Remote Sensing of Environment 280 (2022) 113205

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Mapping tree species proportions from satellite imagery using spectral–spatial deep learning

Corentin Bolyn [a,b,*], Philippe Lejeune [a], Adrien Michez [a], Nicolas Latte [a]

[a] Uliège - Gembloux Agro-Bio Tech. TERRA Teaching and Research Center - Forest Is Life, Passage des Déportés 2, BE-5030 Gembloux, Belgium
[b] European Commission, Joint Research Centre (JRC), Ispra, Italy

ARTICLE INFO

Edited by Marie Weiss

Keywords:
Spectral–spatial deep learning
Convolutional neural network
Satellite
Land cover
Proportions
Tree species

ABSTRACT

Remote sensing can be used to collect information related to forest management. Previous studies demonstrated the potential of using multispectral satellite imagery for classifying tree species. However, methods that can map tree species in mixed forest stands on a large scale are lacking. We propose an innovative method for mapping the proportions of tree species using Sentinel-2 imagery. A convolutional neural network was used to quantify the per-pixel basal area proportions of tree species considering the neighbouring environment (spectral–spatial deep learning). A nested U-shaped neural network (UNet++) architecture was implemented. We produced a map of the entire Wallonia Region (southern Belgium). Nine species or groups of species were considered: *Spruce* genus, *Oak* genus, *Beech*, *Douglas fir*, *Pine* genus, *Poplar* genus, *Larch* genus, *Birch* genus, and remaining species. The training dataset for the convolutional neural network model was prepared using a map of forest parcels extracted from the public forest administration's geodatabase of Wallonia. The accuracy of the predicted map covering the region was independently assessed using data from the regional forest inventory of Wallonia. A robust assessment method for tree species proportions maps was proposed for assessing the (1) majority species, (2) species composition (presence or absence), and (3) species proportions (proportion values). The achieved value of indicator OA_{maj} (0.73) shows that our approach can map the majority tree species in mixed and pure forest stands. Indicators MS (0.89), MPS (0.72) and MUS (0.83) support that the model can predict the species composition in most cases in the study area. *Spruce* genus, *Oak* genus, *Beech*, and *Douglas fir* achieved the best results, with PAs and UAs close to or higher than 0.70. Particularly, high performance was achieved for detecting *Oak* genus and *Beech* in low area proportions: PAs and UAs higher than 0.70 from the 0.4 proportion. Predicted proportions had a R^2_{adj} of 0.50. The proposed method, which uses spectral–spatial deep learning to map the proportions of tree species, is innovative because it was adapted to the complexity of mixed forests and spatial resolution of current satellite imagery. Additionally, it optimises the use of available forest data in the model conception by considering all pixels from pure stands to highly mixed forest stands. When forest inventories are available in a broad sense, that is, georeferenced areas with the proportions of tree species, this method is highly reproducible and applicable at a large scale, offering potential for use in forest management.

1. Introduction

Forest ecosystems provide many services; to maintain their ecological and socio-economic functions, appropriate policies ensuring sustainable management must be implemented. Understanding current challenges and forest evolution at the regional or national level is essential for enacting a well-thought-out forest policy, particularly in the context of global climate change. Tree species are a key source of information for ecologists and forest managers.

In addition to field inventories, remote sensing is useful for collecting information on forested areas. When applied to satellite data, machine learning methods (Maxwell et al., 2018) and, more recently, deep learning (Yuan et al., 2020) have shown potential in environmental remote sensing research. Tree species classification has evaluated in numerous studies, such as those reviewed in Fassnacht et al. (2016). Open access to high-quality satellite images, including those from Sentinel-2 (S2) satellites, has enabled land cover and land-use mapping in recent years (Phiri et al., 2020).

Most recent studies focusing on medium-resolution satellite imagery involved using pixel-based classification approaches with machine learning methods such as random forests, support vector machines, and artificial neural networks (Zagajewski et al., 2021; Xie

* Corresponding author at: European Commission, Joint Research Centre (JRC), Ispra, Italy.
E-mail address: Corentin.BOLYN@ec.europa.eu (C. Bolyn).

https://doi.org/10.1016/j.rse.2022.113205
Received 17 December 2021; Received in revised form 7 July 2022; Accepted 27 July 2022
Available online 19 August 2022

et al., 2021; Bjerreskov et al., 2021; Grabska et al., 2020; Immitzer et al., 2019; Grabska et al., 2019; Hoscilo and Lewandowska, 2019; Persson et al., 2018; Wessel et al., 2018; Bolyn et al., 2018). Two studies were conducted to evaluate the potential of S2 time series using Bayesian inference (Axelsson et al., 2021) and deep neural networks (Conv1D, AlexNet, and LSTM) (Xi et al., 2021). These studies required the availability of a reference database with pixel-level labels for tree species. In general, the necessary pixels were extracted where the forest was known to be pure for a specific species (mono-specific) or when the species was considered as dominant. For example, in some studies (Bjerreskov et al., 2021; Persson et al., 2018), when the reference data were in the form of a field forest inventory, a proportion criterion based on basal area was used to select only plots with a dominant species (typically >70%). There are two reasons for this. First, the spatial resolution of S2 imagery (10 or 20 m depending on the band) does not allow for delineation of the distinct tree crowns of a forest stand. Thus, one pixel may cover several trees of different species in mixed forest stands. Second, even if the spatial resolution is high, field data describing the forest at the tree level over large areas with a sufficiently precise location are typically not available. Therefore, to label pixels with a high level of certainty, a simple solution is to consider only pure forest stands for a specific species.

Although such studies have successfully classified and mapped tree species, a model built using such data for pure forest stands may not show the same performance when applied to mixed stands. Strahler et al. (1986) proposed an explicit framework of remote sensing models in which pixel-based classification is implied considering a discrete scene model, with the scene composed of discrete elements with boundaries. This discrete scene model is nearly always of the "H-resolution" type. Thus, the element may be individually resolved, as the resolution cells of the image are smaller than the elements. Indeed, classification assumes that measurements are samples of energy-exiting objects that are larger than the resolution cells of the image (Strahler et al., 1986). For tree species mapping, the elements of the scene model were trees. Because local variance of an image is linked to the number of mixed pixels, low local variance for pixel-based classification is assumed. Woodcock and Strahler (1987) studied the local variance of an image as a function of resolution. They highlighted that a peak in local variance occurs at a resolution cell size that is somewhat smaller than the size of objects in the scene, which was approximately 1/2–3/4 in their study. Therefore, the imagery used for classification per pixel should have resolution cells smaller than 1/2–3/4 of the tree size to avoid theoretical accuracy limits in mixed forest stands.

Regarding this limitation, the number of studies that include robust evaluation of the resulting tree species map is limited because mixed pixels are not integrated in the validation scheme. Although the same reference data are typically used to train and validate the model, such data cannot fully represent the mapped area, as pixels are extracted only from pure stands. Therefore, purposive sampling and pragmatic site selection may be acceptable for training a classifier; however, a probabilistic sampling design would improve accuracy assessment (Stehman and Foody, 2019).

Several approaches can be used to match the levels of observation to the levels of organisation: dividing a pixel into constituent elements, aggregating the pixels to match a higher level of organisation, or considering the environment of individual pixels without classifying them (Girard and Girard, 2010). Another possibility is to abandon the semantic classification and quantify the proportions of tree species at the pixel level. Mixture models are the most appropriate approach for high local variance conditions (Woodcock and Strahler, 1987). Using this strategy, Gudex-Cross et al. (2017) used spectral unmixing of multi-temporal Landsat images to quantify the basal area percentage of ten tree species/genera using a stepwise linear regression model. The basal area maps were then refined using a set of object-based rules to produce a thematic forest classification.

Table 1
Forest stand types in Wallonia (southern Belgium). From the regional forest inventory, period 1994–2008; (Alderweireld et al., 2015, 2016).

Forest type	Area (ha)	Percentage
Norway spruce stand	163450	34.09%
Oak stand	85200	17.77%
Other broad-leaved stand	56200	11.72%
Beech stand	43750	9.12%
Noble broad-leaved stand	40100	8.36%
Mixed stand Beech–Oak	21200	4.42%
Other needle-leaved	16850	3.51%
Douglas fir stand	13950	2.91%
Pine stand	12600	2.63%
Poplars stand	9800	2.04%
Mixed stand Norway Spruce–Douglas fir	8850	1.85%
Larch stand	7550	1.57%

In recent years, deep learning, particularly convolutional neural networks (CNNs), have been widely used for remote sensing (Ghanbari et al., 2021), primarily for classification and object detection. CNNs are the most common type of NNs for computer vision and image analysis because of their excellent performance and effectiveness. CNNs are particularly robust because of their specific architecture characterised by local receptive fields, shared weights, and subsampling (Kattenborn et al., 2021). Their potential for tree species classification has been tested in some studies and found to achieve high accuracy (Mäyrä et al., 2021; Xi et al., 2021; Cue La Rosa et al., 2021; Illarionova et al., 2021), outperforming traditional machine learning methods.

The objective of this study was to use spectral–spatial deep learning to map tree species proportions over a large area, including all types of forest compositions from pure stands to highly mixed stands, using S2 imagery and available georeferenced forest areas with tree species proportions. There were two challenges to using the proposed method .

- Mapping tree species in pure and mixed forest stands using images with a pixel size too coarse to resolve individual trees ;
- Using forest inventory data that include both pure and mixed stands.

We used a CNN to quantify the per-pixel proportions of the tree species while considering the neighbouring environment. Methods for directly predicting the per-pixel proportions of tree species using both spectral and spatial information are lacking. The method was performed for an entire administrative region, the Wallonia region (southern Belgium), whose forest presents diversity relevant to the study in terms of its structure (mixed and pure forest stands) and tree species composition. A robust and transparent evaluation of the map accuracy was performed using external data from the regional forest inventory (plots) covering the same region.

