Remote Sensing of Environment 223 (2019) 257–273

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region

Heikki Astola*, Tuomas Häme, Laura Sirro, Matthieu Molinier, Jorma Kilpi

VTT Technical Research Centre of Finland, Tekniikantie 1, Espoo, P.O. Box 1000, FI-02044 VTT, Finland

ARTICLE INFO

Keywords:
Sentinel-2
Landsat 8
Forestry
Boreal forest
Forest variables

ABSTRACT

We compared the performance of Sentinel-2 and Landsat 8 data for forest variable prediction in the boreal forest of Southern Finland. We defined twelve modelling setups to train multivariable prediction models with either multilayer perceptron (MLP) or regression tree models with the brute force forward selection method. The reference data consisted of 739 circular field plots that had been collected by the Finnish Forest Centre concurrently with the Sentinel-2 and Landsat 8 acquisitions. The input data were divided into training, validation and test sets of equal sizes for 100 iterations in each modelling setup. The predicted forest variables were stem volume (V), stem diameter (D), tree height (H) and basal area (G), and their species-wise components for pine (Pine), spruce (Spr) and broadleaved (BL) trees. We recorded the performance figures and the best predictive image bands for each modelling setup.

The best average performance over the 100 modelling iterations was obtained using all Sentinel-2 bands. The plot-level relative root mean square errors (RMSE%) of the field observed mean were 38.4% for average stem diameter, 42.5% for stem basal area/ha, 30.4% for average tree height, and 59.3% for growing stock volume/ha with variables including all tree species. The corresponding best figures with all Landsat 8 bands were RMSE % = 44.6%, 50.2%, 36.6% and 72.2%, respectively. The Sentinel-2 outperformed Landsat 8 also when using near-equivalent image bands and Sentinel-2 data down-sampled to 30 m pixel resolution. The relative systematic error (bias%) did not show any significant differences between Sentinel-2 and Landsat 8 data: the average of the absolute value of bias% was 0.8% for Sentinel-2 and 1.2% for Landsat 8. The best predictive Sentinel-2 image band was the red-edge 1 (B05_RE1), when variable totals including all species were estimated. The short-wave infrared bands (B11_SWIR1 & B12_SWIR2) and the visible green band (B03_Green) were also among the best predictors. The median number of predictors in the best performing models was 4–6 for the Sentinel-2 and 4–5 for the Landsat 8 models, respectively.

We conclude that Sentinel-2 Multispectral Instrument (MSI) data can be recommended as the principal Earth observation data source in forest resources assessment.

1. Introduction

Forests cover almost one third of the Earth's land surface (FAO, 2015) and they play a major role in global carbon and water cycles (IPCC, 2006; UNFCCC, 2014; UNFCCC, 2016), but are also an important source of raw material for industry, fuel and other ecosystem services (Binder et al., 2017). The multiple functions of forests can lead to conflicting requirements for their management, with e.g. sustainability as the leading demand from environmental point of view. Reliable data on forest resources is required to provide objective information for sustainable forest management.

Forest inventory using traditional methods is slow and expensive. Traditionally, inventory uses a statistical sample of field inventory plots to provide statistical data on forest resources using unbiased procedures. Such inventory is typically done for natural forest resources assessment, but a full-scale inventory is available only in a fraction of countries globally (FAO, 2000). The inventory is done by wall-to-wall mapping for forest management purposes, and in some cases at a national level (Reese et al., 2003). Such inventory, when applicable for forest management planning, does not necessarily fulfil requirements of unbiased procedures because no statistical sampling is involved. Alternatives, such as participatory sensing, can provide faster and cost-effective collection of in-situ forest and biomass data (Molinier et al., 2016), however not yet at a national scale with the desired density of observations.

