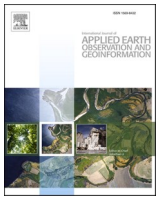
International Journal of Applied Earth Observation and Geoinformation 152 (2026) 105452

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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



Federated learning in forest resource modelling and monitoring: Bridging data confidentiality and collaborative research

Johannes Schumacher [a,*], Alessandro Cescatti [b], Gherardo Chirici [c], Giovanni D'Amico [c], Saverio Francini [d], Johannes Hertzler [e], Lauri Mehtätalo [f], Gert-Jan Nabuurs [g], Mats Nilsson [h], Juho Pitkänen [f], Johannes Breidenbach [a,*]

[a] NIBIO, Norwegian Institute of Bioeconomy Research, National Forest Inventory, NO-1400 Ås, Norway
[b] European Commission, Joint Research Centre (JRC), Ispra, Italy
[c] University of Florence, Department of Agricultural, Food, Environmental and Forestry Sciences and Technologies (DAGRI), Via San Bonaventura 13, IT-50145 Florence, Italy
[d] University of Bologna, Department of Agricultural and Food Sciences (DISTAL), Viale Giuseppe Fanin, 40-50, IT-40127 Bologna, Italy
[e] Thünen Institute of Forest Ecosystems, Alfred-Möller-Straße 1, Haus 41/42, Eberswalde 16225, Germany
[f] Natural Resources Institute Finland (LUKE), Bioeconomy and Environment Unit, Yliopistokatu 6 B, FI-80100 Joensuu, Finland
[g] Forest Ecology and Management Group, Wageningen University, Droevendaalsesteeg 3, 6708 PB Wageningen, the Netherlands
[h] Swedish University of Agricultural Sciences, Department of Forest Resource Management, SE-901 83 Umeå, Sweden

ARTICLE INFO

Keywords:
Forest resource mapping
Satellite remote sensing
Earth observation
Deep learning
Distributed modelling
National forest inventory
Data privacy

ABSTRACT

The availability of reliable ground-truth data is one of the main bottlenecks for improving high-resolution forest attribute maps from Earth observation data. This is underpinned by the European Union (EU) Forest Strategy for 2030 that underscores the need for harmonized, cross-border forest resource assessments that integrate both remote sensing and field-based National Forest Inventory (NFI) data. However, confidentiality constraints on NFI plot coordinates present a significant barrier to aligning these datasets, thereby limiting the development of unified forest monitoring systems that can fully leverage the potential of Earth Observation data. To overcome these data-sharing limitations we explored the effectiveness of a privacy-enhancing technique, known as Federated Learning (FL), that is a form of distributed computing aimed at preserving the privacy and confidentiality of data owned by different organizations. This methodology has been tested for the collaborative modelling and mapping of forest timber volume across four European countries: Norway, Sweden, Finland, and Italy.

We employed a time-series convolutional neural network (CNN) architecture tailored to integrate 40 years of Landsat or 7 years of Sentinel imagery and terrain variables with harmonized NFI data from more than 85,000 sample plots. This model architecture was used for the FL approach and compared to traditional country-specific and centralized modelling strategies.

FL models achieved predictive performances comparable to the traditional models, which proofs the effectiveness of the proposed approach. Centralized or global models showed slightly reduced performance compared to the national models, highlighting the value of fine-tuning with local ground-truth data.

By aligning with the EU's forest monitoring objectives, FL facilitates the generation of harmonized models and maps of forest features, like timber volume and biomass, that are critical to support evidence-based forest policy and management. The findings underscore the potential of FL to transform collaborative environmental monitoring, particularly in domains where data confidentiality and interoperability are critical.

1. Introduction

The availability of reliable and representative ground-truth data remains one of the main bottlenecks for improving high-resolution forest attribute maps derived from Earth observation, limiting both model performance and uncertainty quantification (Abdi and Wang,

* Corresponding authors.
E-mail addresses: Johannes.schumacher@nibio.no (J. Schumacher), Johannes.breidenbach@nibio.no (J. Breidenbach).

https://doi.org/10.1016/j.jag.2026.105452
Received 10 April 2026; Received in revised form 23 June 2026; Accepted 29 June 2026
Available online 1 July 2026

2026; Fassnacht et al., 2023; Kuronen et al., 2025). This is recognized by the European Union (EU) Forest Strategy for 2030, which specifies the needs for a comprehensive forest monitoring framework (European Commission, 2021). This initiative emphasizes the harmonization of data collection across member states and their integration with satellite observations made available for the entire continent by the EU Copernicus program. National governments will be responsible for complementing satellite retrievals with field data collected during National Forest Inventory (NFI) campaigns, with the objective to ensure comparability between countries and increase spatial resolution and timeliness of forest statistics. This approach aims to address existing disparities in data availability, scope and methodologies among NFIs, promoting a more integrated and coherent forest monitoring across the EU. By harmonizing NFI data and reporting standards, the EU seeks to enhance the effectiveness and timeliness of forest monitoring, thereby supporting sustainable forest management and contributing to broader environmental and climate objectives.

The requirement to integrate remote sensing data (e.g., satellite imagery) with NFI data presents a unique challenge due to the confidentiality of plot coordinates in many NFI systems. Plot coordinates in NFIs are often kept confidential to protect landowners' privacy, sensitive ecological data, and prevent any unintended impact on the monitored sites due to vandalism, unauthorized exploitation, or experimental manipulation (US Forest Service, 2023; Schadauer et al., 2024). On the other hand, the fusion of satellite and ground data requires geospatial coordinates to accurately extract remote sensing features corresponding to the same areas where field data is collected. Without access to exact coordinates, aligning the two data sources is not possible, severely limiting the integration of field data with remote sensing data. In particular, the missing fusion of remotely sensed and ground data complicates the ability to create harmonized spatial datasets that meet the EU's needs for comprehensive cross-border forest monitoring. Addressing these challenges and finding a workable solution is therefore critical not only to achieving the EU's goal of a unified and actionable forest monitoring system that relies on both field and remote sensing data, but also for developing robust models capable of capturing a broad spatial range, diverse growing conditions, and varying ecological contexts across large regions. One approach for solving this issue is implementing a centralized analysis platform where all data are collected and where models are calibrated on the combined dataset of satellite and ground information with stringent rules to secure data confidentiality. For this purpose, first remote sensing data must be linked to field data on a national level without revealing exact plot coordinates to third parties. Afterwards, all extracted data is collected in one central place where all further analyses are conducted (Miettinen et al., 2025). However, this builds on the trust in the security of the central organization and hardware storing the data. In addition, this approach is exposed to the risk that plot-coordinates are unveiled from the reverse-engineering of the extracted remote sensing data.

The limitation of centralized analysis could be addressed with a federated approach, based on the concept of distributed computing, where model calibration is performed locally (in this case on country level) and only calibrated model parameters are shared. Federated Learning (FL), first introduced by McMahan et al. (2017), is a framework which enables machine learning approaches on a large, distributed dataset, without sharing the data itself (Ramírez et al., 2023). While the centralized approach is based on manual model building on the servers of analysts, FL is based on automated model building on the servers of data owners. The centralized approach centralises both data and computation but increases exposure to privacy risks and requires high trust in the central authority's security. The FL workflow, in contrast, distributes computation and minimises data sharing, offering technological safeguards for privacy and compliance while still enabling the integration of diverse datasets for improved modelling across national boundaries.

FL has previously been applied for tasks on mobile devices (Bonawitz et al., 2019), in medical research (e.g. Adnan et al., 2022; Casella et al., 2023; Gupta et al., 2023), and in environmental sciences in the context of air quality (Le et al., 2022), and weather forecasting (Chen et al., 2023). Moreno-Álvarez et al. (2024) reviewed federated learning methodologies for remote sensing data with a specific focus on crowdsourced high-resolution optical remote sensing (RS) applications. They found that FL is an effective approach for RS related tasks, which can generalize well even with heterogenous local datasets and privacy requirements. Gallios et al. (2026) utilized FL with Sentinel-2 satellite data to map soil characteristics at the farm scale throughout Brazil. Poudel et al. (2025) demonstrated the methodological potential of FL for deforestation detection, referencing the publicly available Global Forest Change dataset and employing Sentinel-2 imagery for prediction. However, their approach relied exclusively on remotely sensed inputs and did not integrate any field-based reference measurements. In contrast, our study enables collaborative cross-border modelling between four countries while retaining the confidentiality of sensitive plot-level NFI data, thereby addressing the practical challenge of integrating field and remote sensing information without compromising landowner privacy or national data security.

A wide range of machine learning approaches have been successfully applied to RS tasks, each offering specific advantages depending on the nature of the data and the classification problem. Among these, convolutional neural networks (CNNs) are widely used for image processing because they effectively learn spatial feature hierarchies (LeCun et al., 2015; Prince, 2025). CNNs work well with sequential data and can be effectively integrated into FL frameworks. Several studies confirm that these methods effectively extract temporal patterns from satellite image time series in a data-driven way (Miller et al., 2024; Pelletier et al., 2019; Perbet et al., 2024; Zhong et al., 2019).

This article presents a method using FL in combination with NFI data and time series RS data to model forest timber volume, while ensuring confidentiality of critical information. Rather than focusing on the highest-performing model, it highlights FL's advantages in generating large-scale maps by combining sensitive NFI with RS data, and compares this with centralized methods. The study maps tree volume across Norway, Sweden, Finland, and Italy, demonstrating a workflow that enables cross-border NFI-RS modelling without compromising plot-coordinate privacy – a common barrier in centralized studies. This FL approach addresses challenges of limited ground-truth data in forest resource mapping (e.g. Papucci et al., 2026). The main contributions of this work are:

- the initial use of FL for collaborative modelling in the field of forest resource assessment, specifically integrating sensitive NFI data with RS time series across multiple countries, without compromising field sample plot locations
- demonstrating a scalable framework for harmonising and integrating ground-based NFI data with satellite observations, addressing the challenges of data confidentiality and interoperability across Norway, Sweden, Finland, and Italy
- alignment with the EU's Forest Strategy and Member States requirements by offering a technological pathway for bottom-up, harmonised, high-resolution, and timely forest resource mapping for evidence-based policy and sustainable forest management.

2. Materials

In the following we describe the input data in sections 2.1 and 2.2, and data pre-processing in section 2.3. Modelling will be described in section 3. These steps are graphically depicted in Fig. 1 steps 1–4. The modelling approaches (individual, centralized, and federated) are separately explained and depicted in section 3.1 and Fig. 4.