2. Materials and methods

2.1. Study area

The study area was the Wallonia region (Fig. 1) in the southern part of Belgium, which covers 16,901 km^2, of which 33% is covered by forests (Alderweireld et al., 2015, 2016). 52% of the forest is private and 87% of the forest area is managed for harvesting. The first complete cycle of the regional forest inventory of Wallonia was performed in 1994–2008. An official report (Alderweireld et al., 2015) divided the forest into 12 stand types that are characteristic of the study area (Table 1). The figures from this report illustrate the variety of forests in Wallonia, with a large area of mixed stands containing a variety of tree species. The three most common genera are Spruce, Oak, and Beech.

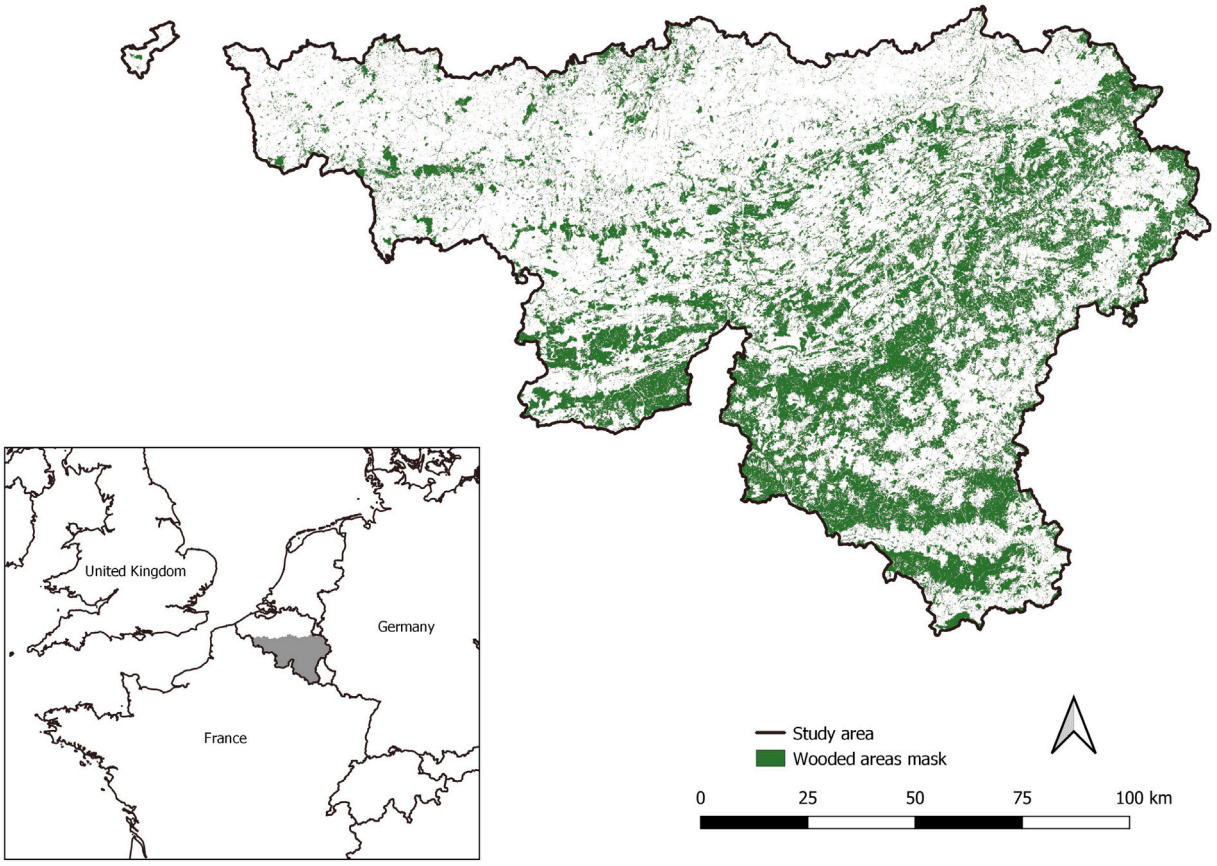

Fig. 1. Geographical location and wooded areas of the study area, Wallonia region (southern Belgium). The study area is presented in grey in the inset map. The wooded areas mask is in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.2. Workflow and datasets

In this study, classes were not mapped individually but rather as vectors of tree species proportions. The tree species proportions are represented as a vector of nine values in the same order as the tree species classes (Table 3). For each proportion, the values ranged from 0 to 1, and the sum of the vectors was equal to 1. This vector of species proportions was the target variable of a CNN model built to produce the tree species proportions map at 2.5 m within the wooded areas of the study area. A pre-existing map for the reference year 2018, at a spatial resolution of 2 m, was used to mask the wooded areas within Wallonia (Fig. 1). This wooded area mask was made as part of a project aimed at generating a "forest mask"(described and freely available at http://geoportail.wallonie.be). The wooded areas mask corresponded to forest areas more than 20 m wide and 0.5 hectares in area. The 10 bands of the S2 imagery were super-resolved at 2.5 m as described by Latte and Lejeune (2020) and used as predictor variables in the CNN model. Two databases were used as references for the target tree species proportion vectors. The first, named as "forest parcel polygons", was used for model training. The second, named as "assessment plots", was used to assess the accuracy of the map (Fig. 2).

2.2.1. S2 super-resolved imagery

The S2 imagery, level 2 A flat reflectance, which was produced and distributed by the Theia Data Center (https://www.theia-land.fr/en/), was used as predictor data (Hagolle et al., 2020) (Fig. 2). The Wallonia region is covered with eight S2 tiles. To obtain a cloud-free and hole-free mosaic with homogeneous radiometry, surface reflectance synthesis was generated considering the vegetation period (from 15 May to 15 September 2018). For this period, all available tiles with cloud cover of less than 50% were downloaded. A total of 86 tiles was selected for 13 dates from 18 May to 19 August. The cloud mask computed using MAJA software was used (Baetens et al., 2019). A time-weighted average of the cloud-free pixels was calculated to give more weight to the mid-summer pixels. The dates of the S2 tiles were converted to numerical values corresponding to the day number in the considered time period of 125 days (from 15 May to 15 September 2018). Values from 1 to 125 were rescaled from −6 to 6. The weights of S2 tiles were then calculated as the densities of a normal distribution (mean = 0, sd = 2.5) for the corresponding date values. Finally, the ten S2 bands (four at 10 m and six at 20 m) of the S2 time-weighted mosaic were super-resolved at 2.5 m as described in Latte and Lejeune (2020). PlanetScope scenes "analytic-sr", covering the study area from 26 June to July 02, were used in this process. The super-resolution was not intended to improve the model performance but rather the geometric accuracy of the map, that is, the boundaries and edges of the mapped forest patches, in accordance with the wooded areas mask (2 m, Fig. 1). Indeed, tree species proportions were still modelled using stand-level information for training (see Section 2.2.2).

2.2.2. Forest parcel polygons

A map of forest parcels extracted from the forest administration's geodatabase (Department of Nature and Forests, Public Service of Wallonia, 2017) was used as the target data for model training (Figs. 2 and 3). This database contains almost 120,000 digitised forest polygons distributed over the entire region (Table 3). For each polygon, the basal area proportion of each tree species was estimated by the responsible forest agents. The proportions of tree species classes (Table 3) were calculated by summing the species-specific proportions. The poplar class was poorly represented in the public geodatabase because a large portion of the poplar stands is managed by private owners in the study area. Thus, 178 additional polygons were digitised by visual

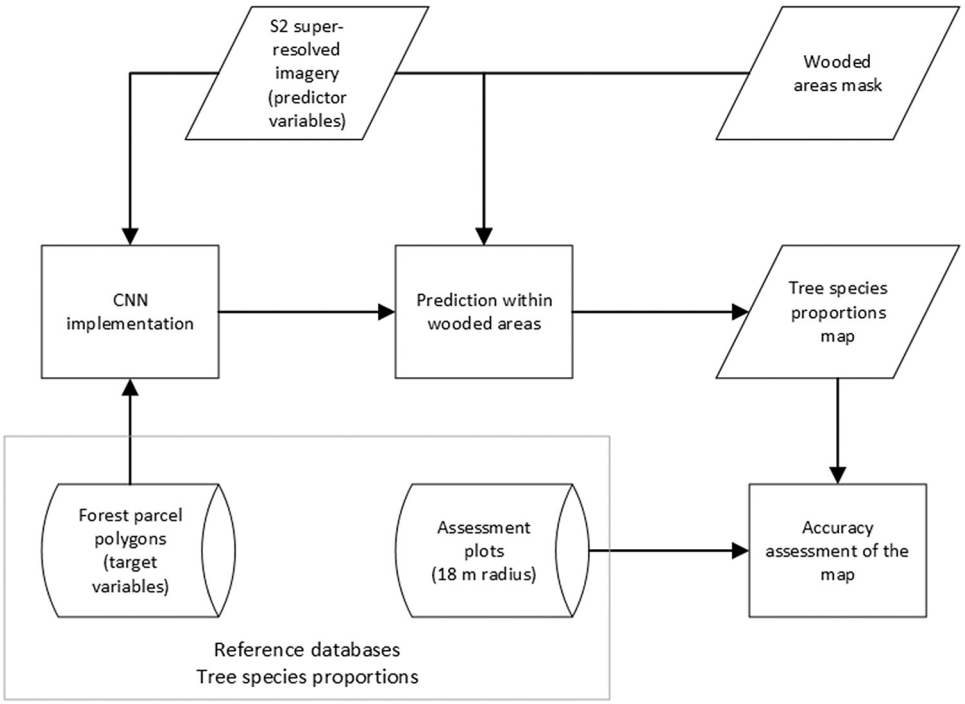

Fig. 2. Workflow of the study.

Table 2
Number of selected plots from the regional forest inventory of Wallonia by year of inventory and their cumulative percentages.