The forest variable gaining the most interest in the remote sensing

* Corresponding author.
E-mail address: heikki.astola@vtt.fi (H. Astola).

https://doi.org/10.1016/j.rse.2019.01.019
Received 15 October 2017; Received in revised form 10 December 2018; Accepted 15 January 2019
Available online 28 January 2019

research community in recent years has been the growing stock volume (Chrysafis et al., 2017; Antropov et al., 2018; Lindberg and Hollaus, 2012). However, growing stock volume is not the only adequate variable for estimation for forest management purposes, nor for fine resolution forest primary productivity modelling (Mäkelä et al., 2007; Härkönen et al., 2011). A high number of variables have been estimated in national forest inventories in Finland and Sweden, but much less has been published for variables other than volume globally.

Satellite imagery has offered an information source for forest resources information with satisfactory resolution since the Launch of Landsat-1 in 1972, and in particular Landsat-4 with 30-meter spatial resolution of the Thematic Mapper sensor in 1982. Today, long and dense Landsat data time series are available worldwide, which enables advanced land and forest cover monitoring (Cohen and Goward, 2017; Wulder et al., 2012; Zhu and Woodcock, 2014; Griffiths et al., 2014).

The relative root mean square error of growing stock volume obtained with medium resolution optical satellite data has varied depending on the geographical area and the type of reference data that have been used. Typical values for growing stock volume have been 48–56% in boreal forest (Hyyppä et al., 2000; Mäkelä and Pekkarinen, 2004; Tokola et al., 2007). In tropical forest with anthropogenic influence, a relative RMSE of 44.2% was achieved using regression analysis with ALOS AVNIR data (Häme et al., 2013).

Higher accuracies have been achieved with very high resolution (VHR) optical data that enables extraction of information of forest structure through textural features and stereo imaging (Persson, 2016). The relative RMSE of the growing stock volume predictions has varied from 31.5% (Peuhkurinen et al., 2008) to 37.4% (Astola et al., 2004) in boreal forest with Ikonos-2 data of one-meter resolution. A combination of optical data with data from other sources has resulted in even better accuracies: Persson et al. (2013) reported stand-level relative RMSE's of 22.4%–29.2% for aboveground biomass using optical stereo imaging with SPOT5 and ALOS PRISM with a combined digital terrain model retrieved from laser scanning data.

Of other information sources, space-borne radar data has the benefit of being almost independent of weather conditions. With L-band imaging space-borne radar, the results for growing stock volume have been similar to those with optical bands, with relative RMSE's being 44.2% at sample plot level in managed tropical forest (Häme et al., 2013) and 43% at the stand level of several hectares in boreal forest (Antropov et al., 2013). The potential of the higher frequency C-band has been low in forest inventory. Santoro et al. (2011) obtained relative RMSE values ranging from 47.7%–96.2% with 100 m pixel size using C-band Envisat ASAR in study areas in Russia and Sweden. Application of aerial LiDAR data enables forest canopy height prediction with a stand level RMSE's for stem volume of 13.7% (Persson and Fransson, 2017) and plot level RMSE's of 21.3% (Varvia et al., 2017), 24.9%–34.9% (Kankare et al., 2013) and 37.3%–41.9% (Lindberg and Hollaus, 2012). However, the price, poorer availability and more complex processing decrease the practical usefulness of VHR, radar and LiDAR data for operational applications and for wall-to-wall mapping of wide areas.

The European Copernicus[1] program with its Sentinel satellites is an operational environment monitoring system providing free satellite data for decades to come. The main Copernicus instrument for forestry purposes is the Sentinel-2 satellite series.[2] Two satellites provide images from 10- to 60-meter spatial resolution with 13 spectral bands and five-day revisit frequency at the equator (Drusch et al., 2012). The operational nature of the Copernicus program facilitates the development of sustainable services built on Sentinel-2 data. The Landsat program is also operational, but the 16-day image acquisition frequency is much lower than with Sentinel-2 (https://landsat.usgs.gov/landsat-8-l8-data-users-handbook-section-2). All other publicly available data with adequate resolution have been acquired with satellites from programs with uncertain future continuation.