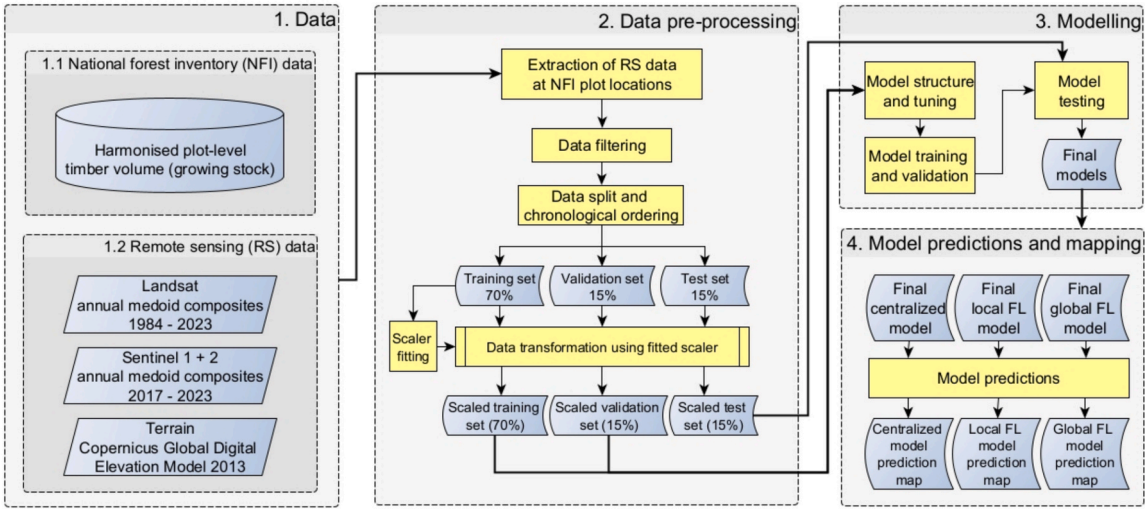


Fig. 1. Overview of processing steps. 1. Data consisting of Remote sensing (RS) based data (Landsat, Sentinel 1 and Sentinel 2, terrain data) and NFI data (timber volume). 2. Data pre-processing. 3. Model training and evaluation. 4. Model prediction and mapping. FL = Federated Learning.

2.1. National forest inventory (NFI) field data

We used harmonized data from the sample plots of the NFIs of four European countries: Norway, Sweden, Finland, and Italy (Fig. 2).

NFI field surveys were conducted according to national procedures and designs vary from country to country. These are described in detail for Norway in Breidenbach et al. (2020), for Sweden in Fridman et al. (2014), for Finland in Korhonen et al. (2024), and for Italy in Fattorini et al. (2006), Gasparini et al. (2022), and D'Amico et al. (2025).

Plot-level timber volume (growing stock) from these NFIs were harmonized following detailed definitions and instructions (nFIESTA, 2024), assuring consistent data across all countries. Timber volume was defined as the above-ground volume of living trees with a minimum dbh of 7 cm, including all parts of the stem from the stump up to a diameter of 7 cm and including all branches with a minimum diameter of 7 cm. It was not possible for Finland to include the branch volume. The omission of branch volume in Finland may lead to a slight underestimation of total timber volume relative to countries including

Fig. 2. Study area comprising Norway, Sweden, Finland, and Italy.

branches, but this effect is expected to be small given that branches typically comprise a modest proportion of stand-level volume, especially considering the prevailing tree species groups and tree dimensions in Finland. A total of 85,199 sample plots measured in the years between 2017–2023 and located inside forest defined according to the international forest definition (FAO, 2023) were used in this study (Table 1, Table 2).

2.2. Remote sensing data

Remote sensing data consisted of a time series of 40 years of Landsat images collected between the years 1984–2023, 7 years of Sentinel 1 and Sentinel 2 images collected between the years 2017–2023, and additionally terrain variables including elevation, slope and aspect. Landsat and Sentinel satellite data were processed to generate annual composites suitable for multi-temporal analysis.

For Sentinel-2 (2017–2023), annual medoid composites were created from a total number of 119,346 images acquired between June and October, applying a 70% cloud cover threshold to limit location errors caused by excessive cloud contamination (Francini et al., 2023; Kennedy et al., 2018). Each composite was created at 10 m resolution including surface reflectance from the red, green, blue, near-infrared (NIR), and short-wave infrared band 2 (SWIR2), and derived spectral indices such as the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Normalized Burn Ratio (NBR), Moisture Stress Index (MSI), Normalized Difference Moisture Index (NDMI), and Tasseled Cap Brightness (TCB), Greenness (TCG), and Wetness (TCW) (Parisi et al., 2023).

Sentinel-1 data (2017–2024) were processed separately for ascending and descending orbits. Annual median composites were computed from a total number of 76,132 ascending and 75,534 descending images (Mullissa et al., 2021), both for the VV and VH polarization bands at 10-meter resolution.

For Landsat (1984–2023), annual medoid composites were generated from a total number of 77,061 scenes acquired between May and September, using the same cloud threshold as for Sentinel-2. Preprocessing included de-spiking and temporal gap-filling via linear interpolation. From these composites, the bands blue, green, red, NIR, SWIR1, and SWIR2 were used, along with a suite of indices: NDVI, EVI, NBR, Normalized Difference Water Index (NDWI), Normalized Difference Built-up Index (NDBI), Enhanced Mangrove Vegetation Index (EMVI), and Tasseled Cap components TCB, TCG, TCW, and the Tasseled Cap Angle (TCA) (Huete, 2012; Parisi et al., 2023).

For the terrain variables, the Copernicus Global Digital Elevation Model was used to derive elevation above sea level, aspect, and slope (Copernicus, 2015). Additionally, the Inspire grid cell coordinate (lower left corner) of the 1 km by 1 km Inspire grid cell in which an NFI plot was located was added to the feature set of each plot.

All remote sensing data were available wall-to-wall for the four countries (Fig. 2) and were used for the features that served as predictors in the models with F = {F_L, F_S2, F_S1S2} features for Landsat and terrain, Sentinel 2 and terrain, and Sentinel 1 and Sentinel 2 and terrain, respectively, and where the number of features were F_L = 21, F_S2 = 18, and F_S1S2 = 22.

Table 1
Details of NFI data for each of the four countries: number of primary sampling units (n PSUs), number of sample plots (n plots), minimum, mean, and maximum timber volume, and years of data collection.

	n PSUs	n plots	Timber volume ($m^{3}\ ha^{-1}$)		Years of data collection
			mean	max	
Finland	7,295	40,828	129	1,121	2019 – 2023
Italy	6,671	6,671	173	1,836	2017 – 2020
Norway	10,271	10,271	118	1,158	2019 – 2023
Sweden	5,663	27,429	148	2,543	2017 – 2021
Total	29,900	85,199	137	2,543	2017 – 2023

2.3. Data pre-processing

During data preprocessing (as described in Step 2 of Fig. 1), all satellite time series pixels and terrain data intersecting a 100 m^2 buffer (radius of 5.64 m) around the center of each NFI field plot were extracted. This was done by each country using exact plot coordinates. For each plot and year, weighted means were computed. The pixel proportions intersecting the circle were used as weights. We chose this approach because the NFI center may be located close to a pixel border (Fig. 3 A). Using weighted means of all intersecting pixels best represents the field plot location. All observations with at least one pixel value per plot and a minimum of 50% plot coverage were included in the analysis.

The data were further pre-processed to ensure that the input was formatted and scaled appropriately for optimal model performance with a CNN designed to handle time series data. We split the data into three subsets: training and validation data sets for model training, and a test data set, which remained unseen by the models and served as an independent benchmark for final model evaluation (Prince, 2025). We performed plot-grouped, random splitting ensuring that all observations from a plot occurred in a single split and that input features for a measurement in year t only use satellite acquisitions at or before t. The identical training, validation, and test splits were used for all models. The training and validation datasets consisted of 70% and 15% of the total data, respectively, while the independent test set accounted for the remaining 15%. All data were organized chronologically according to the year of remote sensing acquisition to preserve temporal consistency. The remote sensing time series data for each plot was concluded in the same year as the field data collection, ensuring that both datasets were temporally aligned. The target variable and features representing only one point in time (terrain features and Inspire grid cell coordinates) were added to the data. Single missing values within the time series were interpolated. To complete 40 years of Landsat data for field plots collected before 2023, missing years at the start (before 1984 for Landsat, before 2017 for Sentinel) were back-filled with the first valid observation to standardize input shape. Data dimensions were (*1, n features, t years*), where "1″ represents a single pixel for each NFI sample plot (one weighted mean value if multiple pixels intersect); "n features" includes all variable types such as spectral, terrain, and spatial data; and "t years" indicates the number of temporal observations recorded for each feature (Fig. 3 B).

Furthermore, the Landsat and Sentinel data sets were cut into different time series lengths T = {T_L, T_S} by cutting years off from the start of the time series to test the effect of time series length on timber volume modelling. For Landsat data, the 40-year time series was cut into shorter lengths such that T_L = {40, 30, 20, 7, 4}, where T_L indicates the time series lengths in years for the Landsat data set. The 7-year Sentinel time series was cut into one shorter length such that T_S = {7, 4}. The 4 years correspond to the maximum common time series length using the common newest field data collected in 2020. This was done to avoid back-filled data at the beginning of the already relatively short time series for countries with older NFI data.

Features and target variables were scaled using

$$x' = \frac{x - \mu}{\sigma} \quad (1)$$

where x' is a normalized feature or target variable, x is its unscaled version, μ is the mean and σ the standard deviation of the feature variable in question in the training data. After fitting the scaler on the training data, the scaler was then used to transform all datasets.

3. Methods

In this section we will first describe the modelling approaches

Table 2
Number of NFI sample plots per country and year.

