variable for another (the example of volume-for-basal-area substitution), the relationship between the substituting and substitution variable must be constant over the data sets.

5. Does the inclusion of abiotic variables improve the results?

In case NL-1 and IT-3 the inclusion of abiotic variables for the improvement of the synthetic data sets was tested. In case NL-1, it was concluded that adding soil information does not change the allocation of structural variables, while it only slightly improves the assignment of dominant tree species. In case IT-3, we saw that the inclusion of elevation, slope and aspect in a study area with heterogeneous relief slightly improves the allocation of structural variables, and considerably improves the allocation of dominant tree species. Therefore, the use of abiotic variables generally seems most useful to improve the allocation of tree species.

6. How do the synthetic data sets influence the results of EFISCEN-Space, compared to runs with original inventory data?

The results of experiment IT-3 show that if the synthetic data set is biased (total growing stock too high or too low) this error will propagate in the outputs. Although the initial starting values differ between the datasets, the temporal development of growing stock follows a similar pattern, indicating that species-specific growth dynamics are consistent. Results from experiment IT-3 further show that, even when the initial growing stock of individual species in the synthetic datasets matches the baseline, their development over time can diverge considerably if, for example, species are assigned to geographic locations with unsuitable environmental conditions. In conclusion, when we want to use synthetic data sets for EFISCEN-Space runs, it is important to initialize with a similar overall growing stock and to ensure that tree species are matched to suitable geographical locations. The inclusion of abiotic variables can help with this.

Implications

One of the main takeaways from these experiments is that the quality of the synthetic data set is mainly limited by the quality of the plot-level product that is used. We see that the best synthetic data sets are created by using two independent sources of field inventory, as in case NL-1, NL-4, and FI-1. This is useful when a country has an NFI, but does not provide access to the tree-level data. However, this is also the most difficult data to access, due to the requirement of exact plot locations. Additionally, these data are not spatially continuous, which makes them inappropriate for the purpose of increasing the density of existing data.

Remote sensing interpretations of inventory variables, on the other hand, have the advantage of spatial continuity and can therefore be used to synthesise a higher density of forest inventory plots. Based on our results, we conclude that any maps with a high accuracy and a detailed spatial resolution should be preferred. Moreover, map products should be known to be accurate for the region of interest, which is more likely to be the case if local data has been used to create the product; although in our case studies we saw no clear

evidence that national maps perform better than European maps. From the results of the EFISCEN-Space runs in IT-3, it mainly appears that it is important to use unbiased data. If a dbh distribution is too narrow this is likely to even itself out over time, but a bias in the growing stock or mis-matched tree species will not disappear. In topographically heterogeneous areas, the inclusion of abiotic variables can mitigate the latter point.

In this chapter we explored a method to improve the completeness and density of data available to run EFISCEN-Space. However, the data sets that were used do not allow for timely updates of tree-level data (e.g. a disturbance has taken place but no new inventory data is available yet). To address this temporal gap, data products that will be updated at regular intervals would be preferred. A synthetic data method that is not dependent on compiled maps but uses remote sensing products which have been processed at a lower level could be explored.

Finally, the selection of variables that are available in both plot- and tree-level data make a difference in the quality of the synthetic data set. Ideally, the structural variables should include two variables out of basal area, quadratic mean diameter (qmd), and stem density, because these are most directly linked to the inputs of EFISCEN-Space. Basal area can be replaced by volume and possibly above ground biomass, if the relationship between these variables is relatively similar in both data sets. Qmd can also be replaced by the more commonly available mean dbh. For species related variables, a species-group product like the European genus map is recommended, because it contains similar groups to the EFISCEN-Space model. Abiotic variables like elevation, slope and aspect can offer further improvement, especially in mountainous regions.

3.3. Other developments in EFISCEN-Space

In addition to the development of a synthetic data approach, we extended EFISCEN-Space functionalities to allow the simulation of natural disturbances (Patacca et al. 2025; Forzieri et al. 2021) and forest management (Feliciano et al., 2025; Filipek et al., 2025). These developments allow for forest development simulations across European countries, including various management scenarios, climate change scenarios and disturbance simulations.