Trustworthy data service is of central importance for forest industries and administrations. The improved spatial resolution of the multispectral data from 30 m of Landsat to ten meters of Sentinel-2 make a big difference for the operational actors since this enables estimation of variables at the level of forest stands. Furthermore, the operational organizations have changed their forest management systems with increased level of automation and decreased demand of labor. Their geographic information systems can now utilize continuously updating information from space borne sensors. All these developments are changing the landscape in operational forest management planning.

Only a few studies exist where the performance of Landsat 8 and Sentinel-2 data has been addressed for the purpose of quantitative monitoring of the environment so far. Mandanici and Bitelli (2016) studied the correlation of Sentinel-2 and Landsat 8 data in six test areas around the world. They reported good correlation of the near-equivalent instrument channels enabling the combined use of the two instruments, as long as the differences in the radiometric characteristics of the two sensors are taken into account in the target application. Chrysafis et al. (2017) examined the relationships between growing stock volume (GSV) and Sentinel-2 Multispectral Instrument and Landsat 8 Operational Land Imager (OLI) imagery. Using random forest (RF) regression models, they found Sentinel-2 to be slightly better for GSV estimation. They also found the Sentinel-2 red-edge band B5 to be significant in terms of correlation with forest parameter estimation. Other studies for Sentinel-2 have been published for the retrieval of forest canopy cover (CC), effective canopy cover (ECC), and leaf area index (LAI) (Korhonen et al., 2017) for land cover/land use classification (Topaloğlu et al., 2016; Forkuor et al., 2017) and for crop and tree species classification (Immitzer et al., 2016). Most of the studies comparing Sentinel-2 and Landsat 8 reported equal or better performance of Sentinel-2-based models. The usefulness of Sentinel-2 red-edge band 1 was also highlighted.

The work for this study was a part of a research project jointly funded by the Finnish Ministry of Agriculture and Forestry and VTT Technical Research Centre of Finland (project OH300-S42100-03). The motivation of the project was to utilize new data sources and methods for future development of forest related products and services for forest owner's and forest industry needs. The objective of the project was to apply new sources of both remotely sensed and field measured data (satellite data and harvester data) to forest variable prediction in boreal forest (Finland) and to assess their added value for prediction performance. The objective of this study was to compare the potential of the two main medium resolution data sources, Landsat 8 and Sentinel-2, for forest monitoring to understand their potential to provide information for forest management. We included the main variables needed in forest management and in fine-scale primary productivity estimation in our comparison. The detailed objectives were

1) To compare the forest variable prediction performance of Landsat 8 and Sentinel-2 using all spectral bands of the instruments and near-equivalent bands
2) To examine the significance of the better spatial resolution of Sentinel-2
3) To examine the significance of Sentinel-2 red-edge bands (RE1, RE2 & RE3) on prediction performance

The predicted forest variables were stem volume (V), stem diameter (D), tree height (H) and basal area (G). The species-wise components of these variables for pine (Pine), spruce (Spr) and broadleaved (BL) trees were also included. Two modelling methods, a multi-layer perceptron (MLP) neural network (Rumelhart and McClelland, 1986) and

[1] http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Overview3.

[2] http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-2/Introducing_Sentinel-2.