	2017	2018	2019	2020	2021	2022	2023
Finland	–	–	8,100	8,066	8,184	8,159	8,319
Italy	56	3,953	2,643	19	–	–	–
Norway	–	–	2,053	2,016	2,043	2,062	2,097
Sweden	5,545	5,464	5,540	5,571	5,309	–	–
Total	5,601	9,417	18,336	15,672	15,536	10,221	10,416

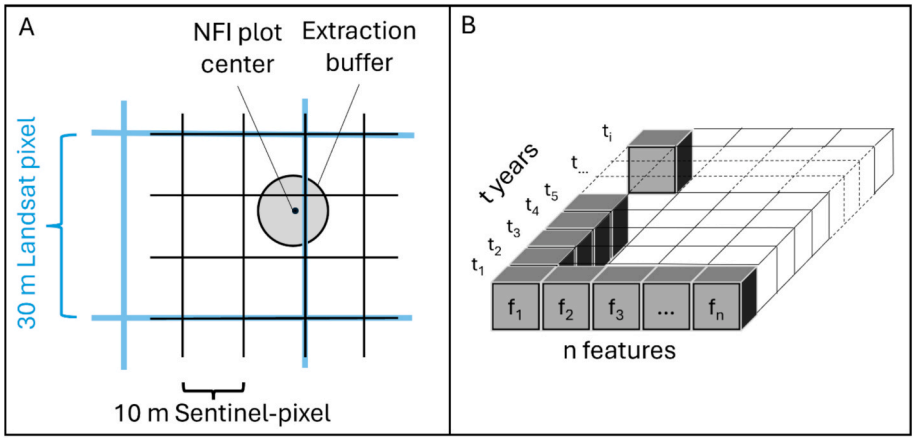


Fig. 3. A: Overlay of the extraction buffer around an NFI field plot with the Landsat (30 m, in blue) and the Sentinel (10 m, in black) grid. B: Data matrix structure for one observation.

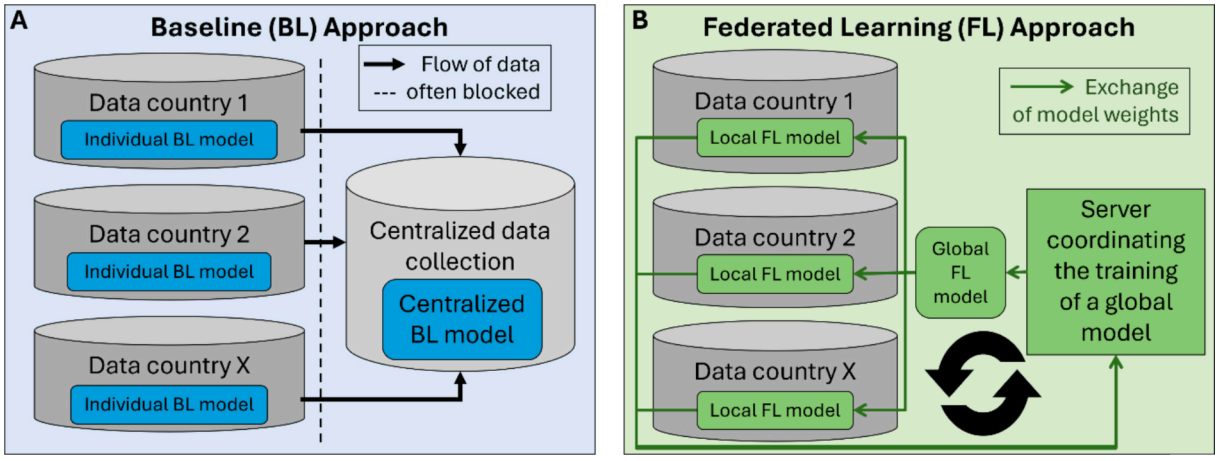


Fig. 4. Overview of the two modelling approaches and fitted models. A) Flow of data in the baseline, non-federated approach. Here, country specific models are fit individually, and data are collected centrally and a centralized model using a combined data set is fit. This flow of data is often blocked due to data confidentiality indicated by a stipulated vertical line. B) Federated learning approach, where the data remain with its owner and only model weights are iteratively exchanged (symbolized by the black circular arrows) with a centralized server training a global model. Local models benefit through the weights of the global model indirectly from the data of the other countries.

baseline (BL) versus Federated Learning (FL), since this is the main focus in this article. Afterwards we will describe the specific model architecture, which we used for all models.

3.1. Modelling approaches

We trained non-federated BL models for each individual country using each country's individual data set separately, and a large, centralized model using the combined data of all countries. These two model types we call BL models. Subsequently, we fitted federated models using the FL framework. All models were fitted with the same data splits, and the same random seeds were set before model training.

3.1.1. Individual and centralized baseline (BL) models

As a baseline and for comparison with the federated approach we fitted individual country specific models and, in addition, one centralized model using a combined data set of all country data sets (Fig. 4 A). The models were trained using the training and validation data sets, and the final models were evaluated using the independent test data sets that were unseen by the model during training. The centralized model was additionally applied to each country's independent test dataset individually. These models were used as baseline, to which we compared the results of the federated models.

3.1.2. Federated learning (FL) models

We employed FL using Flower (Beutel et al., 2020). To implement the

FL models, we employed the four participating countries, each acting as a client in a cross-silo manner with horizontally partitioned data (Moreno-Álvarez et al., 2024). A central server orchestrated the training process by initializing a global model with randomly assigned weights. This initial model was distributed to all clients, who then performed local training using their respective datasets. Upon completion of local training, each client transmitted its updated model weights back to the central server. The server aggregated these weights to update the global model, which was subsequently redistributed to all clients. This iterative process enabled each local model to benefit from the collective knowledge of all clients, without direct data sharing, thereby preserving confidential field plot coordinates (Fig. 4 B). The training proceeded through a pre-defined number of rounds (60 FL rounds for the Landsat and Terrain data set) of communication between the clients and the server, with the server continuously refining the global model through successive aggregations. Each client saved the locally fitted models resulting in customized fine-tuned local models (Fig. 4 B), and the server saved the global aggregated models. Global validation metrics were calculated by aggregating the sums of squares of each client.

While FL is normally characterized by decentralized data storage and communication across multiple remote client devices or servers, we chose to simulate this setting by hosting all client datasets on a single server while still treating each client separately. To evaluate the practical viability of the FL approach, we conducted a real-world test using two clients, representing Norway and Sweden.

We tested the model aggregation algorithms FedAvg (McMahan et al., 2017) and FedProx (Li et al., 2020) to address data heterogeneity across clients. FedProx introduces a proximal term μ to the loss function, which stabilizes training by mitigating the effects of non-IID (non-independent and identically distributed) data across clients. We report results for FedProx with $\mu = 0.05$.

3.1.3. Technical implementation of the FL communication

To implement FL in Flower, the central server needs to be reachable from the internet and listen on two ports (defaulting to 9092 and 9093). Clients contact the server on these two ports but are never contacted by the server. Authentication is ensured using certificates that are exchanged beforehand. More details can be found in the Flower Framework Documentation (Beutel et al., 2020).

3.2. Model building

Modelling (step 3 in Fig. 1) consisted of building a model architecture and configuring the model by tuning the model's hyperparameters. All models were trained including terrain variables, and with various satellite time series lengths of T years. The same hyperparameters across all models using the same features and time series lengths were used.

3.2.1. Time series CNN architecture

CNNs have become a widely used approach for processing and analysing two-dimensional imagery (LeCun et al., 2015; Prince, 2025). However, they are also well-suited for time series data, particularly when adapted to capture temporal dependencies (Ismail Fawaz et al., 2019). CNNs' ability to extract local patterns through convolutional operations makes them effective for time series analysis by applying convolutions along the temporal axis. Several studies have demonstrated the utility of temporal CNNs for satellite image time series. For instance, (Pelletier et al., 2019; Perbet et al., 2024) applied temporal CNNs for classification tasks, while (Perbet et al., 2024) used them for forest disturbance detection. In this study, we applied a Time Series CNN (tCNN) architecture implemented in PyTorch (Paszke et al., 2019) for timber volume prediction, customized for the given time series dataset. The architecture treats each plot as a single-pixel multivariate "movie" through time and learns filters that slide along the temporal axis while looking across spectral features at once. Early convolutional filters act like short-window detectors that respond to local patterns, e.g., sudden drops from disturbance, or short-term recovery, whereas deeper layers (together with pooling) integrate these cues over longer spans to capture trends and legacy effects. The model takes as input a single-channel tensor with 21 features and 40 time points per sample plot. The first convolution spans all features and a temporal window of width 5, producing 8 feature maps while preserving the temporal dimension. Max-pooling reduces the temporal length from 40 to 20. A second convolution with kernel size (1, 5) increases the number of channels to 16 while maintaining the temporal dimension. For the maximum time-series length, adaptive average pooling reduces the representation to (1, 2), followed by dropout regularization and a fully connected linear output layer producing one prediction value per sample. This model architecture is detailed in Table 3. This description assumes the case where F = F_L and T = T_L = 40, but the structure generalises to all datasets and varying time series lengths.

Batch normalizations after the first and second convolutional layers for normalizing the activations from the convolutional layers and enhancing training dynamics by stabilizing the mean and variance of layer inputs were tested but did not improve the training.

3.2.2. Time series CNN configuration

The model based on Landsat data had F_L = 21 features, the model based on Sentinel 2 data had F_S2 = 18 features, and the model based on Sentinel 1 and Sentinel 2 data had F_S1S2 = 22 features. We trained the tCNN model and optimized its hyper parameters kernel time, pool time, dropout layer, and weight decay by minimizing pixel-level root mean squared error (RMSE) of the validation data of a combined data set of all countries. We performed the hyperparameter tuning for different models with input data of time series lengths T_L and T_S years for Landsat and for Sentinel data, respectively. The final models use the mean squared error (MSE) as the loss function, which is suitable for regression tasks such as predicting continuous variables from time-series data. Optimization was performed using the Adam optimizer with a learning rate of 0.001. Overfitting was controlled by incorporating a dropout layer in the model architecture and by setting a low maximum epoch limit of 15. Additionally, we monitored validation loss at the end of each epoch. Training was automatically terminated via an early stopping mechanism if the validation loss failed to decrease for three consecutive epochs, at which point the optimal weights were restored.

3.3. Model evaluation

We selected the overall best performing global FL model, and for each client the best performing local FL model based on validation loss. Subsequently, we produced model predictions for the independent test datasets using the following models: i) individual, country specific BL models applied to the corresponding country-specific test data sets (individual BL models in Fig. 4 A), ii) the centralized BL model applied to the combined country test data set (Centralized BL model in Fig. 4 A), iii) the centralized BL model applied to each country-specific test data set, iv) the local FL models, and v) the global FL model applied to the corresponding country-specific test data sets (Fig. 4 B). The coefficient of determination (R^2) and RMSE were calculated, and the relative RMSE was obtained by dividing the RMSE by the mean of the actual, observed

Table 3
Time series CNN model architecture, shown for F = F_L and T = T_L = 40. This structure applies to all datasets and time series lengths.