Year	2018	2017	2016	2015	2014	2013	2012	2011	2010	2009	2008
Number of plots	62	295	739	672	714	601	491	408	275	278	211
Cumulative percentage	1	8	23	37	52	65	75	84	90	96	100

interpretation from the orthophotos of the Public Service of Wallonia (http://geoportail.wallonie.be). Only pure plantations were digitised to avoid misclassifications. Poplar plantations are easily recognisable from the sky because of the large and constant distance between plants. A proportion value of 1.0 was assigned to the Poplars class for these polygons.

Using this type of reference polygon to prepare a training dataset improves the reproducibility of the study because these data are often available for forest areas with a formal management plan. Based on the polygon database, the tree species proportions used for model training in the study were calculated using the basal areas of the dominant and understory trees. Notably, diameter at breast height of trees cannot be observed on satellite imagery; thus, we assumed that a CNN model would learn the relationship between the tree crown proportions and basal area proportions, as well as the variation of this relationship between tree species. This likely added uncertainty to the predictions, which is a limitation of this study.

2.2.3. Assessment plots

Another data source, the regional forest inventory of Wallonia (Alderweireld et al., 2015, 2016), was used to independently evaluate the map quality (Fig. 2). Field data from 4,746 plots with an 18 m radius visited since 2008 were used in this assessment (Table 2). More than half of the selected plots were visited after 2013. Plots with young plantations or regenerations, as well as plots without field measurements, were discarded. The selected plots were systematically distributed throughout Wallonia on a one-year basis (Fig. 4).

For each plot, the basal area per hectare (m^2/ha) was calculated by tree species class using field measurements. These values were used to determine the proportions of tree species. A large proportion of these plots were mixed forest stands (a quarter presented a maximum proportion <= 0.64). The representative dataset of this study area confirmed the need to develop a suitable method for mapping mixed forests. The plots were located using geographic positioning system (GPS).

A few points must be noted regarding the data used to assess the map accuracy. First, there is overlap between the assessment plots and forest parcel polygons. Indeed, forest parcel polygons cover a large part of the public forest, and assessment plots are in both public and private forest stands. In the study area, 52% of the forest is private. This may imply an optimistic bias in the map quality assessment. However, the forest parcel polygons used for training and assessment plots differ in nature. The first database consisted of polygons covering variable areas and provided average information for a complete forest stand (Fig. 3). However, locally, this information is approximately false in the worst case. The second database provides local information for a constant area (radius 18 m). This database was used to locally assess the accuracy of the complete final map predicted by pixel of 2.5 m. Second, there were time lags between measurement of the assessment plots and the period covered by the S2 imagery. Finally, regarding the GPS location of the plots, the root mean squared error (RMSE) reached 3–7 m under the canopy, depending on the forest type (Andersen et al., 2009). For assessment, we therefore assumed that in most of the mapped forest stands, the species composition did not change significantly within a maximum radius of 7 m and over a period of up to ten years (Table 2).

2.2.4. Tree species classes

Nine tree species classes were defined (Table 3) to map the basal area proportions of tree species, which are exhaustive with respect to tree species in the study area. These classes were defined according to the most frequent stand types in Wallonia (Table 1).

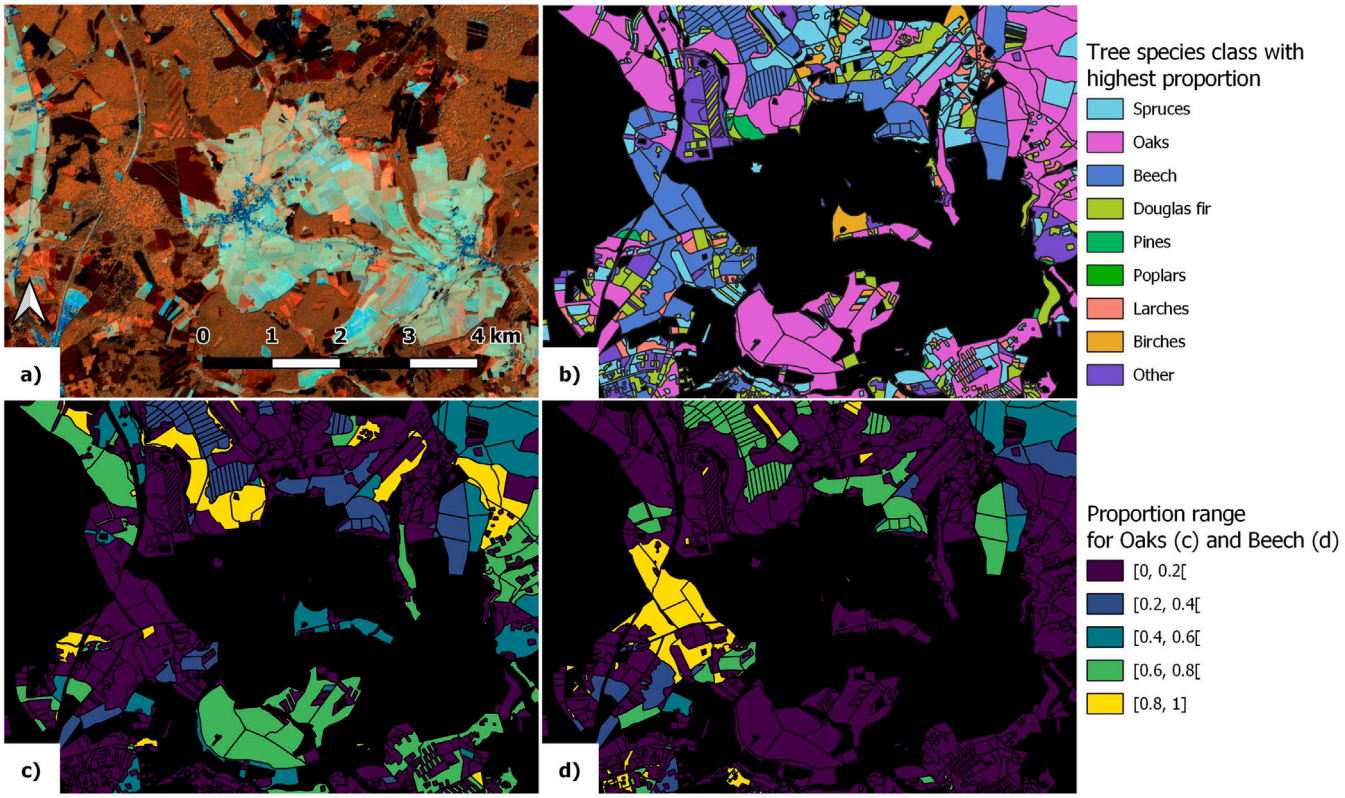

Fig. 3. (a) False colour image of S2 super-resolved imagery (red: B8 A, green: B11, blue: B12). (b) Forest parcel polygons displayed with the tree species class with the highest proportion. (c) Forest parcel polygons displayed with the proportion of Oaks. (d) Forest parcels polygons displayed with the proportion of Beech. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

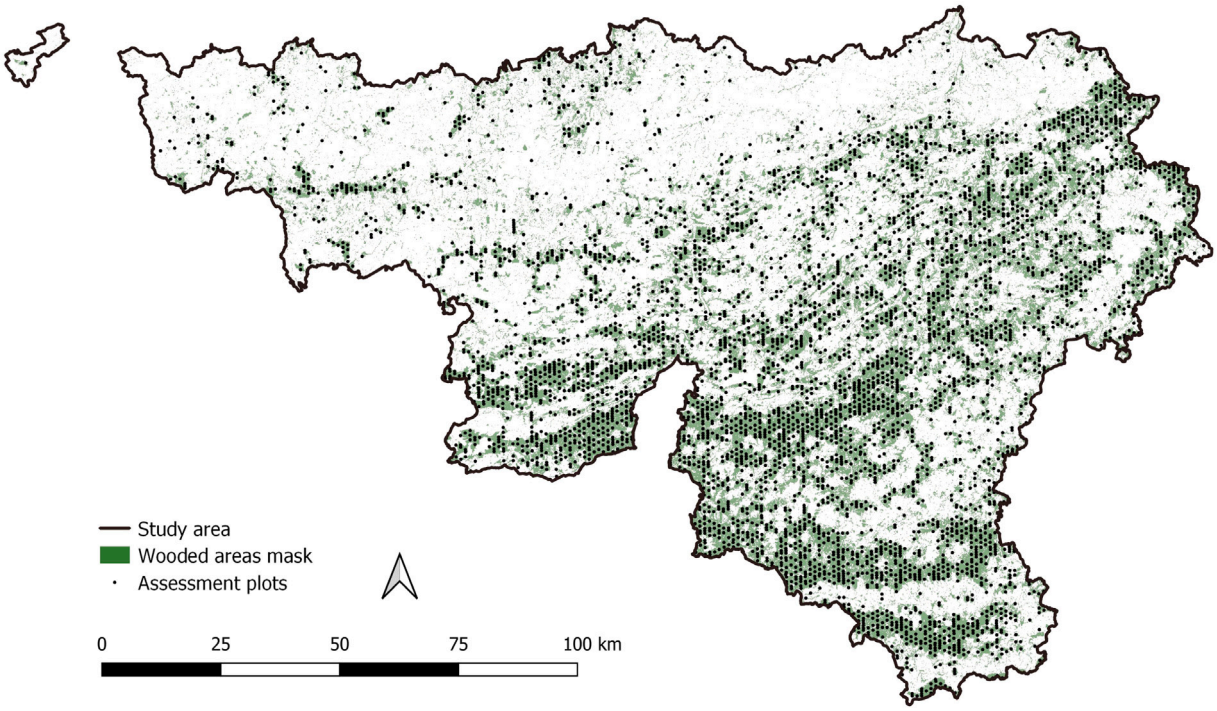

Fig. 4. Accuracy assessment of the tree species proportions map. We selected 4,746 plots of the regional forest inventory of Wallonia (Alderweireld et al., 2016). The wooded areas mask is displayed in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Definitions of tree species classes, number of forest parcel polygons used as target data in the model training (map of forest parcels extracted from the forest administration's geodatabase of the Department of Nature and Forests, Public Service of Wallonia) where the class is present, and number of plots used in map assessment (selected from the regional forest inventory of Wallonia) where the class is present. As a discrete element can contain several tree species classes, a forest parcel polygon or an assessment plot can be counted in more than one class.