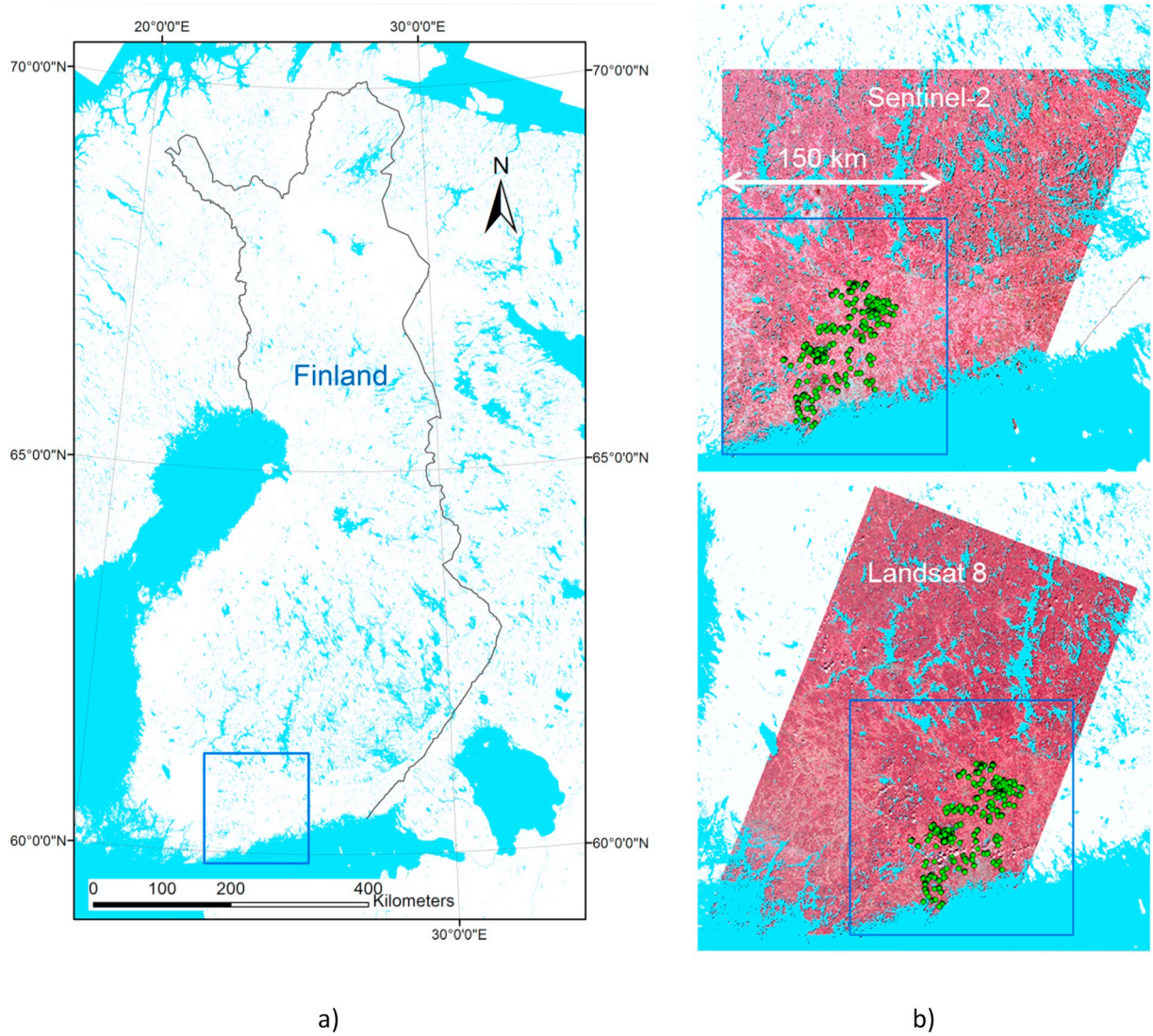


Fig. 1. a) The test site location in Southern Finland, b) The Sentinel-2 (top) and Landsat 8 images used in the study shown as false color images (R = NIR, G = red, B = green). The blue line shows the delineation of the test site image window, the reference data locations shown with green dots. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

regression tree (Breiman, 1998) were used for forest variable prediction. Their results were compared to verify that the results were not method-dependent. A set of 739 plots of field reference data were used to train, validate and test the performance of the prediction models. The test results were produced using the 10 m and 20 m resolution bands of Sentinel-2 MSI (S2) and the bands 1–7 of Landsat 8 OLI (L8) imagery.