	Layer	Kernel/Unit	Output shape
1	Input	–	(1, 21, 40)
2	Convolution 1	8 filters × kernel size (21, 5)	(8, 1, 40)
3	Max-Pool	kernel size (1, 2), stride 2	(8, 1, 20)
4	Convolution 2	16 filters × kernel size (1, 5)	(16, 1, 20)
5	Adaptive Avg Pool	output size (1, 2)	(16, 1, 2)
6	Dropout	–	(16, 1, 2)
7	Dense (linear output)	32 → 1	(1)

values and multiplying by 100. All metrics were calculated at plot level. These results were compared across models to assess their performances.

We performed the same steps and analyses for the datasets composed of Landsat data, Sentinel 2, and Sentinel 1 and 2 data.

3.4. Model predictions and mapping

For applying the models and creating prediction maps (step 4 in Fig. 1), we used the final local FL models to create wall-to-wall prediction maps of timber volume for the countries. We generated predictions for the Landsat and terrain-based models, and for the Sentinel and terrain-based models.

4. Results

4.1. Full time series

The results of the model predictions using test data are illustrated in the plots of observed versus predicted timber volume using the Landsat data for the 40-year time series (Fig. 5). For the individual BL models using the full 40-year time series of satellite data, RMSE values ranged from 72–109 m^3/ha (62%−75%, R^2: 44–55). The centralized BL model applied to the country-specific data sets performed slightly worse than the individual BL models trained for each country, with RMSE values ranging from 1%−13% higher and lower R2 values. The local FL models performed for each country comparable to the individual BL models with RMSE values ranging from 72−110 m^3/ha (56%−74%). The global FL model performed comparably to the individual and centralized BL and local FL models in Sweden and Finland but showed reduced accuracy in Norway and Italy. Overall, the results showed that the local FL models achieve similar performance as the BL models across all countries. In Norway, the global FL model performed slightly worse, and in Italy, both the centralized BL and the global FL models performed slightly worse than the other models. The accuracy metrics of all models are presented in Table 4.

We compared global validation loss for each client using FedAvg and FedProx (Fig. 6). In Scandinavia (Norway, Sweden, Finland), both algorithms showed similar results. For Italy in southern Europe with different growing conditions varying from Mediterranean to Alpine climates and diverse tree species, FedProx was stabilizing after round 40 and outperformed FedAvg.

The average round processing time for FedProx was with 47.2 s 5.2 s longer than for FedAvg with 42.0 s average round processing time. Local average training time was 7.9 s, 14.2 s, 15.8 s, and 6.1 s for Norway, Sweden, Finland, and Italy, respectively.

Fig. 7 shows how loss changes over 60 FL rounds in the four countries using FedProx. The training loss, validation losses (last and best epochs), and global model loss before local training are plotted. Across all countries, global model validation loss drops sharply at first, then levels off as models converge. Norway and Sweden show steady training loss reductions, with validation losses just above training loss, while the global model always performs worse than locally adapted models. In Finland, all losses quickly converge to low values, indicating stable learning. Italy displays a strong need for local training, as the global model consistently performs worse, though improves slightly after round 40. In Finland and Italy, the training loss remains higher than the validation loss, which may be explained by the FedProx proximal penalty being included in the training objective or by the fact that the reported training loss is averaged over all local batches and epochs within a round, thus also reflecting earlier and less well-adapted model states before local improvement from the incoming global model. Overall, most improvement occurs early, with country-specific benefits from local adaptation, especially for Italy.

The communication overhead in federated learning was calculated based on the 1537 trainable parameters of the time series CNN used for Landsat and terrain data (F_L = 21 input features, T_L = 40 time steps). With 32-bit floats, a full model transmission is 6148 bytes. Each client sends and receives this amount per round, totalling 12,296 bytes. With four clients, each round incurs 49,184 bytes; over 60 rounds, cumulative exchange reaches about 2.95 MB (Table 5). This estimate excludes minor additional overheads from protocol metadata and evaluation metrics. Overall, the approach's communication demand (2.95 MB) is minimal and scalable to larger setups. A real-world trial involving Norwegian and Swedish clients was completed successfully.

4.1.1. Effect of time series length

RMSE, RMSE%, and R2 metrics were used to evaluate prediction accuracy across different time series lengths. For Sweden, Finland, and Italy, model performance improved with increasing time series length until 20 years. Increasing the time series length to 30 or 40 years resulted in little improvement or even a slight decrease in accuracy (Fig. 8). Both the individual BL and local FL models exhibited gains in accuracy with longer time series. However, this effect could not be observed in Norway. The local FL models benefitted from longer time series, with RMSE gains between the 40-year and the 4-year models of 1%, 7%, 5%, 7% for Norway, Sweden, Finland, and Italy, respectively. The gains of the 20-year models compared to the 4-year models were 2%, 5%, 5%, 8% for Norway, Sweden, Finland, and Italy, respectively. The global FL model with the same time series length showed similar performance in Sweden and Finland. Accuracy gains in the global FL model with longer time series in Italy was small. For the models based on Sentinel data, these trends were similar but not as pronounced due to the shorter time series (results not shown).

4.1.2. Bias

Bias was calculated as predicted minus observed values of the test data sets. The overall bias of the local FL models (40 years Landsat) was 8.26, 0.91, −1.38, and −1.82 for the four counties Norway, Sweden, Finland, and Italy, respectively. Similar values were obtained for the individual BL models. Overall bias of the global FL models was generally larger. Similar patterns were observed for the models based on the other data sets (Table 6).

4.1.3. Timber volume maps

In Fig. 9 we present an overview with Landsat based predictions, and two close-ups of our prediction maps together with corresponding satellite images. The first close-up shows Landsat based predictions and RGB in southern Norway around the Norwegian and Swedish border. The second close-up shows Sentinel based predictions and RGB in north-central Italy.

5. Discussion

This study demonstrates the potential of FL as a viable and privacy-preserving approach for forest resource modelling and monitoring with the combination of field plot and remotely sensed data, which has not been explored in previous studies in this manner. The main advancement here is not a new deep learning module, but rather a modelling workflow that allows cross-border NFI-RS learning without revealing plot coordinates, which is a common restriction that often hinders centralized international research (Papucci et al., 2026). Our findings show that local FL models can achieve predictive performance comparable to centralized models across the participating countries. This is an important result, as it highlights FL's capacity to address data confidentiality concerns related to field plot coordinates, while maintaining model accuracy across countries.

5.1. Comparison with previous work

Our results are consistent with those of Miettinen et al. (2025), who produced Pan-European forest maps using a combination of satellite-based earth observation data and NFI plots using the centralized

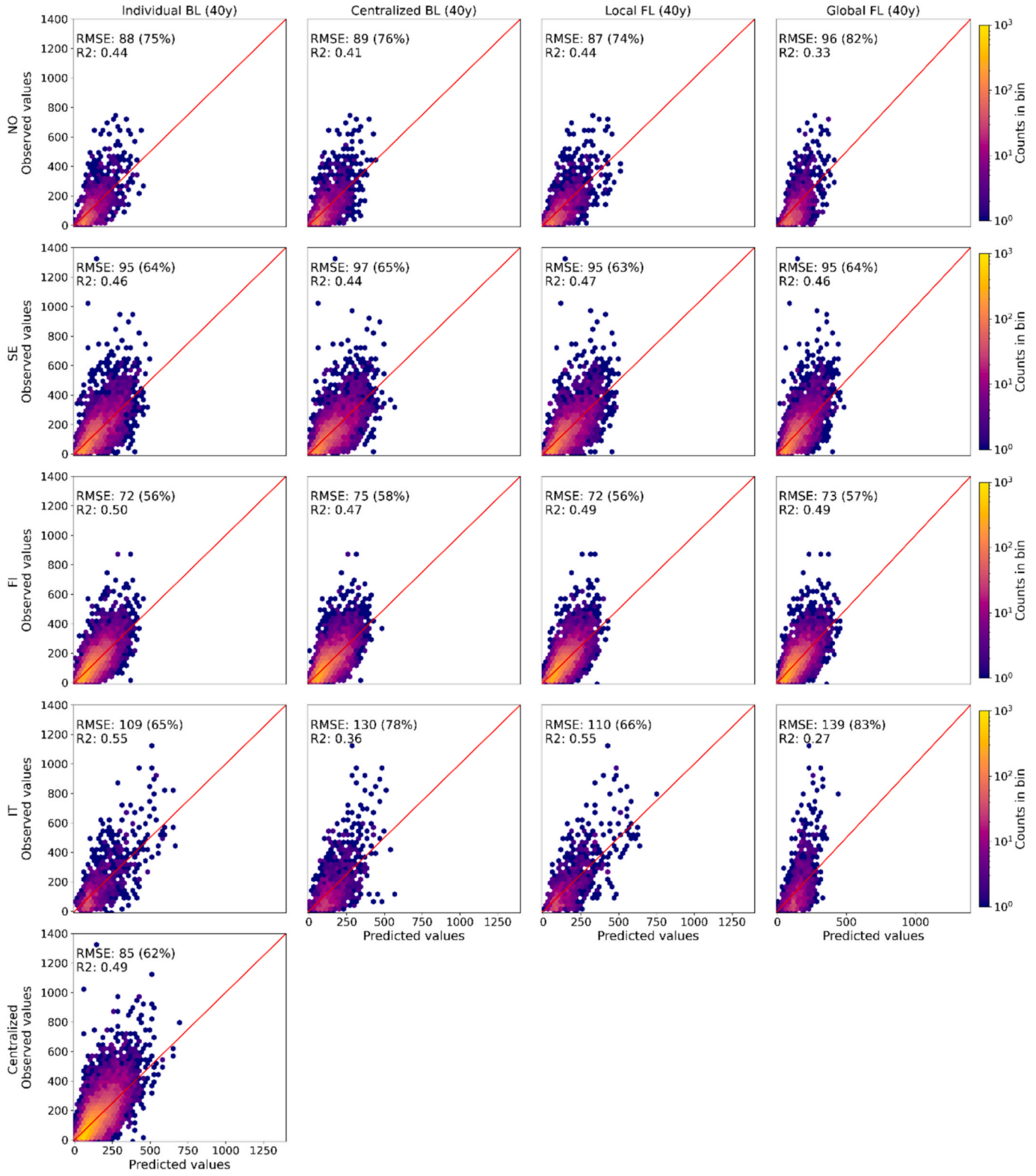


Fig. 5. Observed (y-axis) versus predicted (x-axis) timber volume [m^3/ha] of the Time Series CNN models using the Landsat and terrain data, including RMSE (RMSE %) and R^2 values. Rows represent the results for each individual country and for all country data combined (centralized) in the bottom row. Columns: Individual baseline (BL) (40y): models fitted using 40 years of country specific remote sensing data and centralized model fitted with combined data; Centralized BL (40y): the centralized model fitted with the combined country data (from column 1 bottom row) and applied to 40 years of country specific remote sensing data; last two columns: Results of the local and global Federated Learning (FL) models using 40 years of remote sensing data.