Tree species class	definition	forest parcel polygons	assessment plots
Spruces	Tree species belonging to the genus *Picea*.	51560	1517
Oaks	Native oak in the study area, *Quercus robur L.* and *Quercus petraea (Mattuschka) Lieblein.*	21657	2285
Beech	*Fagus sylvatica L.*	25804	1454
Douglas fir	*Pseudotsuga menziesii (Mirb.) Franco.*	19119	344
Pines	Tree species belonging to the genus *Pinus*.	6926	264
Poplars	Hybrid black poplar, *Populus x euramericana (Dode) Guinier.*	839	129
Larches	Tree species belonging to the genus *Larix*.	8539	186
Birches	Tree species belonging to the genus *Betula*.	10806	866
Other	Tree species not included in other tree species classes.	32548	1958
Total number of discrete elements		119991	4746

2.3. CNN

2.3.1. Preparation of predictor and target variables

The predictor data (S2 super-resolved imagery, Section 2.2.1) were normalised by considering the entire region using Eq. (1). Normalising the data is highly recommended, as it speeds up learning and convergence and improves the CNN performance. The data were normalised using the following equation:

$$NormVal_b = \frac{Val_b - P_{1st_b}}{P_{99th_b} - P_{1st_b}} \quad (1)$$

where Val_b denotes the values of the considered band b, P_{1st_b} and P_{99th_b} denote the first and 99th percentiles of the values of band b, and $NormVal_b$ denotes the normalised values of the band b.

Based on the entire set of forest parcel polygons (Section 2.2.2), a multi-band raster was generated (same grid as the predictor data) and used as target data in CNN training. Each band of this raster corresponded to a proportion of tree species (sum of all bands equalling to 1). A 'No Data' value was assigned to all pixels not covered by the forest parcels polygons or outside the wooded areas mask.

2.3.2. CNN architecture

To model the tree species proportions in the form of a vector of nine values as a function of the ten super-resolved S2 bands, a U-shaped neural network (UNet) was implemented. The UNet architecture was initially developed by Ronneberger et al. (2015) and is now among the most widely used architectures for pixel-wise image segmentation, including remote sensing (Iglovikov et al., 2017). Typically, the UNet architecture contains two paths: the contraction path (also known as the encoder) and symmetric expanding path (also known as the decoder). The encoder comprises of several convolution and pooling operations (downsampling). The decoder comprises several transposed convolutions (upsampling). Features of the encoder and decoder were concatenated at each level via skip connections. The output showed the same spatial dimensions as the input.

Since 2015, numerous adaptations and improvements of the UNet architecture have been proposed. In this study, we selected a nested UNet, named as UNet++, which significantly improves the segmentation accuracy (Zhou et al., 2018). Compared with the original UNet, UNet++ has two major modifications: redesigned skip pathways and deep supervision. The former reduces the semantic gap between the encoder and decoder features of subnetworks, and the latter enables the outputs of each level to be combined. This architecture was described in detail in Zhou et al. (2018). The data reconstruction part of the architecture was slightly modified to obtain a vector of nine values for which the sum equals 1 for each pixel as the final output (Softmax activation function).

2.3.3. Patches and batches preparation

When using a CNN in remote sensing, the data sample unit is the patch, which can be described as a small multi-band image extracted from a portion of the study area. In this study, the patch size was set as 400 × 400 pixels. To produce a patch dataset for CNN training, each patch extent was generated by randomly selecting a class by randomly selecting a forest parcel polygon where this class is present, regardless of the proportion; in this polygon, a point was randomly generated. This point was the centre of the 400 × 400 pixel window used to extract data from the predictor raster, i.e. the ten S2 bands, and the target raster, representing the proportions of the nine tree species . As explained in Section 2.3.1, the target values were "No Data" for all pixels not covered by the forest parcel polygons.

Because the CNN could not be trained using all patches together, as it would require too much computer memory, these patches were grouped into distinct batches. Each batch was used separately to train the CNN to update the architecture weights progressively. Therefore, preparing patches and batches is an important step. A total of 7,500 batches was generated, with each batch comprised of three randomly generated patches.

The nine tree species classes were highly imbalanced in terms of the area covered in the region (Table 3). This imbalance was not compensated when constructing the training dataset, as the spatial context is an essential component of learning. Nevertheless, the manner by which the patch extents were generated ensured that each tree species class had an equal chance of being represented in a patch.

The data within the patches were augmented by applying rotations (22.5°, 45°, or 67.5°) and flips (vertical and/or horizontal). For each patch, the probability of augmentation was 50%. Data augmentation acts as a regulariser and helps to reduce overfitting.

2.3.4. Training

UNet++ includes weight regularisations to reduce overfitting and improve model generalisation and robustness. For training, the Adam optimiser (Kingma and Ba, 2014) was used, with a relatively low learning rate of $1e^{-4}$. The number of epochs was 5,000 and number of steps per epoch was 50. For each step, one of 7,500 batches was randomly selected. When reaching the learning plateau (i.e., no further model improvement) (Yoshida and Okada, 2019), the learning rate was progressively reduced by a step of −10% every 100 epochs.

A specific loss implementation is defined. The 'No Data' pixels were not used to compute loss. Thus, only pixels with forest parcel polygon data and inside the wooded areas mask (see Section 2.3.1) were used. Five functions were tested: mean absolute error, mean squared error, pseudo-Huber (PH) (Barron, 2019), focal loss (Lin et al., 2017), and Kullback–Leibler divergence (Erven and Harremos, 2014). The best results were obtained using PH (Eq. (2)). The loss of each batch was

computed by averaging the PH values of all selected pixels (all classes considered).

$$PH = \delta^2 \times (\sqrt{1 + (\frac{RS}{\delta})^2} - 1) \quad (2)$$

where δ (=2) denotes the delta value and RS denotes the residual value for a pixel (true proportion minus predicted proportion).

2.4. Map prediction

Mapping prediction was performed within the wooded areas mask using tiles of 400 × 400 pixels. To avoid undesirable effects at the tile borders resulting from successive convolutions, a grid with an 8-pixel overlap between tiles was used. Before merging the tiles, the 8-pixel overlap of each tile was removed (resulting in 392 × 392 pixels remaining). To further improve the visual aspect, the map was obtained from the average of 16 predictions corresponding to combinations of four shifts (one-quarter left and/or up, or none) and four flips (vertical and/or horizontal, or none).

As the loss function used (Eq. (2)) minimised the error in tree species proportions but did not prevent small error in absent classes (absent class = proportion 0.0), false-positives for small proportions were frequent. Thus, a proportion threshold of 0.2 was applied to improve the consistency of the map. Predicted proportion values below 0.2 were reduced to zero, and the proportions were recalculated to sum to one. The final map was a 9-band raster, with each band corresponding to the proportion of a tree species class.

2.5. Accuracy assessment of the map

2.5.1. Assessment protocol

Classification maps are typically evaluated using a confusion matrix and indices derived from it, such as overall accuracy and producer's and user's accuracies per class (Stehman and Foody, 2019). For this study, because of the specific nature of the prediction, a vector of the proportions of the nine tree species, such traditional methods were not adapted. In the absence of a consensus for assessing compositional data in the literature, we defined a new specific protocol that included new indicators. The evaluation was divided into three parts, each focusing on one specific element: (1) the majority class, (2) species composition (presence or absence), and (3) species proportions (proportion values).

To compare the predicted tree species proportions with those observed in the assessment plots, the predicted values of pixels inside the 18 m radius were extracted and averaged.

2.5.2. Majority class assessment

Plots with a majority class in the field were selected. These plots corresponded to those whose tree species proportion vector contained a class value higher than 0.6 according to the assessment data. Observed and predicted classes with the highest proportion were used to build a traditional confusion matrix and compute three values: the overall accuracy (OA_{maj}), producer's accuracy per class (PA_{maj}), and user's accuracy per class (UA_{maj}). This assessment is most comparable to other studies of tree species classification.

2.5.3. Species composition assessment

The species composition, that is, the presence or absence of tree species classes, was derived from the vector of tree species proportions by applying a simple Boolean filter (vector > 0). Three new indicators were introduced: (1) mean score (MS), (2) mean producer's score (MPS), and (3) mean user's score (MUS). The MS was computed by averaging the proportion of correct attributions within a plot (Eq. (3)). Similar to MS, MPS focused on the presence relative to the reference, and MUS focused on the presence relative to the prediction (Eqs. (4) and (5)):

$$MS = \frac{1}{n}\sum_{i=1}^{n}\frac{c_i}{9} \quad (3)$$

where n denotes the total number of assessment plots and, for plot i, c_i denotes the number of correct attribution between the presence or absence among the nine classes.

$$MPS = \frac{1}{n}\sum_{i=1}^{n}\frac{t_i}{cr_i} \quad (4)$$

where n denotes the total number of assessment plots and, for plot i, t_i denotes the number of true-positives among the number of classes cr_i present in the plot i according to the reference.

$$MUS = \frac{1}{n}\sum_{i=1}^{n}\frac{t_i}{cp_i} \quad (5)$$

where n denotes the total number of assessment plots and, for plot i, t_i denotes the number of true-positives among the number of classes cp_i present in plot i according to the prediction.