2. Materials

2.1. Test site

The study area was located in Southern Finland, centered at 60°35′07.2″N, 24°44′25.9″E and covered roughly an area of 100 km × 100 km (Fig. 1). The test site is a typical Finnish managed boreal forest landscape with some agricultural areas and small towns and villages. The topography is relatively flat with only small changes in elevation. The forest area is conifer-dominated by Norway spruce (*Picea abies* (Karst.)) and Scots pine (*Pinus sylvestris* (L.)) as the major species. The deciduous trees usually occur as mixed species, mainly birches (*Betula* spp.). The regional proportions of the different species are 48%, 28% and 23% of the mean growing stock volume, respectively (Ylitalo, 2012). Classified by stand site class, the area consists of mainly fertile forest types (Cajander, 1949).

2.2. Reference data

The test area was covered with forest field sample plots that had been measured by the local Finnish Forest Center in the field. The measured variables included stem volume, diameter at breast height, tree height, basal area, age, stem number, and additional stand information like development class, dominant species, proportion of sawn timber, regeneration situation, etc. These reference data had been collected during the period from May 6th 2015 to Sep 23rd 2015. Three different plot radii had been used: The plot radius was 9 m in young and advanced managed forests with a relatively high tree density, and 12.62 m in forest with a low stem density but usually high volume due to the mature development stage. In seedling stands the radius was 5.64 m. Table 1 shows the mean and standard deviation of the main variables for all plots, and for the field plot groups with different dominating species.

2.3. Satellite data

Landsat 8 surface reflectance Level-2 data product of an image scene

Table 1
Mean and standard deviation of main forest variables for all plots, and for field plots grouped according to dominating species.

Forest variable	All plots		Pine		Spruce		Broadleaved	
	Mean	Stdev	Mean	Stdev	Mean	Stdev	Mean	Stdev
Mean stem diameter (D [cm])	16.6	10.1	17.9	8.2	18.1	11.3	11.3	7.9
Basal area (G [m^2/ha])	18.7	12.0	19.2	9.4	20.8	13.7	13.4	10.2
Mean tree height (H [m])	14.3	7.4	14.4	5.7	15.2	8.5	11.9	6.7
Stem volume (V [m^3/ha])	159.8	140.0	150.0	97.1	193.6	167.1	101.1	110.0
Number of stems (N [stems/ha])	2268	3082	1691	1962	2074	2886	3661	4370
Age (T [a])	43.4	32.7	49.4	35.9	47.4	32.5	24.5	17.2
Nbr of plots:	739		252		336		151	

Table 2
Definition of the different test setups.

Test setup	Test setup description	Feature set(s) (see Table 3)	Sentinel-2 pixel resolution	Modelling method
S2_All	S2 data with all instrument bands	A	10/20 m	MLP
L8_All	L8 data with all instrument bands	A	30 m	MLP
S2_Eqv	S2 data with near-equivalent bands	B	10/20 m	MLP
L8_Eqv	L8 data with near-equivalent bands	B	30 m	MLP
S2_wo_RE	S2 data without the red-edge bands	C	10/20 m	MLP
S2_10	S2 data with 10 m bands only	D	10 m	MLP
S2_20	S2 data with 20 m bands only	E	20 m	MLP
S2_30_All	S2 data resampled to 30 m, all bands	A	30 m	MLP
S2_30_Eqv	S2 data resampled to 30 m, near-equivalent bands	B	30 m	MLP
S2_L8_All	All S2 and L8 bands combined	A	10/20/30 m	MLP
S2_All_rTree	S2 data with regression tree method	A	10/20 m	Regression tree
L8_All_rTree	S2 data with regression tree method	A	30 m	Regression tree