Table 4
Accuracy metrics for time series models based on 40 years of Landsat, 7 years of Sentinel 2, and 7 years of Sentinel 1 and 2 data. Results are presented for the four countries Norway (NO), Sweden (SE), Finland (FI), and Italy (IT), and additionally for the combined data set labelled 'Centralized'. RMSE values are presented in m^3/ha (relative RMSE in %).

		Baseline		Federated Learning	
Country		Individual	Centralized	Local	Global
40 years Landsat and terrain data					
NO	RMSE (%)	88 (75%)	89 (76%)	87 (74%)	96 (82%)
	R2	0.44	0.41	0.44	0.33
SE	RMSE (%)	95 (64%)	97 (65%)	95 (63%)	95 (64%)
	R2	0.46	0.44	0.47	0.46
FI	RMSE (%)	72 (56%)	75 (58%)	72 (56%)	73 (57%)
	R2	0.50	0.47	0.49	0.49
IT	RMSE (%)	109 (65%)	130 (78%)	110 (66%)	139 (83%)
	R2	0.55	0.36	0.55	0.27
Centralized	RMSE (%)	85 (62%)	–	–	–
	R2	0.49	–	–	–
7 years Sentinel 2 + terrain					
NO	RMSE (%)	90 (75%)	95 (78%)	91 (75%)	95 (79%)
	R2	0.48	0.43	0.48	0.42
SE	RMSE (%)	97 (66%)	99 (68%)	96 (66%)	97 (66%)
	R2	0.43	0.40	0.44	0.44
FI	RMSE (%)	71 (55%)	78 (61%)	72 (56%)	72 (56%)
	R2	0.51	0.41	0.50	0.49
IT	RMSE (%)	122 (70%)	161 (93%)	125 (72%)	156 (90%)
	R2	0.49	0.11	0.47	0.17
Centralized	RMSE (%)	88 (64%)	–	–	–
	R2	0.47	–	–	–
7 years Sentinel 1 and 2, and terrain data					
NO	RMSE (%)	92 (76%)	94 (78%)	91 (75%)	96 (80%)
	R2	0.47	0.44	0.48	0.41
SE	RMSE (%)	96 (66%)	99 (68%)	96 (66%)	97 (66%)
	R2	0.45	0.41	0.44	0.43
FI	RMSE (%)	71 (55%)	76 (59%)	72 (55%)	72 (56%)
	R2	0.51	0.44	0.50	0.49
IT	RMSE (%)	123 (70%)	156 (90%)	123 (71%)	153 (88%)
	R2	0.49	0.17	0.48	0.20
Centralized	RMSE (%)	88 (64%)	–	–	–
	R2	0.47	–	–	–

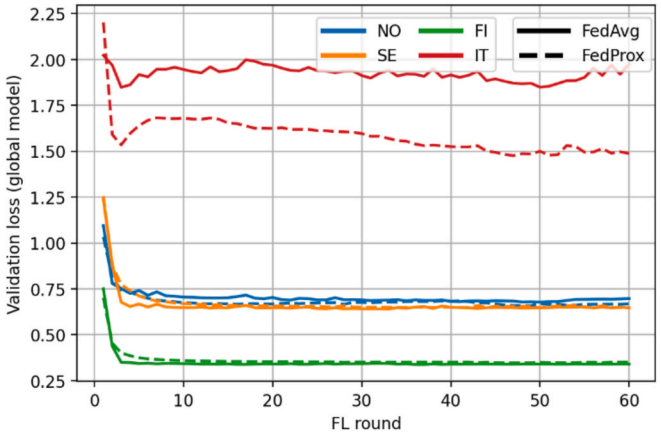


Fig. 6. Global validation loss for each client using the FedAvg (solid line) and the FedProx (dashed line, with μ=0.05) algorithms. In Italy (IT, in red), FedProx outperforms FedAvg, while no differences between the two algorithms can be observed in the Scandinavian countries Norway (NO, in blue), Sweden (SE, in orange), and Finland (FI, in green).

approach, where all data were collected and stored in one data base. Using the local FL model, we achieved RMSEs for Norway, Sweden, and Finland of 72 – 95 m^3/ha (56% – 74%), with a combined value of 83 m^3/ha (62%) for these three countries, while Miettinen et al. (2025) reported RMSEs for timber volume of 76 m^3/ha (53%) in the boreal region (Norway, Sweden, and Finland combined. Using the local FL model in Italy, we achieved an RMSE of 110 m^3/ha (66%), while Miettinen et al. (2025) in a southern region (southern Germany, Switzerland, Austria, south-east France, and entire Italy, Area 11 in that study) reported an RMSE of 178 m^3/ha (62%).

Koma et al. (under review) found comparable results for modelling timber volume in Norway and Finland using Sentinel 1, Sentinel 2, and PALSAR-2 data and a UNet-based deep learning model. Using only Sentinel 2 data, they found RMSE ranging from 61% to 73%, bias ranging from −7.9% to 2.1%, and R^2 ranging from 0.41 to 0.58. Combining Sentinel 1, Sentinel 2 and PALSAR data, they reported RMSE ranging from 57% to 72%, bias ranging from −2.5% to 2.7%, and R^2 ranging from 0.42 to 0.63. These results are in line with our findings.

Ceriani et al. (2025) modelled timber volume in mountain forests in northern Italy using spaceborne sensors and several machine learning approaches. With Sentinel-2 data alone, they reported RMSE values of 98–139 m^3/ha across the tested models. Our results for the whole of Italy falls within this range.

5.2. Client Contributions and Model Dynamics in Federated Learning

The influence of individual clients on the global FL model is a critical aspect of system behaviour. Our study highlights several mechanisms that govern this dynamic: i) choice of aggregation algorithm: using the FedAvg algorithm for aggregating model weights, client updates are weighted by the number of local training samples (McMahan et al., 2017), which naturally reduces the influence of clients with limited data; FedProx introduces a proximal term to the loss function, which stabilizes training by mitigating the effects of non-IID data across clients (Li et al., 2020); ii) local fine-tuning: the ability of clients to fine-tune the global model locally is important. It allows each client to adapt the shared model to its specific data distribution, ensuring that useful local patterns are still captured even if the global model is slightly biased; iii) fairness and collaborative gains: while FL may not yield uniform performance improvements for all clients, it fosters fairness by enabling clients with small or noisy datasets to benefit from the collective knowledge of the group.

In the present study we observed lower predictive performance of the global FL model in Norway and Italy. As discussed above, this corresponds to smaller sample sizes in Norway and Italy (6671and 10,271, respectively) compared to Sweden and Finland (40,828 and 27,429, respectively). However, additional factors may also have contributed to these differences. In federated learning, heterogeneity between clients is known to affect performance of the aggregated global model. Differences in forest structure, dominant tree species, climatic conditions, and forest management regimes may have resulted in country-specific feature-target relationships that deviated from the global average learned during federated optimization. Differences in spatial and temporal alignment between predictor and target data may also have contributed to reduced predictive accuracy in some countries. Due to the weighted aggregation function in FL, the larger clients dominate over the smaller clients. However, this effect does not hold for the local FL models, where accuracies comparable to the individual baseline models were consistently observed.

The aggregation method FedProx did not outperform FedAvg in the Scandinavian countries of Norway, Sweden, and Finland; however, it produced better outcomes than FedAvg in Italy. A μ value of 0.05 was applied throughout the experiments. The observed differences between countries may be related to varying degrees of heterogeneity among local datasets. In federated learning, FedProx is specifically designed to mitigate the effects of non-identically distributed (non-IID) data by constraining local model updates from diverging excessively from the global model during training. The Italian dataset likely differed more strongly from the aggregated global data distribution than the