For fire and wind disturbances, we implemented a procedure that allows to use two different approaches to estimate forest vulnerability to disturbances. The first approach builds on the vulnerability approach developed in the RESONATE project by Patacca et al. (2025). The authors implemented an approach for wind disturbances using the ForestGALES impact model. ForestGALES (FGR, Gardiner et al., 2008) is a hybrid empirical-mechanistic model (Hale et al., 2015) and one of the most widely used tools for simulating wind effects on forest stands. It incorporates both stand and tree level parameters to calculate the Critical Wind Speed (CWS) required to damage a tree within a stand (Locatelli et al., 2016). Wildfire impact is modelled at the stand level and evaluated based on 1) the climatic conditions determining the degree of flammability of the available fuel at the forest floor, 2) the structural characteristics of the forest and 3) tree species specific characteristics. At each annual time step, fuel load is derived from leaves and woody biomass, and the FWI. FWI is the Canadian Fire Weather Index of the plot during the European fire season (June, July, August). The FWI captures drought-driven changes in fuel flammability. Finally, fire induced mortality is quantified as the percentage (%) of crown loss per tree, expressed as a function on fuel load (fire intensity) and tree size (using DBH).

The second approach builds on the random decision forest models developed by Forzieri et al. (2021) to estimate disturbance vulnerability, based on specific forest, climate and landscape parameters. This approach is also used in the Eco2Adapt project. In the approach, vulnerability, defined as relative biomass loss, is then combined with the probability of a disturbance event and the actual available biomass per plot to determine the absolute biomass loss per plot per time step. The approach is implemented in EFISCEN-space in collaboration with the Eco2Adapt project but is still in a testing phase.

3.4. References

- Aussenac, R.**, Monnet, J. M., Klopčič, M., et al. (2023). Diameter, height and species of 42 million trees in three European landscapes generated from field data and airborne laser scanning data [version 1; peer review: 2 approved with reservations]. Open Research Europe, 3, 32. <https://doi.org/10.12688/openreseurope.15373.1>.
- De Keersmaecker, W.**, Zanaga, D., Senf, C., Viana-Soto, A., Klapper, J., Blickensdörfer, L., Govaere, L., Lerink, B., Leyman, A., Schelhaas, M.-J., Teeuwen, S., Verkerk, P. J., & Van De Kerchove, R. (2025). European Tree Genus Map 2020 (Early Access Release) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.13341104>
- De Laurentis, D.**, Papitto, G., Gasparini, P., Di Cosmo, L., & Floris, A. (2021). Italian Forests—Selected Results of the Third National Forest Inventory INFC2015. Council for Agricultural Research and Economics, Research Centre for Forestry and Wood (CREA-FL).
- Feliciano, D.**, Franzini, F., Schelhaas, M.J., Haltia, E., Bacciu, V., Boonen, S., Filipek, S., Häyrinen, L., Lindner, M., Menini, A., Nieberg, M., Ofoegbu, C., Peltoniemi, M., Stancioiu, T., Staritsky, I., Uzquiano, S., Wiersma, H. (2025). Decision rules, parameters, and narratives for modelling. ForestPaths project deliverable D1.3.
- Filipek, S.** et al. 2025. Holistic management practices, modelling and monitoring for European forest soils, HoliSoils Del 6.3.
- Finnish Forest Centre (Metsäkeskus).** (2025). Forest resource lattice data (Hila-aineisto) [2019–2021]. Retrieved from <https://www.metsakeskus.fi>.
- Forzieri, G.**, Girardello, M., Ceccherini, G. et al. Emergent vulnerability to climate-driven disturbances in European forests. Nat Commun 12, 1081 (2021). <https://doi.org/10.1038/s41467-021-21399-7>.
- Francini, S.** (unpublished, A). Dutch national forest structure maps.
- Francini, S.** (unpublished, B). Italian national forest map of growing stock.
- Fridman, J.**, Holm, S., Nilsson, M., Nilsson, P., Ringvall, A. H., & Ståhl, G. (2014). Adapting National Forest Inventories to changing requirements – the case of the Swedish National Forest Inventory at the turn of the 20th century. Silva Fennica, 48(3), Article 1095. <https://doi.org/10.14214/sf.1095>.
- Gardiner, B.**, Byrne, K., Hale, S., Kamimura, K., Mitchell, S.J., Peltola, H., Ruel, J-C. (2008). A review of mechanistic modelling of wind damage risk to forests. Forestry 81: 447-463.
- Gasparini, P.**, Di Cosmo, L., Floris, A., & De Laurentis, D. (2022). Italian National Forest Inventory—Methods and Results of the Third Survey: Inventario Nazionale delle Foreste e dei Serbatoi Forestali di Carbonio—Metodi e Risultati della Terza Indagine (p. 576). Springer Nature.
- Gower, J. C.** (1971). A general coefficient of similarity and some of its properties. Biometrics, 27(4), 857–871.