acquired on 20.8.2015 was ordered and downloaded from the USGS Earth Explorer data portal (https://earthexplorer.usgs.gov/). Sentinel-2 Top-Of-Atmosphere (TOA) Level-1C orthoimage product for a scene acquired on 17.8.2015 was downloaded from the Copernicus Open Access Hub https://scihub.copernicus.eu/dhus/#/home. The Sentinel-2 Bottom-Of-Atmosphere (BOA) or surface reflectance image was computed using the Dense Dark Vegetation (DDV) algorithm (Kaufman and Sendra, 1988). A cloud mask was manually drawn for the L8 image according to visual inspection. The S2 image was totally cloud-free. In addition to the original pixel resolution image, the S2 image was down-sampled to 30 m resolution by nearest neighbor sampling, in order to match the spatial resolution of L8 data. The input data set for the analysis included the spectral signatures (BOA) extracted from the pixels located at the reference data plot coordinates. Only target pixel data was recorded, i.e. no spatial averaging of image data was performed at this point. Fig. 1 shows the test site location and S2 and L8 satellite imagery with the extents of the study area and field sample plot data overlaid.

3. Methods

3.1. Test setups

Twelve test setups (Table 2) were defined to investigate the factors causing the presumed differences in the forest variable prediction performance between Landsat 8 and Sentinel-2. They included the L8 data, the S2 data in original pixel resolution and in down-sampled 30 m resolution, and two modelling methods, MLP and regression tree. The instrument bands were grouped into different channel combinations, feature sets, to investigate the significance of the S2 red-edge bands and pixel resolution for prediction accuracy.

- Instrument channel combinations

Two different instrument channel combinations were defined for L8 (feature sets A and B) and four for S2 (feature sets A–E) for testing. The feature set A included all the available instrument channels,[3] and feature set B the near-equivalent wavelength bands from both instruments. As in Korhonen et al. (2017) the S2 band B8a_nNIR instead of B08_NIR was used in feature set B as it corresponds better to L8 band 5 characteristics in terms of bandwidth and pixel size. Three more feature sets were defined for the S2 data. Feature set C included all S2 bands except the red-edge bands (B05_RE1, B06_RE2 & B07_RE3), feature set D included the S2 bands with 10 m pixel resolution and feature set E the S2 bands with 20 m pixel resolution. For the summary of the feature sets, see Table 3.

- Sentinel-2 data re-sampled to 30 m resolution

The modelling was repeated using S2 data resampled to 30 m in order to evaluate the significance of S2 better spatial resolution with respect to L8. The S2 30 m resolution was tested with two band combinations: including all S2 image bands (feature set A) and bands near-equivalent with L8 (feature set B).

- Sentinel-2 data with either 10 m or 20 m bands only

The feature sets D and E were defined to test S2 model performance using either only the 10 m resolution, or the 20 m resolution image bands to investigate the significance of pixel resolution to prediction accuracy.

- Sentinel-2 and Landsat 8 data combined

To compare the significance of S2 and L8 image bands as forest variable model predictors, all bands (feature set A) from the two satellites were combined into a 17-element input vector for the modelling process.

[3] All available instrument channels meaning the 10 m & 20 m bands of Sentinel-2 MSI, and bands 1–7 of Landsat 8 OLI.

Table 3
Spectral bands of Sentinel-2 MSI and Landsat 8 OLI instruments and the definition of feature sets.

Sentinel-2 MSI

Band	Description	Wavelengths (nm)	Resolution (m)	Feature set				
				A	B	C	D	E
1	Coastal aerosol	433–453	60	–	–	–	–	–
2	Blue	458–523	10	A	B	C	D	
3	Green	543–578	10	A	B	C	D	
4	Red	650–680	10	A	B	C	D	
5	Vegetation Red Edge (RE1)	698–713	20	A				E
6	Vegetation Red Edge (RE2)	733–748	20	A				E
7	Vegetation Red Edge (RE3)	773–793	20	A				E
8	Near-Infrared (NIR)	785–900	10	A		C	D	
8a	Narrow NIR (nNir)	855–875	20	A	B	C		E
9	Water vapor	935–955	60	–	–	–	–	–
10	Shortwave infrared - Cirrus	1360–1390	60	–	–	–	–	–
11	Shortwave infrared (SWIR1)	1565–1655	20	A	B	C		E
12	Shortwave infrared (SWIR2)	2100–2280	20	A	B	C		E