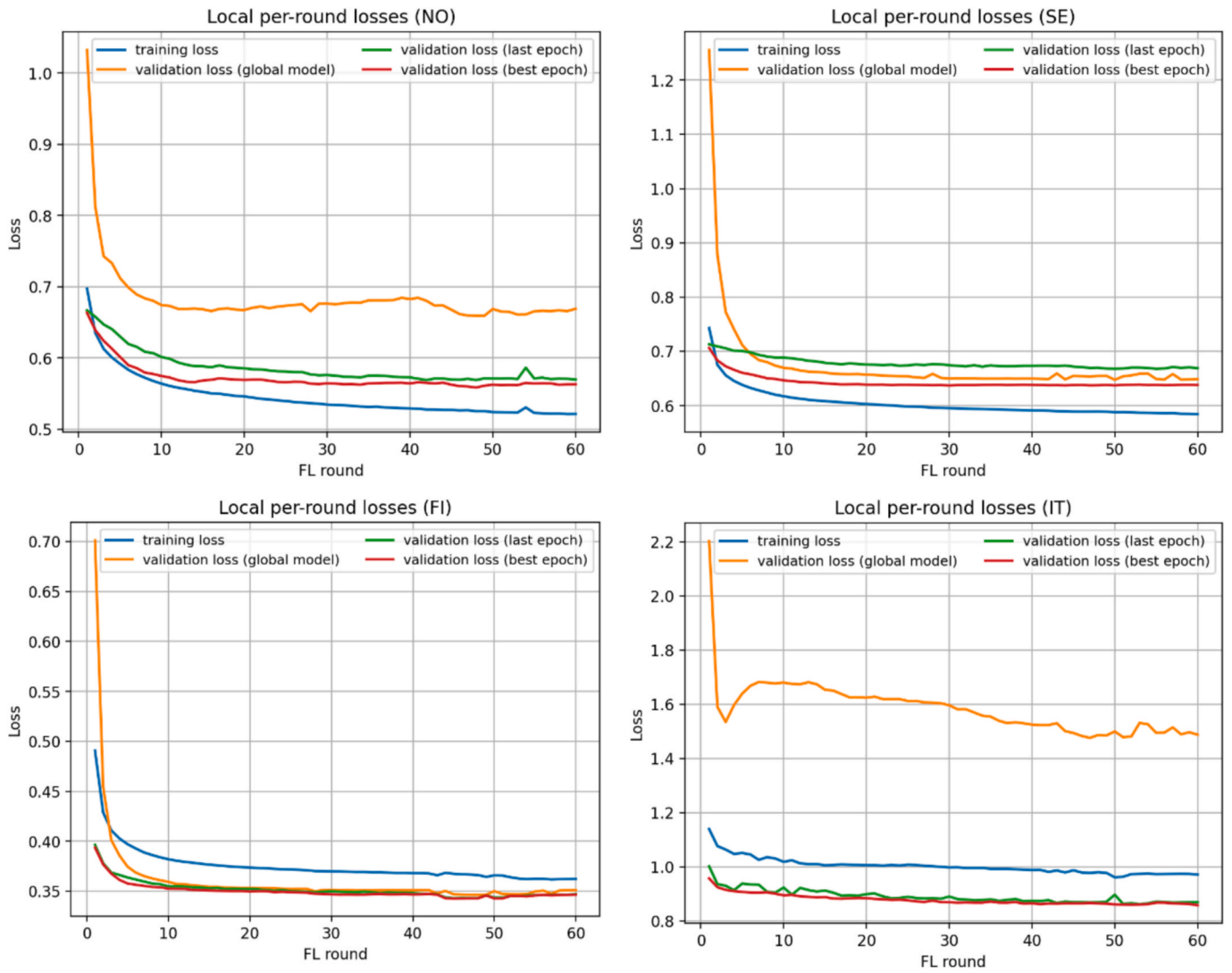


Fig. 7. Loss over federated learning rounds for Norway (top left), Sweden (top right), Finland (bottom left), and Italy (bottom right). Note different y-axis scales. The blue curve shows training loss, the green curve shows loss of the last epoch, the red curve shows loss of the best epoch, and the orange curve shows the loss of the global model before local training.

Table 5
Communication overhead of the 40-years Landsat and Terrain tCNN model with 4 clients and 60 FL rounds.

Communication quantity	Calculation	Size
Model size per transmission	(1537 x 4) bytes (float32)	6148 bytes
Per client per round (download + upload)	(2 x 6148) bytes	12,296 bytes
Total per round for 4 clients	(4 x 12,296) bytes	49,184 bytes
Total for 60 rounds and 4 clients	(60 x 49,184) bytes	~2,95 MB

Scandinavian datasets due to differences in forest structure, dominant tree species, climatic conditions, productivity, and forest management regimes. Furthermore, the smaller sample size may have contributed to less stable local optimization in Italy. Under such conditions, the proximal regularization term in FedProx may help stabilize training and reduce excessive local model drift, thereby improving predictive performance. In contrast, the Scandinavian countries likely exhibited more similar data distributions and ecological conditions, allowing the standard FedAvg aggregation approach to converge effectively without additional regularization. In these cases, the FedProx constraint may have slightly limited local adaptation to country-specific patterns, reducing its potential benefit. More generally, reducing the proximal regularization parameter μ tended to benefit larger clients that predominantly influenced the global optimization process, whereas increasing μ provided advantages to smaller or more heterogeneous clients by constraining local updates more strongly.

5.3. *Implications for forest monitoring*

A critical aspect in forest monitoring is the protection of field plot locations, which is a barrier to fully integrate remote sensing and ground data (Papucci et al., 2026). These plots are often situated on private land and are revisited over long time periods. While demands for freely available NFI data are made (Nabuurs et al., 2022), public disclosure of their locations could lead to vandalism, unintentional disturbance, bias in harvesting regimes, or (under)exploitation of sensitive ecological or economic resources, which could violate landowner privacy and lead to undermining trust in national inventory programs (US Forest Service, 2023; Schadauer et al., 2024). Maintaining the confidentiality of plot locations is therefore essential for ensuring the scientific integrity, security, and continuity of long-term forest monitoring. Our findings suggest that FL can play a pivotal role in overcoming these criticalities

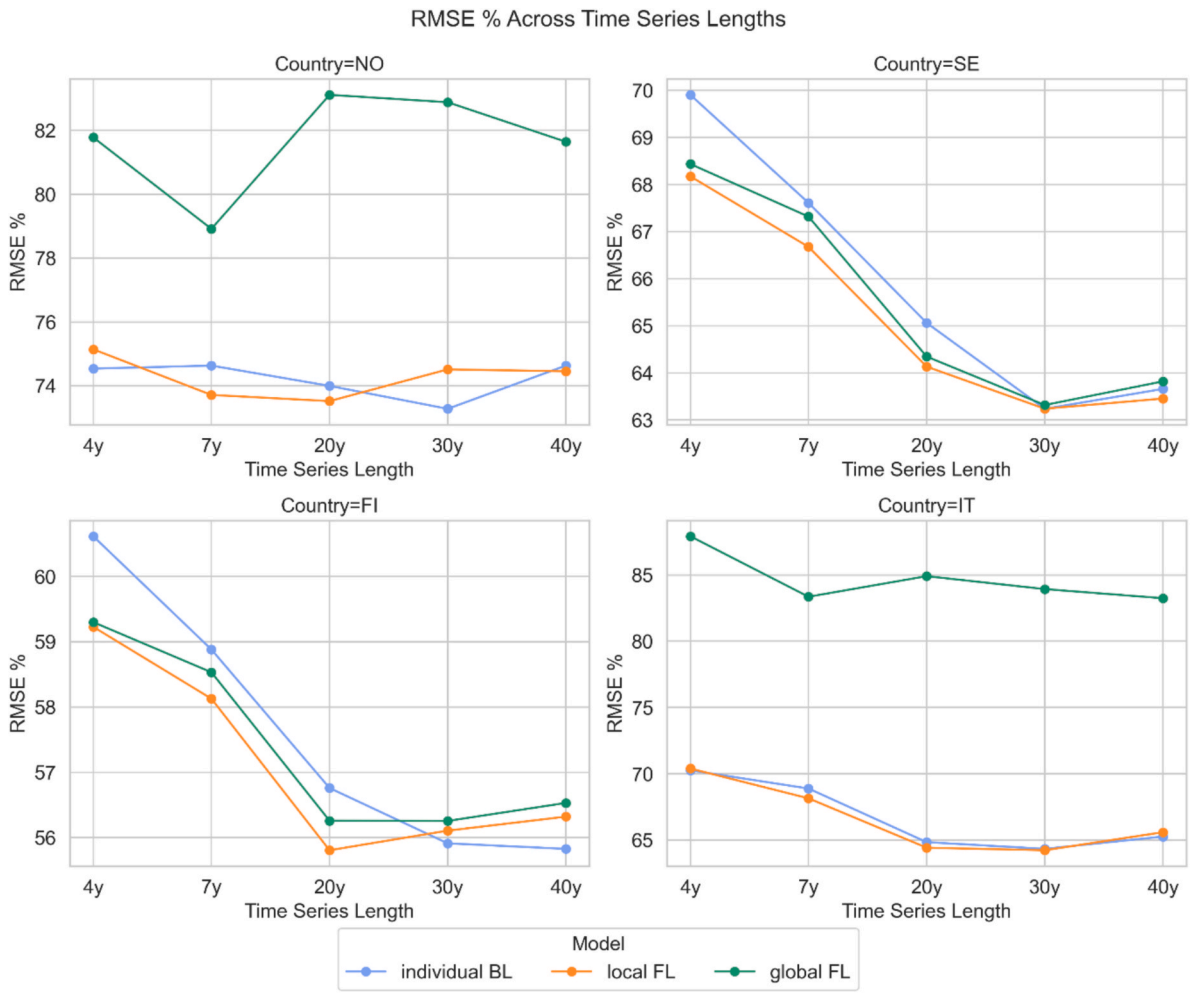


Fig. 8. Line plots of RMSE% values of independent test data for model results (individual BL, local FL, global FL) for data with time series lengths of 4, 7, 20, 30 and 40 years. Note individual y-axes scales for each country.

Table 6
Overall bias, calculated as predicted minus observed values, for the different models and corresponding test data sets. Results are presented for the four countries Norway (NO), Sweden (SE), Finland (FI), and Italy (IT), and additionally for the combined data set labelled 'Centralized'.

Overall bias	Baseline		Federated Learning	
	Individual	Centralized	local	global
40 years of Landsat and terrain				
NO	3.36	−4.50	8.26	19.43
SE	−5.38	0.13	0.91	−4.59
FI	−2.32	−2.57	−1.38	−1.42
IT	−2.40	−1.16	−1.82	−11.51
Centralized	−2.63	–	–	–
7 years of Sentinel 2 and terrain				
NO	−4.5	−8.71	−1.31	11.75
SE	0.35	1.57	1.36	−0.82
FI	1.82	−8.29	1.05	5.85
IT	−4.51	−5.01	10.02	−11.91
Centralized	−1.04	–	–	–
7 years of Sentinel 1 and 2 and terrain				
NO	3.79	−4.10	−5.75	14.82
SE	−1.28	3.02	0.02	−2.33
FI	2.07	−1.29	3.90	8.02
IT	−0.32	0.11	−0.89	−28.68
Centralized	−1.90	–	–	–

and achieving the objectives of a European Forest Monitoring system. Furthermore, by enabling secure and collaborative modelling across national datasets, FL supports the integration of remote sensing and field data for generating harmonized forest resource maps across large spatial domains. These outputs are essential for supporting the formulation and implementation of evidence-based policy related to the forest sector, the development of forest adaptation plans, the sustainable management, and international reporting related to climate mitigation and biodiversity conservation.

5.4. *Time series lengths and model architecture*

Both the individual and local FL models benefitted from longer time series. This was most obvious for the 40 years Landsat data in Sweden, Finland, and Italy, where most improvements were observed until a time series length of 20–30 years. In Norway, there was only little improvement until a time series length of 20 and 30 years for the BL and local FL models, respectively. In Norway and Italy, RMSE and R^2 of the global FL models did not improve with longer time series. This corresponds to the general poorer performance of the global FL model in these two countries and might be related to smaller data sets and potentially different rotating times in forest management practices in these countries. Tree species and forest structure in the Scandinavian countries Norway, Sweden, and Finland is similar, and, therefore, are unlikely the reason for the poorer performance of the global FL model in Norway compared to Sweden and Finland.