- Hale, S.E., Gardiner B., Peace, A., Nicoll, B., Taylor, P. and Pizzirani, S., 2015.** Comparison and validation of three versions of a forest wind risk model. *Environmental Modelling and Software* 68, 27-41.
- Hijmans R (2025).** terra: Spatial Data Analysis. R package version 1.7-65, <https://rspatial.org/>.
- Kangas, A., Pitkänen, T. P., Mehtätalo, L., & Heikkinen, J. (2023).** Mixed linear and non-linear tree volume models with regional parameters to main tree species in Finland. *Forestry*, 96(2), 188-206.
- Korhonen, K. T. (2023).** National Forest Inventory 12 (Version 1). Natural Resources Institute Finland. <https://doi.org/10.23729/69e33efe-aefc-47df-ad37-f268389b56a9>.
- Korhonen, K. T., Rätty, M., Haakana, H., Heikkinen, J., Hotanen, J.-P., Kuronen, M., & Pitkänen, J. (2024).** Forests of Finland 2019–2023 and their development 1921–2023. *Silva Fennica*, 58(5), Article 24045. <https://doi.org/10.14214/sf.24045>.
- Kramer, H. and S. Los. (2022).** *Basiskaart Natuur 2021; een landsdekkend basisbestand voor de terrestrische natuur in Nederland*. Wageningen, Wettelijke Onderzoekstaken Natuur & Milieu, WOTtechnical report, Wageningen UR, Wageningen, Nederland.
- Lesschen, J.P., H.I.M. Heesman, J.P. Mol-Dijkstra, A.M. van Doorn, E. Verkaik, I.J.J. van den Wyngaert and P.J. Kuikman. (2012).** *Mogelijkheden voor koolstofvastlegging in de Nederlandse landbouw en natuur*. Alterra-rapport 2396. Alterra Wageningen UR, Wageningen, The Netherlands. <http://edepot.wur.nl/247683>.
- Locatelli, T., Gardiner, B., Tarantola, S., Nicoll, B., Bonnefond, J. M., Garrigou, D., ... & Patenaude, G. (2016).** Modelling wind risk to *Eucalyptus globulus* (Labill.) stands. *Forest Ecology and Management*, 365, 159-173.
- Miettinen, J., Adame, P., Adolt, R., Alberdi, I., Antropov, O., Arnarsson, Ó., Astrup, R., Berger, A., Bogason, J., Chirici, G., Corona, P., D'Amico, G., Fejfar, J., Fischer, C., Gohon, F., Gschwantner, T., Hertzler, J., Koma, Z., Korhonen, K. T., ... Breidenbach, J. (2024).** High-resolution Pan-European forest structure maps: An integration of Earth observation and National Forest Inventory data. Zenodo. <https://doi.org/10.5281/zenodo.13143235>.
- National Forest Inventory, SLU. (2025).** Unpublished data. Department of Forest Resource Management, Swedish University of Agricultural Sciences, Umeå. <http://www.slu.se/en/nfi>.
- Natural Resources Institute Finland (LUKE). (2021).** The Multi-source National Forest Inventory Raster Maps of 2021.
- Natural Resources Institute Finland (LUKE). (2024).** Growing stock volume on forest land and poorly productive forest land by tree species by inventory, region and tree species, NFI12. LUKE Statistics Database. Retrieved from <https://stat.luke.fi>.