The tree species composition was also assessed according to tree species class. Two traditional indicators were used: 1) producer's accuracy (PA, Eq. (6)) and 2) user's accuracy (UA, Eq. (7)).

$$PA_{cl} = \frac{t_{cl}}{pr_{cl}} \quad (6)$$

where, for the evaluated class cl, t_{cl} denotes the number of true-positives among the number of plots pr_{cl} where the class cl is present according to the reference.

$$UA_{cl} = \frac{t_{cl}}{pp_{cl}} \quad (7)$$

where, for the evaluated class cl, t_{cl} denotes the number of true-positives among the number of plots pp_{cl} where the class cl is present according to the prediction.

To evaluate the detection accuracy per species as a function of their proportions, Eqs. (6) and (7) were applied to several subsets of assessment plots by proportion ranges (0–0.2, 0.2–0.4, etc.). To determine PA and UA, the plots were sorted based on the reference and predicted proportions, respectively.

2.5.4. Assessment of tree species proportions

The residuals of the predicted proportions were also analysed. The RMSE, variance of residuals (VAR_{res}), variance of reference proportions (VAR_{tot}), and adjusted coefficient of determination (R^2_{adj}) were calculated. First, the nine classes of the tree species proportion vector were considered (4746 plots × 9 classes). Second, RMSE, VAR_{res}, VAR_{tot}, and R^2_{adj} were calculated by class only using values for the class of interest (4746 plots × 1 class).

2.6. Data processing and analysis tools

Preparation and processing of the predictor and target data, as well as deep learning implementation and map accuracy assessment, were performed in R (R Core Team, 2020), mainly using three R packages: raster (Hijmans, 2019), sf (Pebesma, 2018), and keras (TensorFlow backend) (Allaire and Chollet, 2019), and in connection with the GDAL/OGR library (GDAL/OGR contributors, 2020) and Orfeo ToolBox (Inglada and Christophe, 2009; Grizonnet et al., 2017).

3. Results

3.1. Tree species proportions map

Using our method, we produced a map of tree species proportions in the Wallonia region (Fig. 1). The final map covered 16,901 km^2 at 2.5 m spatial resolution (Fig. 5).

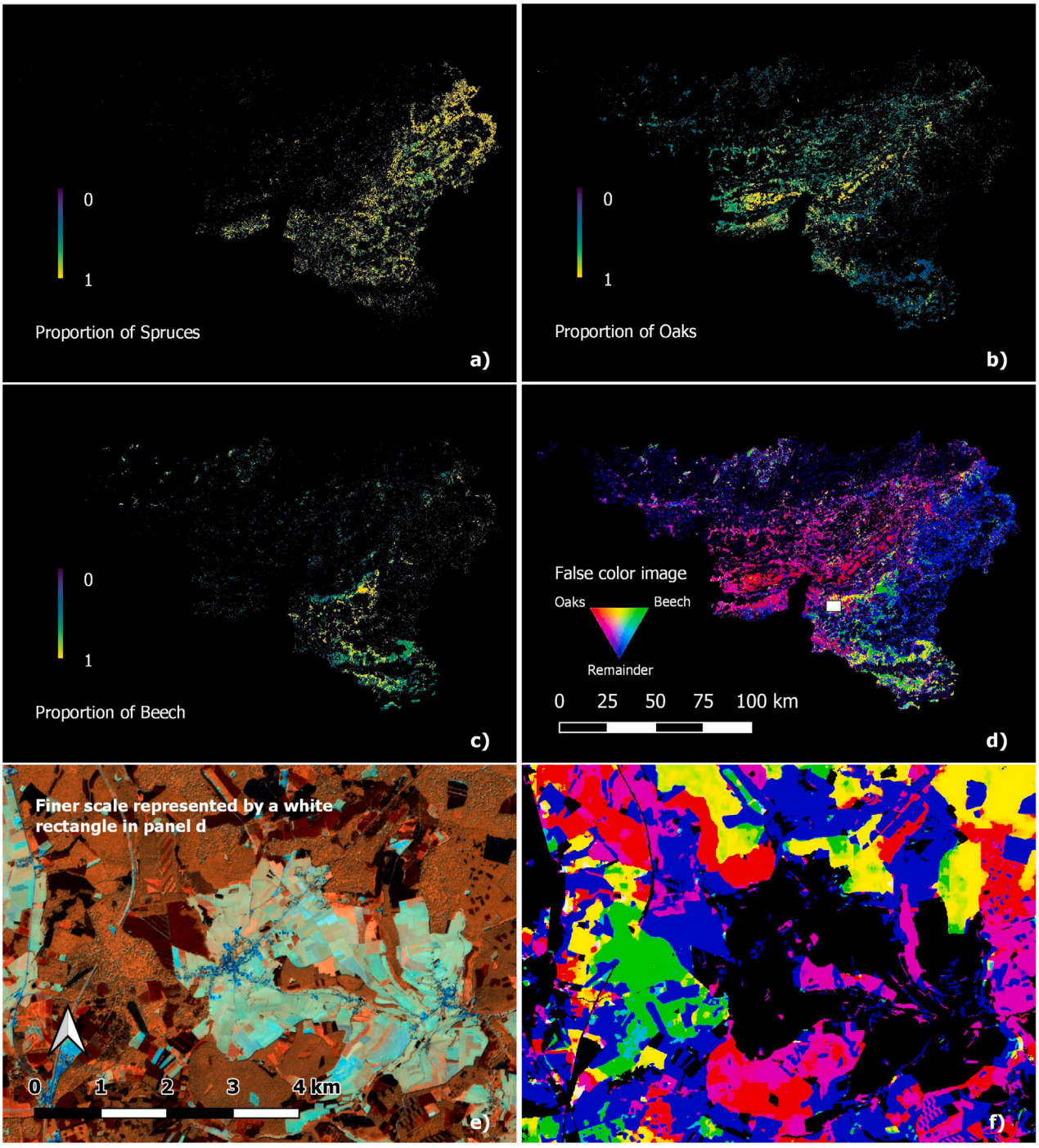

Fig. 5. (a) Tree species proportions map for the Spruces class in the entire study area. (b) Tree species proportions map for the Oaks class. (c) Tree species proportions map for the Beech class. (d) False colour image of the tree species proportions map (red: proportion of Oaks, green: proportion of Beech, blue: sum of proportions for the remaining classes). To display this false colour image, the maximum channel values were set to 0.33. (e) False colour image of the S2 super-resolved imagery (red: B8 A, green: B11, blue: B12) presented for the same extent as panel f. (f) False colour image of the tree species proportions map presented at a finer scale. The extent of the detailed view in panels e and f is represented on the full false colour image (panel d) by a white rectangle. This extent is the same as that in Fig. 3. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. Accuracy assessment of the map

3.2.1. Majority class assessment

The OA_{maj} was 0.73. The Spruces class had the lowest confusion with a PA_{maj} and UA_{maj} above 0.90 (Table 4). The lowest accuracies were observed for Poplars, Birches, and Other classes. The most frequent classes (Spruces, Oaks, Beech, and Douglas fir) showed the highest accuracies. The main confusion was observed between classes often mixed with each other in the assessment plots; such as Oaks and Beech, and Spruces and Douglas fir.

3.2.2. Species composition assessment

The MS value reached 0.89. The MPS and MUS values were 0.72 and 0.83, respectively.

The Spruces, Oaks, Beech, and Douglas fir classes achieved the best global results, with PAs and UAs close to or higher than 0.70 (Table 5).

Table 4
Confusion matrix built to assess detection of the majority class. Assessment plots with an observed class proportion higher than 0.6 were selected. The confusion matrix was calculated by crossing the observed majority class and predicted majority class. The producer's accuracy (PA_{maj}) and user's accuracy (UA_{maj}) were calculated per class.

Prediction	Validation									UA_{maj}
	Spruces	Oaks	Beech	Douglas fir	Pines	Poplars	Larches	Birches	Other	
Spruces	1024	22	20	20	6	0	1	1	21	0.92
Oaks	10	693	77	2	9	1	0	5	83	0.79
Beech	7	74	467	1	3	2	0	0	25	0.81
Douglas fir	45	4	3	128	3	1	2	0	6	0.67
Pines	5	3	2	3	71	0	7	1	5	0.73
Poplars	0	5	1	1	1	27	2	0	20	0.47
Larches	6	4	2	4	2	1	58	0	3	0.72
Birches	12	45	8	1	8	1	2	24	47	0.16
Other	13	256	59	3	2	20	0	3	273	0.43
PA_{maj}	0.91	0.63	0.73	0.79	0.68	0.51	0.81	0.71	0.57	

Table 5
Number of true-positives (t_{cl}), Number of plot where a class is present according to the reference (pr_{cl}), number of plots where a class is present according to the prediction (pp_{cl}); producer's accuracy (PA) and user's accuracy (UA) by class. See Eqs. (6) and (7).

Class	t_{cl}	pr_{cl}	pp_{cl}	PA	UA
Spruces	1178	1517	1312	0.78	0.90
Oaks	1922	2285	2437	0.84	0.79
Beech	967	1454	1224	0.67	0.79
Douglas fir	233	344	332	0.68	0.70
Pines	134	264	202	0.51	0.66
Poplars	59	129	106	0.46	0.56
Larches	77	186	116	0.41	0.66
Birches	87	866	112	0.10	0.78
Other	1163	1958	1526	0.59	0.76

Table 6
Root mean squared error (RMSE), adjusted coefficient of determination (R^2_{adj}), variance of the residuals (VAR_{res}) and variance of the reference proportions (VAR_{tot}).

Class	RMSE	R^2_{adj}	VAR_{res}	VAR_{tot}
Spruces	0.18	0.80	0.03	0.17
Oaks	0.29	0.19	0.09	0.11
Beech	0.21	0.48	0.05	0.09
Douglas fir	0.13	0.54	0.02	0.04
Pines	0.11	0.28	0.01	0.02
Poplars	0.10	0.05	0.01	0.01
Larches	0.09	0.48	0.01	0.02
Birches	0.17	0.05	0.03	0.03
Other	0.29	0.10	0.08	0.09
Overall	0.19	0.50	0.04	0.07

The Oaks class exhibited the best PA with a value of 0.84, whereas Spruces had the best UA with a value of 0.90. The lowest PA and UA were observed for the Birch and Poplars classes with values of 0.10 and 0.56, respectively.