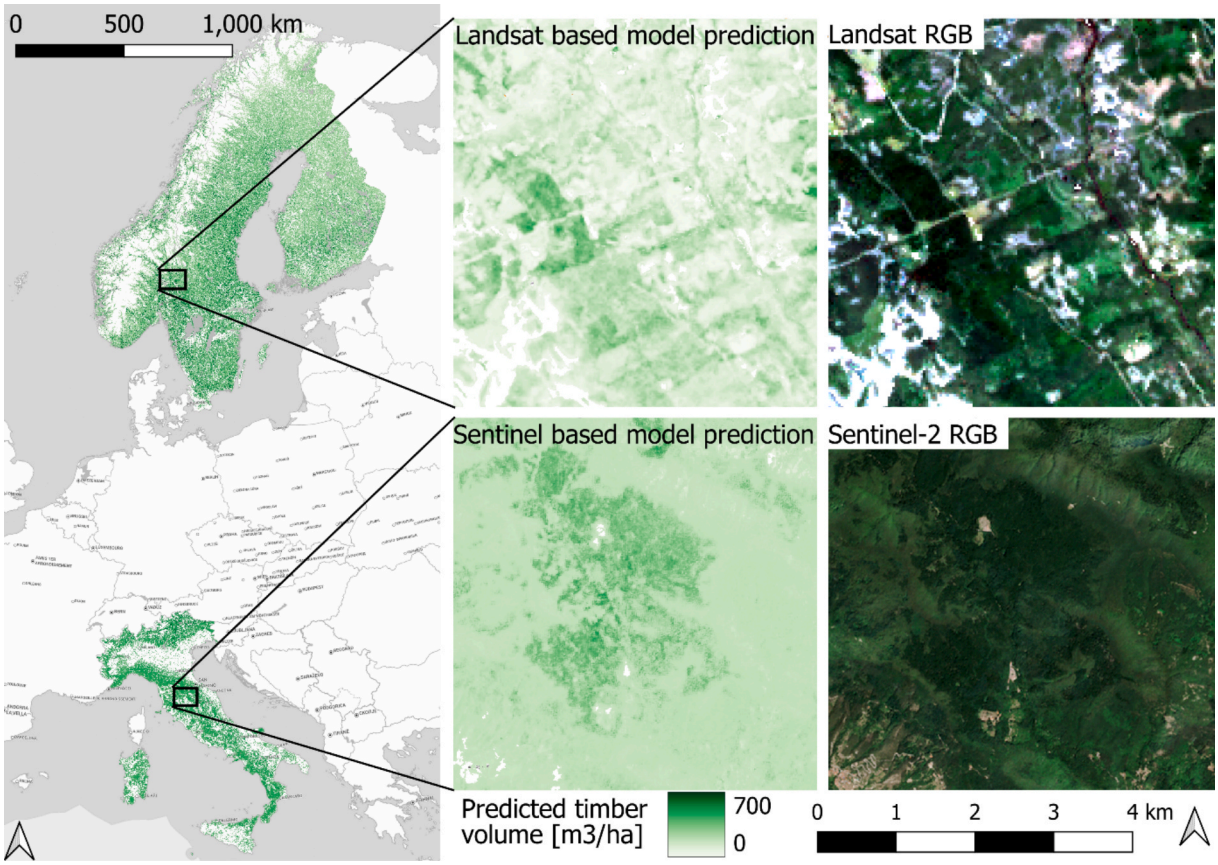


Fig. 9. Overview with Landsat based predictions (left panel), and two close-ups of the prediction maps with corresponding satellite images (center and right panel). The upper close-up represents Landsat based local FL model predictions with corresponding satellite image at the Norway-Sweden border. The lower close-up represents Sentinel based local FL model predictions with corresponding satellite image.

While we used a Time Series CNN model in the present study, the FL approach can be used with other model types as well, where model weights or parameters can be aggregated in a meaningful way. Parametric regression models could be used in FL, however, they are sensitive to multicollinearity among predictors, which is particularly pronounced in multitemporal remote sensing data where spectral indices are highly correlated. Long short-term memory (LSTM) and extreme gradient boosting (XGBoost) models are well-suited for sequential data and should be considered in future studies. However, in the present case with low temporal resolution, potentially noisy time series and limited number of samples per country, LSTMs might be less robust. Extracting remote sensing data in a window around each field plot and using a model architecture as, for example, a UNet can also be considered for future studies. Our scope in this study was to test the FL approach in the context of NFIs and timber volume modelling, and here the tCNN approach provided an efficient, robust way (Miller et al., 2024).

5.5. Limitations and implementation constraints

We did not assess the impact of the number of participating countries on model scalability, nor did we evaluate the generalization of the model to countries or forest types not included in training. The primary aim of this study was to enable collaborative modelling among participating countries while preserving the confidentiality of sensitive NFI data. Extending the framework to additional countries, including those without compatible NFI data, is a relevant and interesting avenue for future work, but was beyond the scope of the present study.

The freely provided FL software implementation (Schumacher, 2026) has been validated between Sweden and Norway. However, due to time constraints in this study, all federated learning clients were simulated on a single server environment, rather than being distributed across multiple independent institutions or devices. Although we tested the framework using various devices within the environment, the implementation did not employ secure aggregation protocols or differential privacy techniques, which are typically used to enhance formal privacy guarantees in federated learning frameworks. Consequently, although this method maintains the confidentiality of raw NFI plot locations and eliminates direct data sharing, it lacks comprehensive privacy protection against potential data leakage. The transmission of model weights from clients to a centralized server may still disclose sensitive information. However, this can be addressed with the framework of differential privacy (Adnan et al., 2022). Furthermore, the modelling relied on pre-processed remote sensing data stacks. These factors should be considered when interpreting the privacy and scalability aspects of the results, and highlight the need for future work to address secure aggregation and privacy-preserving mechanisms in operational deployments.

6. Conclusion

This study establishes FL as a scalable, data confidentiality-preserving solution for forest resource modelling and monitoring. Local FL models demonstrated predictive performance comparable to baseline models, while effectively managing heterogenous data across countries. Our aim was not to find the best performing model for timber volume prediction, but to demonstrate the applicability of FL for combining field and remote sensing data while preserving sensitive data.

A European Forest Monitoring System will require harmonized, cross-border forest assessments while upholding data sovereignty and

privacy. FL directly supports these goals by enabling collaborative model training across national datasets without requiring the exchange of raw data or sensitive geospatial information. This makes FL particularly well-suited for integrating remote sensing and field data across EU member states, facilitating the generation of harmonized timber volume and biomass maps. In doing so, FL not only aligns with the EU's Forest Monitoring objectives but also strengthens the foundation for evidence-based forest policy, sustainable management, and international scientific collaboration.

CRediT authorship contribution statement

Johannes Schumacher: Writing – review & editing, Writing – original draft, Software, Investigation, Formal analysis, Data curation, Conceptualization. **Alessandro Cescatti:** Writing – review & editing. **Gherardo Chirici:** Writing – review & editing, Resources. **Giovanni D'Amico:** Writing – review & editing, Data curation. **Saverio Francini:** Writing – review & editing, Data curation. **Johannes Hertzler:** Writing – review & editing, Data curation. **Lauri Mehtätalo:** Writing – review & editing, Project administration, Funding acquisition. **Gert-Jan Nabuurs:** Writing – review & editing. **Mats Nilsson:** Writing – review & editing, Data curation. **Juho Pitkänen:** Writing – review & editing, Data curation. **Johannes Breidenbach:** Writing – review & editing, Funding acquisition.

Declaration of competing interest

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

Acknowledgements

This study was supported by the Horizon Europe projects MoniFun (GA number: 101134991) and PathFinder (GA number: 101056907).

During the preparation of this work the author(s) used Colab and ChatGPT in order to improve the writing style of some sentences or paragraphs, and for improving and debugging python code, respectively. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Data availability

The data used in this study contains confidential information such as field plot locations and can therefore not be published. Our Federated Learning code example is publicly available at https://flower.ai/apps/johannes/forest-monitoring-example/.

References

Abdi, A.M., Wang, F., 2026. Mapping forest tree species and their uncertainty using Earth observation and National Forest Inventory data: towards operational monitoring in Sweden. Int. J. Remote Sens. 47, 2912–2943. https://doi.org/10.1080/01431161.2026.2625513.

Adnan, M., Kalra, S., Cresswell, J.C., Taylor, G.W., Tizhoosh, H.R., 2022. Federated learning and differential privacy for medical image analysis. Sci. Rep. 12, 1953. https://doi.org/10.1038/s41598-022-05539-7.

Beutel, D.J., Topal, T., Mathur, A., Qiu, X., Fernandez-Marques, J., Gao, Y., Sani, L., Kwing, H.L., Parcollet, T., Gusmão, P.P. de, Lane, N.D., 2020. Flower: A Friendly Federated Learning Framework [WWW Document]. URL https://flower.ai/.

Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., Kiddon, C., Konečný, J.K., Mazzocchi, S., McMahan, H.B., Overveldt, T. Van, Petrou, D., Ramage, D., Roselander, J., 2019. Towards Federated Learning at Scale: System Design, in: Proc. Mach. Learn. Syst. pp. 374–388.

Breidenbach, J., Granhus, A., Hylen, G., Eriksen, R., Astrup, R., 2020. A century of National Forest Inventory in Norway – informing past, present, and future decisions. For. Ecosyst. 7. https://doi.org/10.1186/s40663-020-00261-0.

Casella, B., Riviera, W., Aldinucci, M., Menegaz, G., 2023. MERGE: a model for multi-input biomedical federated learning. Patterns 4. https://doi.org/10.1016/j.patter.2023.100856.

Ceriani, R., Brocco, S., Pepe, M., Oggioni, S., Vacchiano, G., Motta, R., Berretti, R., Ascoli, D., Garbarino, M., Morresi, D., Bassi, F., Fava, F., 2025. Hyperspectral and LiDAR space-borne data for assessing mountain forest volume and biomass. Int. J. Appl. Earth Obs. Geoinf. 141, 104614. https://doi.org/10.1016/j.jag.2025.104614.