- Patacca**, M., Schelhaas, M.-J., Bozzolan, N., Filipek, S., Staritsky, I., v.d. Sande, M., Nabuurs., G.-J. (2025). Deliverable 4.9. Report on EU level forest and wood value chain resilience and assessed by three scenarios. Horizon 2020 project RESONATE, project no. 101000574, Wageningen University & Research.
- Pucher**, C., Neumann, M., & Hasenauer, H. (2022). An improved forest structure data set for Europe. *Remote Sensing*, 14(2), 395. <https://doi.org/10.3390/rs14020395>.
- Probos** (2024). Unpublished data. Systeem Houtmeetkundige Inventarisatie. <https://www.probos.nl/projecten/gegevensvoorziening/3830-syhi>
- Riedel** T., Hennig P., Kroiher F., Polley H., Schmitz F., Schwitzgebel F. (2017): Die dritte Bundeswaldinventur (BWI 2012). Inventur- und Auswertemethoden, 124 S
- Schelhaas**, M. J., Teeuwen, S., Oldenburger, J., Beerkens, G., Velema, G., Kremers, J., ... & Scholten, T. (2022A). Zevende Nederlandse Bosinventarisatie. *WUR Rapport*.
- Schelhaas**, M.J., Hengeveld, G., Filipek, S., König, L., Lerink, B., Staritsky, I., de Jong, A., Nabuurs, G.J., (2022B). EFISCEN-Space 1.0 model documentation and manual. Wageningen, Wageningen Environmental Research, Report 3220. <https://doi.org/10.18174/583568>.
- van der Loo**, M. (2022). gower (version 1.0.1). <https://rdocumentation.org/packages/gower/versions/1.0.1>.

4. Improved inputs for process-based projections of future climate change impacts on forests

4.1. Introduction

This chapter describes updates to the LPJ-GUESS demographic vegetation model to allow improved initialisation of current forest structure as a base for simulations of future climate and management impacts. The updates are made in the context of a systematic upgrade of the LPJ-GUESS model to better represent the state and dynamics of European forests that has been made with contributions from several projects (in addition to FORWARDS, principally the Horizon Europe projects ForestPaths, CLIMB-Forest and wildE and the ERA-Net project FORECO).

One set of major advances concerns improved initialisation of the forest vegetation in LPJ-GUESS using either stand age distributions from National Forest Inventory (NFI) or by directly initialising the tree size and species composition based on tree-by-tree NFI observations (Pugh et al., 2024). This was a major step forward in being able to represent realistic forest structure, but the process can only be updated to the most recently available census of NFI data, which can sometimes be a decade or more in the past. In FORWARDS we have developed a process to update the stand age to the most recent year based on satellite observations of forest disturbances and harvest. Satellite observations from Landsat are capable of providing a timeseries of forest disturbances (attributed to disturbance agents) and clearcut harvest at annual temporal resolution and ca. 30 m spatial resolution. This allows us to replace the prognostic forest disturbance and harvest modules with a definitive timeseries of such events over the period between the date of forest initialisation from NFI (currently 2010) and the present day. As these large events are major factors controlling forest structure, this should allow a much more faithful representation of current forest structure than would be obtained by running the prognostic disturbance modules alone. As the satellite disturbance products developed in WP2 of FORWARDS are not yet pan-European and attributed by disturbance agent, we used the pan-European disturbance maps by Viana-Soto and Senf (2024, 2025) that were developed in the ForestPaths project. The approach applied here enhances the capability to use LPJ-GUESS to add value to monitoring efforts and forward-looking projections by integrating the satellite observations with NFI observations of harvest (Suvanto et al., 2025; Pugh et al., 2024) and propagating this information into carbon cycling assessments – an application that is core to the overall goals of FORWARDS.

In the next sections, we describe the implementation of the new disturbance and harvest scheme based on satellite observations and then provide some example results from the new system. Because this work became a joint effort with the ForestPaths project, the model and method description below have also been included in similar form in ForestPaths Deliverable 3.1 (Pugh et al. 2024).