PAs and UAs increased in the range of proportions for all classes. The best PAs were observed for the Oaks class, with values higher than 0.80 in all proportion ranges above 0.2 (Fig. 6). Except for the Poplars, Larches, and Birches classes, all tree species classes had PAs higher than or equal to 0.75 for the proportion range above 0.6 (Fig. 6). The best UAs were observed for the Oaks, Birches and Other classes with values higher than 0.63 in all proportions above 0.2. In the proportion range of 0.6 to 1.0, the UAs were higher than 0.70 for all classes except Poplars. Only the Oaks and Beech classes achieved high Pas and UAs in low proportions. The PAs and UAs were higher than 0.70 from the 0.4 proportion.

3.2.3. Assessment of tree species proportions

The overall R^2_{adj} value was 0.50. The Spruces class reached the best R^2_{adj}. The lowest R^2_{adj} values were obtained for Poplars, Birches, and Other classes (Table 6).

4. Discussion

The method proposed in this study for mapping tree species proportions using spectral–spatial deep learning expands traditional tree species classification and mapping, as it allows for modelling of the majority class, presence/absence of species, and composition basal area proportions. Furthermore, the method was adapted to the complexity of mixed forests and spatial resolution of current satellite imagery. It also optimises the use of available forest inventory data in the model conception by considering all pixels from pure stands to highly mixed forest stands. This method is highly reproducible and applicable at a large scale in cases where forest inventory data, in a broad sense, are available, that is, georeferenced areas with tree species proportions. The training dataset was derived from a map of forest parcels from the Forest Administration, and the resulting map was assessed and validated using a specific protocol that included new indicators with independent data (plots of the regional forest inventory). The assessment covered the full range of possible forest compositions in the study area (n = 4,746 sample plots; Fig. 4).

4.1. Map accuracy

The balance between MPS and MUS (0.72 and 0.83 respectively) and the MS value (0.89) support that the model can predict the tree species composition (presence or absence) in most cases in the study area. For the most frequent classes, Spruces, Oaks, Beech, and Douglas fir (Table 1), the high PA and UA values (Table 5) demonstrate that the model can detect these tree species classes for the majority of situations encountered in the study area. For all classes, there was an increasing proportion of false-negatives in the decreasing observed proportions and increasing proportions of false-positives in the decreasing predicted proportions (Fig. 6). This highlights the expected difficulty in building a model that performs well for low proportions, particularly for less common species. Nevertheless, the high values of PA and UA in the low proportion ranges for the Oaks and Beech classes reveal the potential of this approach for mapping mixed forest stands.

Because the majority tree species of a stand is an important parameter in forestry considerations, we also examined whether the model could detect the majority class when one exists. The confusion matrix (Table 4) showed a high OA_{maj} of 0.73. Lower PA_{maj} or UA_{maj} derived from this matrix were compensated by very good accuracies for the most frequent classes (Spruces, Oaks, Beech, and Douglas fir). These higher accuracies show that our approach can map the majority of tree species in mixed and pure forest stands in the study area. Furthermore, we achieved accuracies comparable to those computed for pure stands only in the latest study conducted in European temperate forests (Hemmerling et al., 2021).

The overall RMSE of the predicted proportions was 0.19 for an R^2_{adj} value of 0.50. This result was expected, as the data used to train the model were polygons providing average information for a

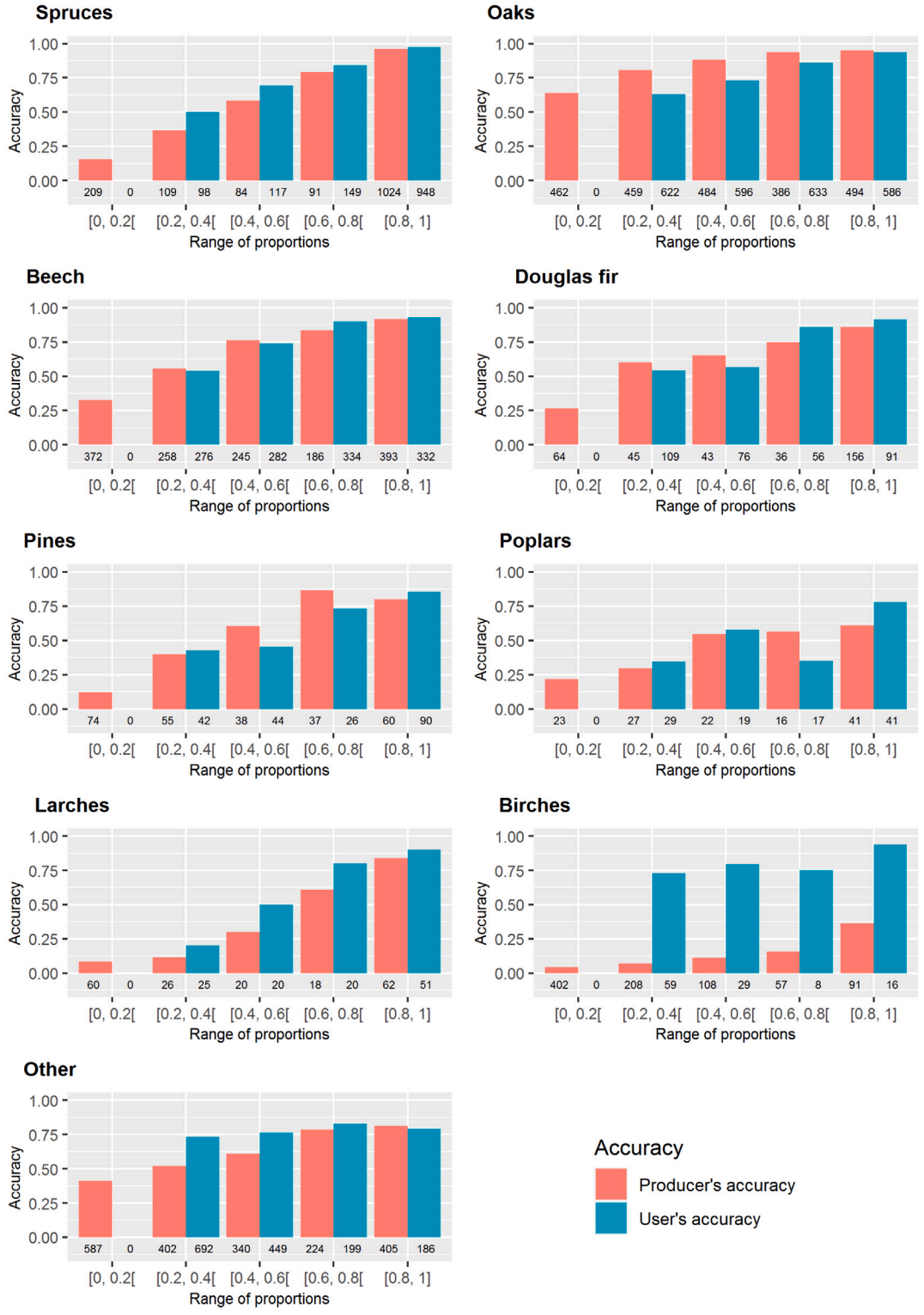

Fig. 6. PAs and UAs by tree species class as a function of the range of proportions. For PAs, the x-axis corresponds to the observed proportions. For UAs, it corresponds to the predicted proportions. The number of plots used to calculate the accuracies (pr_{ci} and pp_{ci}, Eqs. (6) and (7)) is shown below the corresponding bars. As described in Section 2.4, predicted proportion values lower than 0.2 were filtered to prevent false-positives. Thus, the UA bar is absent from the 0–0.2 range in all classes.

complete forest stand. Nevertheless, this accuracy was very satisfactory for analysing the tree species composition and majority tree species in the study area. Spruces, the largest class in the study area, achieved an R^2_{adj} value of 0.80. This class is mostly composed of pure or almost pure stands in the study area. This characteristic is reflected in the distribution of the number of assessment plots per proportion range (Fig. 6). Therefore, the variance in its proportions is higher (Table 6), increasing R^2_{adj}. In addition, this class, as the most common tree species, is more likely to be represented in patches used in CNN training, even when not targeted during patch generation (Section 2.3.3).

Gudex-Cross et al. (2017) performed the most comparable and recent study as the current study. Using spectral unmixing of multi-temporal Landsat imagery, the authors quantified the basal area percentage of ten tree species/genera using a stepwise linear regression model. They also performed validation using 50 plots distributed over their study area (Landsat Row 29, Path 14) and achieved R^2_{adj} values from 0.24 to 0.59 depending on the species. Although these statistics were calculated from a proportionally smaller number of observations, they are comparable to our results, highlighting the difficulty of modelling continuous variables, such as the basal area of an individual tree species. To produce a forest thematic map in their study, post-treatment was applied to the percentage basal area rasters obtained from pixel-based spectral unmixing (object-based hierarchical classification scheme). This step requires the definition of arbitrary clustering choices. Because their final forest thematic map was evaluated after this step, comparison was not relevant. Moreover, the quality of the final map depended on the accuracy of the predicted proportions of tree species. Therefore, to ensure consistency and reproducibility, our method can be used to assess tree species composition and majority classes before a hierarchical classification is chosen, such that the direct focus can be placed on the presence or absence of species in the predicted species proportion vector.