Chen, S., Long, G., Shen, T., Jiang, J., 2023. Prompt Federated Learning for Weather Forecasting: Toward Foundation Models on Meteorological Data. ArXiv. https://doi.org/https://doi.org/10.48550/arXiv.2301.09152.

Copernicus, 2015. Copernicus DEM GLO-30. https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_DEM_GLO30#description.

D'Amico, G., Chirici, G., Campetella, G., Canullo, R., Cervellini, M., Chelli, S., Di Biase, R. M., Fattorini, L., Floris, A., Franceschi, S., Francini, S., Giannetti, F., Marcelli, A., Marcheselli, M., Mattioli, W., Notarangelo, M., Papitto, G., Parisi, F., Pisani, C., Saba, E.P., Quatrini, V., Vangi, E., Corona, P., 2025. National Forest Inventory in Italy: new perspectives for forest monitoring. Ann. for. Sci. 82, 35. https://doi.org/10.1186/s13595-025-01303-9.

European Commission, 2021. Communication from the commission to the European parliament, the council, the European Economic and Social Committee and the Committee of the Regions. New EU Forest Strategy for 2030. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021DC0572.

FAO, 2023. Forest Resources Assessment Working Paper. Terms and Definitions. FRA 2025. Rome.

Fassnacht, F.E., White, J.C., Wulder, M.A., Næsset, E., 2023. Remote sensing in forestry: current challenges, considerations and directions. Forestry. https://doi.org/10.1093/forestry/cpad024.

Fattorini, L., Marcheselli, M., Pisani, C., 2006. A three-phase sampling strategy for large-scale multiresource forest inventories. J. Agric. Biol. Environ. Stat. 11, 296–316. https://doi.org/10.1198/108571106X130548.

Francini, S., Hermosilla, T., Coops, N.C., Wulder, M.A., White, J.C., Chirici, G., 2023. An assessment approach for pixel-based image composites. ISPRS J. Photogramm. Remote Sens. 202, 1–12. https://doi.org/10.1016/j.isprsjprs.2023.06.002.

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 Fenn. 48. https://doi.org/10.14214/sf.1095.

Gallios, G., Demattê, J.A.M., Tsakiridis, N., Cardoso, M.C., Kritharoula, A., Tziolas, N., 2026. Federated earth-observation models for collaborative farm-scale soil mapping. Int. J. Appl. Earth Obs. Geoinf. 146, 105067. https://doi.org/10.1016/j.jag.2025.105067.

Gasparini, P., Di Cosmo, L., Floris, A., De Laurentis Editors, D., 2022. Italian National Forest Inventory—Methods and Results of the Third Survey, 1st ed. Springer Cham. https://doi.org/10.1007/978-3-030-98678-0.

Gupta, S., Kumar, S., Chang, K., Lu, C., Singh, P., Kalpathy-Cramer, J., 2023. Collaborative privacy-preserving approaches for distributed deep learning using multi-institutional data. RadioGraphics 43, e220107. https://doi.org/10.1148/rg.220107.

Huete, A.R., 2012. Vegetation indices, remote sensing and forest monitoring. Geogr. Compass 6, 513–532. https://doi.org/10.1111/j.1749-8198.2012.00507.x.

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., Muller, P.-A., 2019. Deep learning for time series classification: a review. Data Min. Knowl. Discov. 33, 917–963. https://doi.org/10.1007/s10618-019-00619-1.

Kennedy, R., Yang, Z., Gorelick, N., Braaten, J., Cavalcante, L., Cohen, W., Healey, S., 2018. Implementation of the LandTrendr Algorithm on Google Earth Engine. Remote Sens. Basel). 10, 691. https://doi.org/10.3390/rs10050691.

Koma, Z., Antropov, O., Cartus, O., Miettinen, J., Breidenbach, J., n.d. UNet-based deep learning models for estimating timber volume and Lorey's height across Nordic countries using optical and radar EO data. Under review.

Korhonen, K.T., Räty, M., Haakana, H., Heikkinen, J., Hotanen, J.-P., Kuronen, M., Pitkänen, J., 2024. Forests of Finland 2019–2023 and their development 1921–2023. Silva Fenn. 58. https://doi.org/10.14214/sf.24045.

Kuronen, M., Räty, J., Packalen, P., Myllymäki, M., 2025. Uncertainty quantification for forest attribute maps with conformal prediction and k-nearest neighbor method. Remote Sens. Environ. 325, 114758. https://doi.org/10.1016/j.rse.2025.114758.

Le, D.-D., Tran, A.-K., Dao, M.-S., Nazmudeen, M.S.H., Mai, V.-T., Su, N.-H., 2022. Federated Learning for Air Quality Index Prediction: An Overview, in: Proc. Int. Conf. Knowl. Syst. Eng. pp. 1–8. https://doi.org/10.1109/KSE56063.2022.9953790.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.

Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V., 2020. Federated Optimization in Heterogeneous Networks. ArXiv. https://doi.org/10.48550/arXiv.1812.06127.

McMahan, B., Moore, E., Ramage, D., Hampson, S., Arcas, B.A. y, 2017. Communication-Efficient Learning of Deep Networks from Decentralized Data, in: Singh, A., Zhu, J. (Eds.), Proc. 20th Int. Conf. Artif. Intell. Stat., Proceedings of Machine Learning Research. PMLR, pp. 1273–1282.

Miettinen, J., Breidenbach, 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., Krajnc, L., Latte, N., Lejeune, P., McCullagh, A., Mionskowski, M., Moreno-Fernández, D., Myllymäki, M., Nilsson, M., Perin, J., Pitkänen, J., Redmond, J., Riedel, T., Schumacher, J., Seitsonen, L., Sirro, L., Skudnik, M., Snorrason, A., Sroga, R., Traub, B., Traustason, B., Westerlund, B., Wurpillot, S., 2025. Pan-European forest maps produced with a combination of earth

observation data and national forest inventory plots. Data Brief 60, 111613. https://doi.org/10.1016/j.dib.2025.111613.
Miller, L., Pelletier, C., Webb, G.I., 2024. Deep learning for satellite image time-series analysis: a review. IEEE Geosci. Remote Sens. Mag. 12, 81–124. https://doi.org/10.1109/MGRS.2024.3393010.
Moreno-Álvarez, S., Paoletti, M.E., Sanchez-Fernandez, A.J., Rico-Gallego, J.A., Han, L., Haut, J.M., 2024. Federated learning meets remote sensing. Expert Syst. Appl. 255, 124583. https://doi.org/10.1016/j.eswa.2024.124583.
Mullissa, A., Vollrath, A., Odongo-Braun, C., Slagter, B., Balling, J., Gou, Y., Gorelick, N., Reiche, J., 2021. Sentinel-1 sar backscatter analysis ready data preparation in google earth engine. Remote Sens. (Basel) 13. https://doi.org/10.3390/rs13101954.
Nabuurs, G.-J., Harris, N., Sheil, D., Palahi, M., Chirici, G., Boissière, M., Fay, C., Reiche, J., Valbuena, R., 2022. Glasgow forest declaration needs new modes of data ownership. Nat. Clim. Chang. 12, 415–417. https://doi.org/10.1038/s41558-022-01343-3.
nFIESTA, 2024. MoniFun, Codelists of Plot Data Target Variables [WWW Document]. URL https://gitlab.com/monifun/monifun_data/-/wikis/codelist_target_variables#nfi-target-variables (accessed 3.6.25).
Papucci, E., Valbuena, R., Roberge, C., Mensah, A.A., Ståhl, G., 2026. A review of forest biomass assessments based on remote sensing reveals progress in methodological quality—but major challenges remain. Forestry 99. https://doi.org/10.1093/forestry/cpag007.
Parisi, F., Vangi, E., Francini, S., D'Amico, G., Chirici, G., Marchetti, M., Lombardi, F., Travaglini, D., Ravera, S., De Santis, E., Tognetti, R., 2023. Sentinel-2 time series analysis for monitoring multi-taxon biodiversity in mountain beech forests. Front. for. Glob. Change 6–2023. https://doi.org/10.3389/ffgc.2023.1020477.
Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. ArXiv. https://doi.org/10.48550/arXiv.1912.01703.
Pelletier, C., Webb, G.I., Petitjean, F., 2019. Temporal convolutional neural network for the classification of satellite image time series. Remote Sens. Basel). 5. https://doi.org/10.3390/rs11050523.
Perbet, P., Guindon, L., Côté, J.-F., Béland, M., 2024. Evaluating deep learning methods applied to Landsat time series subsequences to detect and classify boreal forest disturbances events: the challenge of partial and progressive disturbances. Remote Sens. Environ. 306, 114107. https://doi.org/10.1016/j.rse.2024.114107.
Poudel, S., Sainju, A.M., Upadhyay, K., 2025. Privacy Meets Conservation: Federated Learning's Revolution in Deforestation Detection, in: 2025 IEEE World AI IoT Congr. (AIIoT). pp. 249–258. https://doi.org/10.1109/AIIoT65859.2025.11105254.
Prince, S.J.D., 2025. Understanding Deep Learning. The MIT Press.
Ramírez, D.H., Díaz, L.P., Rahimian, S., García, J.M.A., Peña, B.I., Al-Khazraji, Y., Alarcón, A.G., Fuente, P.G., Soler Garrido, J., Kotsev, A., 2023. Technological Enablers for Privacy Preserving Data Sharing and Analysis. Publications Office of the European Union, Luxembourg. https://doi.org/10.2760/427718.
Schadauer, K., Astrup, R., Breidenbach, J., Fridman, J., Gräber, S., Köhl, M., Korhonen, K.T., Johannsen, V.K., Morneau, F., Päivinen, R., Riedel, T., 2024. Access to exact National Forest Inventory plot locations must be carefully evaluated. New Phytol. 242, 347–350. https://doi.org/10.1111/nph.19564.
Schumacher, J., 2026. FlowerHub: forest-monitoring-example [WWW Document]. URL https://flower.ai/apps/johannes/forest-monitoring-example/ (accessed 4.2.26).
US Forest Service, 2023. Spatial Data Services [WWW Document]. URL https://research.fs.usda.gov/programs/fia/sds (accessed 3.28.26).
Zhong, L., Hu, L., Zhou, H., 2019. Deep learning based multi-temporal crop classification. Remote Sens. Environ. 221, 430–443. https://doi.org/10.1016/j.rse.2018.11.032.