4.2. Description of LPJ-GUESS relevant to the implementation

LPJ-GUESS is a dynamic global vegetation model which has been widely applied over the last 25 years for studies of the terrestrial biosphere and carbon cycling (Smith et al., 2001, 2014). The model includes detailed representations of forest, grassland, croplands, pasture, peatlands and arctic vegetation, as well as the land-use changes between them (Bayer et al., 2017; Rabin et al., 2020). LPJ-GUESS simulates vegetation and soils as a series of pools of carbon and nitrogen, whose sizes are governed by the fluxes between them and between the pools and the atmosphere. The processes of photosynthesis and respiration, along with updates of hydrology, are carried out on a daily timestep, whilst vegetation growth and mortality are assessed annually. The basic processes of photosynthesis, respiration, carbon allocation and vegetation dynamics are described in Smith et al. (2014). These processes are forced by externally-provided boundary conditions of climate, atmospheric CO₂ mixing ratio, nitrogen deposition and soil type. This information is provided in grid cells, whose size is primarily determined by the spatial scale of the boundary conditions provided. Whilst global simulations are typically run at 0.5° x 0.5°, higher resolution simulations can also be run regionally (e.g., 3km, Lagergren et al., 2024), or even at the local catchment scale (e.g., 50m, Gustafson et al., 2021).

Within a grid cell, the basic unit of an LPJ-GUESS simulation is the 0.1 ha patch. Within patches, age cohorts of different plant functional types (PFTs) compete for light, water and nitrogen. In the European version of LPJ-GUESS, vegetation is represented by 24 plant functional types, 20 of which are directly parameterised based on major tree species (Hickler et al., 2012; Lindeskog et al., 2021). Because the processes of vegetation establishment and mortality have a strong stochastic component, multiple replicate patches are typically simulated to provide a realistic representation of the range of trajectories that patches might take. Replicate patches are forced by the same external boundary conditions of climate, soil type, atmospheric CO₂ mixing ratio and nitrogen deposition. Patches themselves are contained within “stands”, with the group of patches forming a representative sample of the vegetation within the stand. Stands loosely conform to an area of trees established at the same time, similar to the terminology in forestry. However, in LPJ-GUESS, all vegetation areas of the same type and management rules, established at the same time, are spatially aggregated within a grid cell and are contained within the same stand. Whilst patch size is fixed (as patches are samples, analogous to forest inventory plots), stands are initiated covering a fraction of a grid cell, which can be reduced over time as a result of land-use change or management interventions.

LPJ-GUESS includes a wide range of possible management interventions including thinning, clearcuts and planting according to user-specified rules (Lindeskog et al., 2021). It has also recently been updated to include a new approach where the harvest rate and intensity is defined probabilistically based on statistics from NFI observations (Pugh et al., 2024; Suvanto et al., in prep.). LPJ-GUESS contains prognostic modules for stand-replacing (or potentially stand-replacing) disturbances by fire (Rabin et al., 2017; Nieradzick et al., in prep.), wind throw (Lagergren et al., 2012) and spruce bark beetle (Lagergren et al., 2025). These modules have recently been updated as part of the wider European update of the model (Pugh et al., 2024).

4.3. Implementing satellite-driven disturbances in LPJ-GUESS

We added a capability to directly force stand-replacing disturbances in LPJ-GUESS with observations of disturbance occurrence from satellite information. The satellite information is currently provided by the product of Viana-Soto and Senf (2024, 2025). This product uses Landsat observations from 1985 to 2023 to provide 30 x 30 m resolution annual wall-to-wall observations of stand-replacing disturbance occurrence across the European continent. Attribution of disturbances between forestry clear-cut, fire and wind/bark-beetle agents is also provided. This product was aggregated to the standard LPJ-GUESS resolution ($0.5^\circ \times 0.5^\circ$), and corrected based on map accuracy statistics, providing fractions of area disturbed by each agent for each grid cell and year. Because the product cannot separate the wind and bark beetle agents we used country-level statistics from Patacca et al. (2023) to split between these agents. The product marks a 30 m pixel as either disturbed or not for a given year. However, a pixel marked as disturbed might not necessarily suffer 100% loss of trees. To account for this, the 0.5° disturbance fractions were combined with an average disturbance intensity derived from a space-for-time substitution using a 2019 biomass map based on Planet imagery, following the procedure of Kowalski et al. (submitted). This intensity was assumed to be characteristic of the disturbance regime in the grid cell and applied multiplicatively to the total disturbance fraction to provide a corrected timeseries.