4.2. Innovative method for tree species mapping

Compared to the most recent studies using machine learning or deep learning to process satellite images for tree species classification (Xi et al., 2021; Zagajewski et al., 2021; Xie et al., 2021; Bjerreskov et al., 2021; Grabska et al., 2020; Immitzer et al., 2019; Grabska et al., 2019; Hoscilo and Lewandowska, 2019; Persson et al., 2018; Wessel et al., 2018; Bolyn et al., 2018), the proposed method is not limited to spectral information but also includes spatial information, as recently explored by Illarionova et al. (2021). This innovation, enabled by using a CNN, considers the neighbourhood pixels of the targeted pixel for prediction. Therefore, it is an inclusive method for considering groups of pixels for mapping. The prediction was better and more robust because of the nested levels of observation. The mapped object is spatially variable. A forest stand must be considered at several scales of analysis, from the pixel to the whole patch. In contrast to classical segmentation, the CNN allowed us to learn the segmentation parameters directly with regard to the variable to be predicted. The CNN included this step in the model architecture rather than performing pre-processing for object-based analysis.

Predicting tree species proportions at the pixel level based on their surroundings overcame the two challenges of this study. First, it allowed for mapping of the whole forest, including all types of species compositions (pure or mixed stands), even though the resolution of the satellite image was not sufficient to differentiate individual trees. Thus, our method allows reliable characterisation of forest species over large geographical areas and addresses the lack of relevant studies in this field of research raised by Fassnacht et al. (2016).

Second, this approach allowed us to optimise the use of the available forest inventory data in model training and validation by considering all pixels from pure to highly mixed forest stands. The scarcity of spatial data required for remote sensing studies provides an opportunity for a wide range of applications. A recurrent problem is the availability of a representative training dataset. An ideal dataset composed of the precise positioning of tree crowns and their characteristics is rare. However, forest managers often use databases that reference the proportion of tree species per forest plot, stand, or polygon. Thus, Illarionova et al. (2021) proposed an additional strategy. They first trained a CNN model to find homogeneous areas within each forest stand, presenting a tree species class with a proportion higher than 0.5. They then trained a final model for tree species classification in the detected pure areas. Rather than focusing only on homogeneous areas, our reproducible method enables the use of full forest inventory data. Moreover, a mapping approach via a quantitative variable rather than a qualitative approach can be used to consider the gradients that exist in nature, making this model generalisable to many subjects.

Our method requires some improvements. First, Poplars stands grow and change very rapidly compared to other classes (plantation cycle of 20–25 years). Therefore, in the proportion range 0.8–1.0, the assessment plots data were much more likely to be outdated compared to the other mapped classes. Designing a suitable assessment dataset would improve the analysis of this class.

Second, finding reference data for less frequent classes is an intrinsic problem in model-based mapping. Therefore, in this study, special attention was paid to class balancing during the preparation of patches and batches for CNN training (see Section 2.3.3). Nevertheless, lower results were obtained for minority classes (Poplars, Larches, and Birches). Thus, balancing classes and augmenting data cannot completely compensate for a large lack of data. The best dataset would be that which is most representative of the population variability of every class in the study area. Therefore, new reference data should be collected to improve the model .

Finally, a major improvement to the proposed model would be to add a temporal dimension by upgrading spectral–spatial deep learning to ‘spectral-spatial–temporal’ deep learning. This is supported by recent studies that highlighted the importance of using S2 time series for tree species classification (Xi et al., 2021; Hemmerling et al., 2021). Their findings suggest that this upgrade would significantly improve the model performance.

4.3. Deep learning for remote sensing mapping

We demonstrated that deep learning is advantageous for remote sensing mapping. As discussed above, the use of spectral–spatial modelling (rather than spectral modelling alone) makes a substantial difference in this field, which necessarily involves spatial considerations. Few studies have been conducted on tree species classification using deep learning with satellite imagery, and previous studies did not use spatial information in their architecture (Xi et al., 2021; D’Amico et al., 2021). In addition, considering the temporal dimension, as discussed above, further research of tree species mapping using spectral–spatial–temporal deep learning is needed.

Another advantage of NNs is their flexibility, as CNN models and frameworks can be retrained using a custom dataset for any use case (O’Mahony et al., 2020). Based on their customisability, it is possible to adapt the architecture, loss function, and output data type. In this study, only this tool could learn from compositional data (i.e. vectors of tree species proportions as target data). In addition, considering the high learning performance of CNNs, they have vast potential for use in modelling complex objects, particularly in remote sensing for forestry or other environmental applications.

Finally, deep learning techniques are powerful only when coupled with sufficient data (O’Mahony et al., 2020). Ideally, well-distributed spatial data with good coverage should be used but are not always available. The proposed method makes the best use of available reference data in forested areas.

5. Conclusion

This study addressed two major challenges in tree species mapping. First, we propose a method for predicting tree species basal area proportions that is adapted to the complexity of mixed forests and satellite spatial resolutions coarser than the size of tree crowns. Second, this method optimises the use of available forest management and inventory data by considering all pixels of pure and mixed forest stands. These two advances enabled the production of a robust tree species map for a large geographical area. Particularly, high performance was achieved for detecting Oaks and Beech tree species classes in areas with low species proportions.

We also proposed a robust assessment method for tree species proportions maps that allows separate assessment of the 1) majority species, 2) species composition (presence or absence), and 3) species proportions (proportion values). When forest inventory data, in a broad sense, are available, that is, georeferenced areas with tree species proportions, the method is highly reproducible and allows for remote sensing studies at scales comparable to field forest inventories and for all types of forest compositions.

The use of spatial information, in addition to spectral information, was crucial for achieving the objectives of the study and resulted in high performance. As recent studies of tree species classification using remote sensing have highlighted the importance of using time series in model performance, we will next include a temporal dimension in the architecture model to take advantage of species phenology.

CRediT authorship contribution statement

Corentin Bolyn: Conceptualization, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Philippe Lejeune:** Supervision, Writing – review & editing. **Adrien Michez:** Writing – review & editing. **Nicolas Latte:** Conceptualization, Methodology, Software, Writing – review & editing.

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

The authors acknowledge the Department of Nature and Forests (General Operational Directorate for Agriculture, Natural Resources and Environment, Public Service of Wallonia), and Regional Forest Inventory of Wallonia for providing their geodatabases. This work was supported by the Wallonia Region Forest Administration through the CARTOFOR project, which is part of a 5-year forest research and training plan (2019—2024). The views expressed are purely those of the writers and may in no circumstance be regarded as stating an official position of the European Commission.

References

Alderweireld, M., Burnay, F., Pitchugin, M., Lecomte, H., 2015. Inventaire Forestier Wallon - Résultats 1994–2012. SPW, URL: http://orbi.ulg.ac.be/handle/2268/181169.

Alderweireld, M., Rondeux, J., Latte, N., Hébert, J., Lecomte, H., 2016. Belgium (Wallonia). In: Vidal, C., Alberdi, I.A., Hernández Mateo, L., Redmond, J.J. (Eds.), National Forest Inventories. Springer International Publishing, Cham, pp. 159–179. http://dx.doi.org/10.1007/978-3-319-44015-6_8, URL: http://link.springer.com/10.1007/978-3-319-44015-6_8.

Allaire, J.J., Chollet, F., 2019. Keras: R Interface to 'Keras'. URL: https://keras.rstudio.com.

Andersen, H.E., Clarkin, T., Winterberger, K., Strunk, J., 2009. An accuracy assessment of positions obtained using survey- and recreational-grade global positioning system receivers across a range of forest conditions within the tanana valley of interior alaska. West. J. Appl. For. 24 (3), 128–136. http://dx.doi.org/10.1093/wjaf/24.3.128, URL: https://academic.oup.com/wjaf/article/24/3/128/4683488.

Axelsson, A., Lindberg, E., Reese, H., Olsson, H., 2021. Tree species classification using Sentinel-2 imagery and Bayesian inference. Int. J. Appl. Earth Obs. Geoinf. 100, 102318. http://dx.doi.org/10.1016/j.jag.2021.102318.

Baetens, L., Desjardins, C., Hagolle, O., 2019. Validation of copernicus sentinel-2 cloud masks obtained from MAJA, Sen2Cor, and FMask processors using reference cloud masks generated with a supervised active learning procedure. Remote Sens. 11 (4), http://dx.doi.org/10.3390/rs11040433, Number: 4.

Barron, J.T., 2019. A general and adaptive robust loss function. CVPR.

Bjerreskov, K.S., Nord-Larsen, T., Fensholt, R., 2021. Classification of nemoral forests with fusion of multi-temporal sentinel-1 and 2 data. Remote Sens. 13 (5), 950. http://dx.doi.org/10.3390/rs13050950, URL: https://www.mdpi.com/2072-4292/13/5/950.

Bolyn, C., Michez, A., Gaucher, P., Lejeune, P., Bonnet, S., 2018. Forest mapping and species composition using supervised per pixel classification of Sentinel-2 imagery. Biotechnol. Agron. Soc. Environ. 22 (3).

Cue La Rosa, L., Sothe, C., Feitosa, R., Almeida, C., Schimalski, M., Oliveira, D., 2021. Multi-task fully convolutional network for tree species mapping in dense forests using small training hyperspectral data. ISPRS J. Photogramm. Remote Sens. 179, 35–49. http://dx.doi.org/10.1016/j.isprsjprs.2021.07.001.

D'Amico, G., Francini, S., Giannetti, F., Vangi, E., Travaglini, D., Chianucci, F., Mattioli, W., Grotti, M., Puletti, N., Corona, P., Chirici, G., 2021. A deep learning *approach* for automatic mapping of poplar plantations using Sentinel-2 imagery. GISci. Remote Sens. 58 (8), 1352–1368. http://dx.doi.org/10.1080/15481603.2021.1988427, URL: https://www.tandfonline.com/doi/full/10.1080/15481603.2021.1988427.

Erven, T.v., Harremos, P., 2014. Rényi divergence and Kullback-Leibler divergence. IEEE Trans. Inform. Theory 60 (7), 3797–3820. http://dx.doi.org/10.1109/TIT.2014.2320500, Number: 7.

Fassnacht, F., Latifi, H., Stereńczak, K., Lefsky, M., Straub, C., Waser, L., Ghosh, A., Modzelewska, A., 2016. Review of studies on tree species classification from remotely sensed data. Remote Sens. Environ. 186, 64–87.