Within FORWARDS we implemented functions to read this disturbance timeseries into LPJ-GUESS and use it to determine how much of the simulated forest to disturb each year. Each simulation year, all the forest stands within a grid cell are ranked for vulnerability to the different disturbance types. The prescribed forest disturbances were then applied to the stands in the order of vulnerability to the different disturbance types. The different disturbance types were applied in the order: clearcut, fire, wind and bark-beetle mortality. Disturbance vulnerability of stands was determined as follows. For forestry clearcuts, we ranked stands according to their quadratic mean diameter at breast height (DBH), preferentially harvesting stands with higher values. Fire vulnerability was ranked using the BLAZE model (Rabin et al., 2017; Nieradzik et al., in prep.) and wind vulnerability was ranked using the storm module (Lagergren et al., 2012; Pugh et al., 2024). For bark beetles, we assumed that disturbance was primarily spruce bark beetles (Patacca et al., 2023). We thus ranked stands by quadratic mean DBH and further weighted disturbance likelihood according to stand composition, with *Picea abies* individuals given a weight of 100, other needle-leaved trees a weight of 10 and broadleaved trees a weight of 1. It was assumed that all the trees within a disturbed area were killed, with the exception of fire disturbances, where we used the BLAZE model to calculate tree mortality on disturbed areas. Different assumptions can be made about the fate of the disturbed vegetation, depending on the disturbance type. For clearcut harvest we followed the standard model assumption that 90% of the stem biomass and 40% of branches, 30% of leaves and 10% of coarse roots are removed from the ecosystem. For fire, we allowed combustion of vegetation to follow the BLAZE model calculations. For wind and bark beetle disturbances the base assumption is that the disturbed material is salvaged logged, with 75% of the stem biomass, 30% of branches and 5% of coarse roots removed, but it is also possible to specify different settings.

4.4. Example outputs

We ran LPJ-GUESS with the set-up described in Pugh et al. (2024) to initialise stand age to the distribution observed in 2010 based on forest inventory observations. We then applied the new satellite-forced disturbance module over the period 2011-2023. The resulting difference between the tree age structure in 2010 and 2023 is shown in Figure 47. Overall, the total number of simulated trees dropped from 129.8 to 113.5 billion over this period, with the mean age increasing from 38.9 years to 43.3. Such changes can be important in the context of both assessments of biodiversity indicators and carbon sinks.

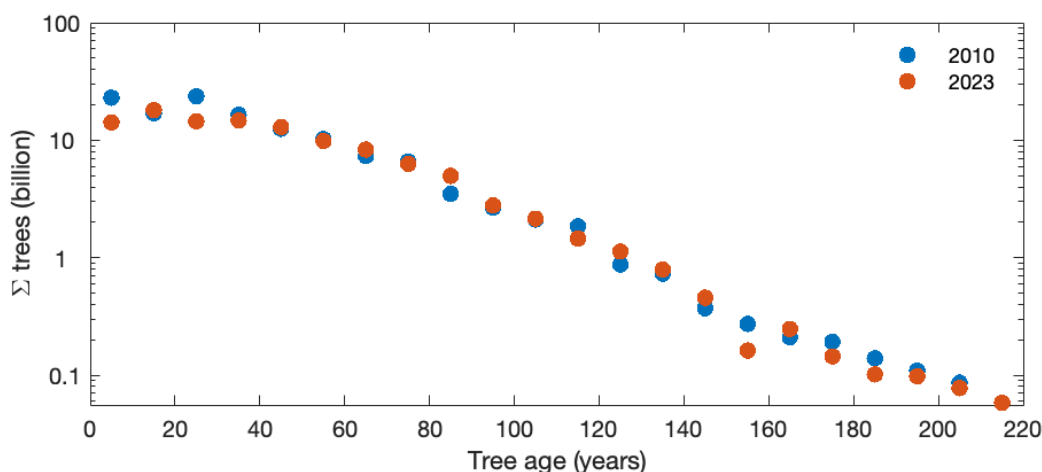


FIGURE 47 TREE AGE STRUCTURE IN 2010 AND 2023 ACROSS 12 EUROPEAN COUNTRIES BASED ON LPJ-GUESS OUTPUT. NOTE THE LOG SCALE ON THE Y AXIS.

Some of the potential of the model results to also be used in assessments of the recent historical period are shown in Figure 48. LPJ-GUESS acts as a consistent integrator for stand-level forest dynamics observations, based by Landsat, and tree level harvest information, based on NFI, combined with the model's internal process-based estimates of non-observed dynamics. As a result, it is possible to give clear attributions of contributions of different events to carbon cycling. Figure 48 puts in context the level of carbon turnover generated by natural disturbances against that generated by harvest. Only in some limited areas of central Germany and Czechia were disturbances high enough to rival the effect of harvest.

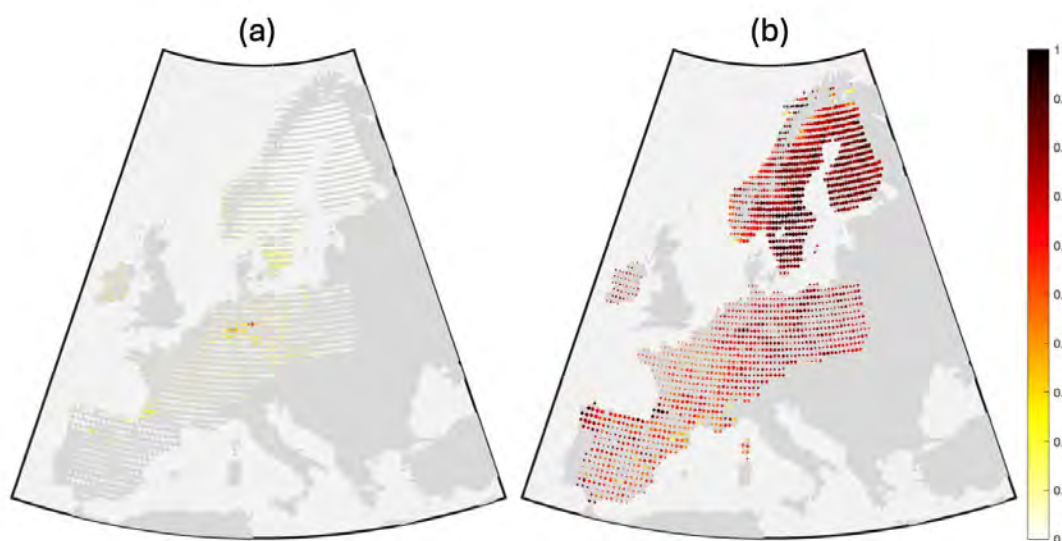


FIGURE 48: FRACTION OF TOTAL CARBON TURNOVER DUE TO NATURAL DISTURBANCES (A) AND HARVEST (B) AVERAGED OVER THE PERIOD 2011-2023, AS ESTIMATED BY LPJ-GUESS FORCED BY SATELLITE AND NFI OBSERVATIONS. HARVEST INCLUDES BOTH CLEARCUT HARVEST AND LESS INTENSIVE EVENTS. PIXEL SIZE IS SCALED BY FOREST AREA.

4.5. Concluding remarks

The new satellite-forced disturbance module in LPJ-GUESS provides capability to update the forest age structure to the most recent year for which disturbance observations are available (currently expected with ca. 1 year lag). This is expected to provide a more realistic base for simulations of future climate and management scenarios.

The new capability also allows LPJ-GUESS simulations to add value to satellite disturbance observations by integrating them with other information on the forest ecosystem and calculating unobserved values such as carbon fluxes. This capability will be explored in a future FORWARDS WP3 deliverable.

We have here used the disturbance product of Viana-Soto and Senf (2024, 2025) as input to the new disturbance module. However, the LPJ-GUESS implementation can work with any similar product calculated at the same spatial resolution. This includes the product currently under development in WP2 of FORWARDS.