GDAL/OGR contributors, 2020. GDAL/OGR Geospatial Data Abstraction Software Library. Open Source Geospatial Foundation, URL: https://gdal.org.

Ghanbari, H., Mahdianpari, M., Homayouni, S., Mohammadimanesh, F., 2021. A meta-analysis of convolutional neural networks for remote sensing applications. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 3602–3613. http://dx.doi.org/10.1109/JSTARS.2021.3065569, URL: https://ieeexplore.ieee.org/document/9376238/.

Girard, M.C., Girard, C.M., 2010. Traitement Des Données de Télédétection. Dunod, Paris, OCLC: 1091131298.

Grabska, E., Frantz, D., Ostapowicz, K., 2020. Evaluation of machine learning algorithms for forest stand species mapping using Sentinel-2 imagery and environmental data in the Polish Carpathians. Remote Sens. Environ. 251, 112103. http://dx.doi.org/10.1016/j.rse.2020.112103.

Grabska, E., Hostert, P., Pflugmacher, D., Ostapowicz, K., 2019. Forest stand species mapping using the sentinel-2 time series. Remote Sens. 11 (10), 1197. http://dx.doi.org/10.3390/rs11101197.

Grizonnet, M., Michel, J., Poughon, V., Inglada, J., Savinaud, M., Cresson, R., 2017. Orfeo ToolBox: open source processing of remote sensing images. Open Geospatial Data Softw. Stand. 2 (1), 15. http://dx.doi.org/10.1186/s40965-017-0031-6, Number: 1.

Gudex-Cross, D., Pontius, J., Adams, A., 2017. Enhanced forest cover mapping using spectral unmixing and object-based classification of multi-temporal Landsat imagery. Remote Sens. Environ. 196, 193–204. http://dx.doi.org/10.1016/j.rse.2017.05.006.

Hagolle, O., Colin, J., Kettig, P., d'Angelo, P., Auer, S., Doxani, G., Desjardins, C., Inglada, J., 2020. Sentinel-2 surface reflectance products generated by THEIA and DLR: methods, validation and applications. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. Publisher: Copernicus Publications.

Hemmerling, J., Pflugmacher, D., Hostert, P., 2021. Mapping temperate forest tree species using dense Sentinel-2 time series. Remote Sens. Environ. 267, 112743. http://dx.doi.org/10.1016/j.rse.2021.112743, URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425721004636.

Hijmans, R.J., 2019. Raster: Geographic Data Analysis and Modeling. URL: https://CRAN.R-project.org/package=raster.

Hoscilo, A., Lewandowska, A., 2019. Mapping forest type and tree species on a regional scale using multi-temporal sentinel-2 data. Remote Sens. 11 (8), 929. http://dx.doi.org/10.3390/rs11080929.

Iglovikov, V., Mushinskiy, S., Osin, V., 2017. Satellite imagery feature detection using deep convolutional neural network: A kaggle competition. ArXiv abs/1706.06169.

Illarionova, S., Trekin, A., Ignatiev, V., Oseledets, I., 2021. Tree species mapping on sentinel-2 satellite imagery with weakly supervised classification and object-wise sampling. Forests 12 (10), 1413. http://dx.doi.org/10.3390/f12101413, URL: https://www.mdpi.com/1999-4907/12/10/1413.

Immitzer, M., Neuwirth, M., Böck, S., Brenner, H., Vuolo, F., Atzberger, C., 2019. Optimal input features for tree species classification in central europe based on multi-temporal sentinel-2 data. Remote Sens. 11 (22), 2599. http://dx.doi.org/10.3390/rs11222599.

Inglada, J., Christophe, E., 2009. The Orfeo Toolbox remote sensing image processing software. In: 2009 IEEE International Geoscience and Remote Sensing Symposium. vol. 4, http://dx.doi.org/10.1109/IGARSS.2009.5417481, pp. IV–733. Journal Abbreviation: 2009 IEEE International Geoscience and Remote Sensing Symposium Num Pages: IV-736.

Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. ISPRS J. Photogramm. Remote Sens. 173, 24–49. http://dx.doi.org/10.1016/j.isprsjprs.2020.12.010, URL: https://www.sciencedirect.com/science/article/pii/S0924271620303488.

Kingma, D., Ba, J., 2014. Adam: A method for stochastic optimization. In: International Conference on Learning Representations.

Latte, N., Lejeune, P., 2020. PlanetScope radiometric normalization and sentinel-2 super-resolution (2.5 m): A straightforward spectral-spatial fusion of multi-satellite multi-sensor images using residual convolutional neural networks. Remote Sens. 12 (15), 2366. http://dx.doi.org/10.3390/rs12152366.

Lin, T.Y., Goyal, P., Girshick, R.B., He, K., Dollár, P., 2017. Focal loss for dense object detection. In: 2017 IEEE International Conference on Computer Vision. ICCV, pp. 2999–3007.

Maxwell, A., Warner, T., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: An applied review. Int. J. Remote Sens. 39, 2784–2817. http://dx.doi.org/10.1080/01431161.2018.1433343.

Mäyrä, J., Keski-Saari, S., Kivinen, S., Tanhuanpää, T., Hurskainen, P., Kullberg, P., Poikolainen, L., Viinikka, A., Tuominen, S., Kumpula, T., Vihervaara, P., 2021. Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks. Remote Sens. Environ. 256, 112322. http://dx.doi.org/10.1016/j.rse.2021.112322, URL: https://www.sciencedirect.com/science/article/pii/S0034425721000407.

O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G.V., Krpalkova, L., Riordan, D., Walsh, J., 2020. Deep learning vs. traditional computer vision. In: Arai, K., Kapoor, S. (Eds.), Advances in Computer Vision. vol. 943, Springer International Publishing, Cham, pp. 128–144. http://dx.doi.org/10.1007/978-3-030-17795-9_10, URL: http://link.springer.com/10.1007/978-3-030-17795-9_10. Series Title: Advances in Intelligent Systems and Computing.

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

Persson, M., Lindberg, E., Reese, H., 2018. Tree species classification with multi-temporal sentinel-2 data. Remote Sens. 10 (11), 1794. http://dx.doi.org/10.3390/rs10111794.

Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V.R., Murayama, Y., Ranagalage, M., 2020. Sentinel-2 data for land cover/Use mapping: A review. Remote Sens. 12 (14), 2291. http://dx.doi.org/10.3390/rs12142291, URL: https://www.mdpi.com/2072-4292/12/14/2291. Number: 14 Publisher: Multidisciplinary Digital Publishing Institute.

R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL: https://www.R-project.org/.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention. MICCAI 2015, In: Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 234–241. http://dx.doi.org/10.1007/978-3-319-24574-4_28.

Stehman, S.V., Foody, G.M., 2019. Key issues in rigorous accuracy assessment of land cover products. Remote Sens. Environ. 231, 111199. http://dx.doi.org/10.1016/j.rse.2019.05.018, URL: https://www.sciencedirect.com/science/article/pii/S0034425719302111.

Strahler, A.H., Woodcock, C.E., Smith, J.A., 1986. On the nature of models in remote sensing. Remote Sens. Environ. 20 (2), 121–139. http://dx.doi.org/10.1016/0034-4257(86)90018-0, URL: https://linkinghub.elsevier.com/retrieve/pii/0034425786900180.

Wessel, M., Brandmeier, M., Tiede, D., 2018. Evaluation of different machine learning algorithms for scalable classification of tree types and tree species based on sentinel-2 data. Remote Sens. 10 (9), 1419. http://dx.doi.org/10.3390/rs10091419.

Woodcock, C.E., Strahler, A.H., 1987. The factor of scale in remote sensing. Remote Sens. Environ. 21 (3), 311–332.

Xi, Y., Ren, C., Tian, Q., Ren, Y., Dong, X., Zhang, Z., 2021. Exploitation of time series sentinel-2 data and different machine learning algorithms for detailed tree species classification. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 7589–7603. http://dx.doi.org/10.1109/JSTARS.2021.3098817, URL: https://ieeexplore.ieee.org/document/9495140/.

Xie, B., Cao, C., Xu, M., Duerler, R., Yang, X., Bashir, B., Chen, Y., Wang, K., 2021. Analysis of regional distribution of tree species using multi-seasonal sentinel-1&2 imagery within google earth engine. Forests 12 (5), 565. http://dx.doi.org/10.3390/f12050565.

Yoshida, Y., Okada, M., 2019. Data-dependence of plateau phenomenon in learning with neural network — Statistical mechanical analysis. In: Wallach, H., Larochelle, H., Beygelzimer, A., Alché-Buc, F., Fox, E., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 32. Curran Associates, Inc., pp. 1722–1730.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Weiwei, T., Yang, Q., Wang, J., Gao, J., Zhang, L., 2020. Deep learning in environmental remote sensing: Achievements and challenges. Remote Sens. Environ. 241, 111716. http://dx.doi.org/10.1016/j.rse.2020.111716.

Zagajewski, B., Kluczek, M., Raczko, E., Njegovec, A., Anca, D., Kycko, M., 2021. Comparison of random forest, support vector machines, and neural networks for post-disaster forest species mapping of the Krkonoše/karkonosze transboundary biosphere reserve. Remote Sens. 13 (13), 2581. http://dx.doi.org/10.3390/rs13132581.

Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J., 2018. UNet++: A nested U-net architecture for medical image segmentation. In: Stoyanov, D., Taylor, Z., Carneiro, G., Syeda-Mahmood, T., Martel, A., Maier-Hein, L., Tavares, J.a.M.R., Bradley, A., Papa, J.a.P., Belagiannis, V., Nascimento, J.C., Lu, Z., Conjeti, S., Moradi, M., Greenspan, H., Madabhushi, A. (Eds.), Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer International Publishing, Cham, pp. 3–11.