4.6. References

- Bayer, A. D., Lindeskog, M., Pugh, T. A. M., Anthoni, P. M., Fuchs, R., and Arneth, A.:** Uncertainties in the land-use flux resulting from land-use change reconstructions and gross land transitions, *Earth System Dynamics*, 8, 91–111, <https://doi.org/10.5194/esd-8-91-2017>, 2017.
- Gustafson, A., Miller, P. A., Björk, R. G., Olin, S., and Smith, B.:** Nitrogen restricts future sub-arctic treeline advance in an individual-based dynamic vegetation model, *Biogeosciences*, 18, 6329–6347, <https://doi.org/10.5194/bg-18-6329-2021>, 2021.
- Lagergren, F., Jönsson, A. M., Blennow, K., and Smith, B.:** Implementing storm damage in a dynamic vegetation model for regional applications in Sweden, *Ecol Modell*, 247, 71–82, <https://doi.org/10.1016/j.ecolmodel.2012.08.011>, 2012.
- Lagergren, F., Björk, R. G., Andersson, C., Belušić, D., Björkman, M. P., Kjellström, E., Lind, P., Lindstedt, D., Olenius, T., Pleijel, H., Rosqvist, G., and Miller, P. A.:** Kilometre-scale simulations over Fennoscandia reveal a large loss of tundra due to climate warming, *Biogeosciences*, 21, 1093–1116, <https://doi.org/10.5194/bg-21-1093-2024>, 2024.
- Lagergren, F., Jönsson, A. M., Lindeskog, M., and Pugh, T. A. M.:** Combining empirical and mechanistic understanding of spruce bark beetle outbreak dynamics in the LPJ-GUESS (v4.1, r13130) vegetation model, *Geosci. Model Dev. Discuss.* [preprint], <https://doi.org/10.5194/gmd-2024-239>, in review, 2025.
- Lindeskog, M., Lagergren, F., Smith, B., and Rammig, A.:** Accounting for forest management in the estimation of forest carbon balance using the dynamic vegetation model LPJ-GUESS (v4.0, r9710): Implementation and evaluation of simulations for Europe, *Geosci Model Dev*, 14, 6071–6112, <https://doi.org/10.5194/gmd-2020-440>, 2021.
- Pugh, T., Schelhaas, M.-J., Arneth, A., Dutca, I., Eckes-Shephard, A., Filipek, A., König, L., Jacobs, S., Jönsson, A. M., Patacca, M., Lerink, B., Lindeskog, M., Munoz Avilez, R., Nabuurs, G.-J., Olin, S., Peltoniemi, M., Piltz, K., Senf, C., Starcevic, A., Staritsky, I., Viana-Soto, A., Wittenbrink, M. (2024).** Enhanced and evaluated forest ecosystem models. ForestPaths project deliverable D3.1.
- Rabin, S. S., Melton, J. R., Lasslop, G., Bachelet, D., Forrest, M., Hantson, S., Kaplan, J. O., Li, F., Mangeon, S., Ward, D. S., Yue, C., Arora, V. K., Hickler, T., Kloster, S., Knorr, W., Nieradzik, L., Spessa, A., Folberth, G. A., Sheehan, T., Voulgarakis, A., Kelley, D. I., Colin Prentice, I., Sitch, S., Harrison, S., and Arneth, A.:** The Fire Modeling Intercomparison Project (FireMIP), phase 1: Experimental and analytical protocols with detailed model descriptions, *Geosci Model Dev*, 10, 1175–1197, <https://doi.org/10.5194/gmd-10-1175-2017>, 2017.
- Rabin, S. S., Alexander, P., Henry, R., Anthoni, P., Pugh, T. A. M., Rounsevell, M., and Arneth, A.:** Impacts of future agricultural change on ecosystem service indicators, *Earth System Dynamics*, 11, 357–376, <https://doi.org/10.5194/esd-11-357-2020>, 2020.
- Smith, B., Wårlind, D., Arneth, a., Hickler, T., Leadley, P., Siltberg, J., and Zaehle, S.:** Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model, *Biogeosciences*, 11, 2027–2054, <https://doi.org/10.5194/bg-11-2027-2014>, 2014.
- Viana-Soto, A. and Senf, C.:** European Forest Disturbance Atlas (Version 2.1.1.), 2024.

Viana-Soto, A. and Senf, C.: The European Forest Disturbance Atlas: a forest disturbance monitoring system using the Landsat archive, *Earth Syst. Sci. Data*, 17, 2373–2404, <https://doi.org/10.5194/essd-17-2373-2025>, 2025.

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or REA. Neither the European Union nor the granting authority can be held responsible for them.