Landsat 8 OLI

Band	Description	Wavelengths (nm)	Resolution (m)	Feature set	
				A	B
1	Violet-deep Blue (V-D Blue)	433–453	30	A	–
2	Blue	450–515	30	A	B
3	Green	525–600	30	A	B
4	Red	630–680	30	A	B
5	Near-Infrared (NIR)	845–885	30	A	B
8	Pan-Chromatic	500–680	30	–	–
9	SWIR - Cirrus	1360–1390	30	–	–
6	Shortwave infrared (SWIR1)	1560–1660	30	A	B
7	Shortwave infrared (SWIR2)	2100–2300	30	A	B

- Two modelling methods

The two different modelling methods, MLP neural network and regression tree, were used to verify that the results were not method-dependent. For this it was considered adequate to repeat the tests with feature set A only.

3.2. Forest variable modelling and feature selection

Artificial neural networks are a set of different methodologies in mathematical problem solving. Typical uses for neural networks are different pattern recognition tasks, or modelling of the complex mappings between inputs and outputs of some, usually non-linear, system (Rumelhart and McClelland, 1986). Neural networks have proved to be efficient in finding complex, often non-linear dependencies between input and output data spaces. The major drawback of neural networks is that they do not provide an explicit formulation of the model, unlike physical-based models. The feed-forward MLP neural networks are known for their ability to interpolate values in target variable space and filter out noise to some extent (Helliwell et al., 1995). They are also well-suited for predicting continuous variables and are simple to implement.

During the last decade, the development of deep learning algorithms along with the increase in computing performance have brought neural networks again into the state-of-the art machine learning methodologies (Hinton et al., 2006; Vincent and Larochelle, 2010). However, for the comparison task of this study, the usage of deep learning methods were not supposed to perform better with the foreseen test setup (this was also verified with rapid tests) and thus a MLP network with one hidden layer was selected as the principal method for the models.

Regression trees, together with classification trees, are nonlinear predictive models commonly called decision trees (Breiman, 1998; Rokach and Maimon, 2007). The advantages of decision trees are, e.g., that they are simple to interpret and understand, they are capable of handling both categorical and continuous-valued data and they scale well with large data sets. On the other hand, decision trees may not be as accurate as other methods and they are prone to overfitting (James et al., 2000).

An iteration wrapper (Kohavi and John, 1997) program was written with Matlab for the modelling process, and applied for all the defined test setups. A flow chart of the iterative process is shown in Fig. 2. The input data set was split to training, validation and test sets by stratified random sampling in the beginning of the modelling process. Total stem volume was used as a stratification variable with five strata. The wrapper program included a brute force forward selection scheme (Blum and Langley, 1997; Reif and Shafait, 2014) to select the best predictive features for the models. The method searched first the best one-feature model for the target variable, and then added the next best feature on the next feature selection loop. The model performance was evaluated for all feature combinations of the set of already selected features and the available unselected features. In this study, the maximum number of input features was seven for L8, and ten for S2, numbers considered low enough to perform the exhaustive feature search. The selection criterion S_j for the best feature set was calculated for all the model input features j = i, …,N as:

$$S_j = \left(rmse\%_V + 10^{*}abs\left(bias\%_V\right)\right) \Big/ \left(0.5 + R_V^2\right) \quad (1)$$

$$c = argmin_j\{S_j\},\ j = i,\ \ldots,N$$

where

$argmin_j$ = argument of the minimum: $argmin_j\{S_j\} = \{j \mid \wedge j,\ l \in [i,N],$

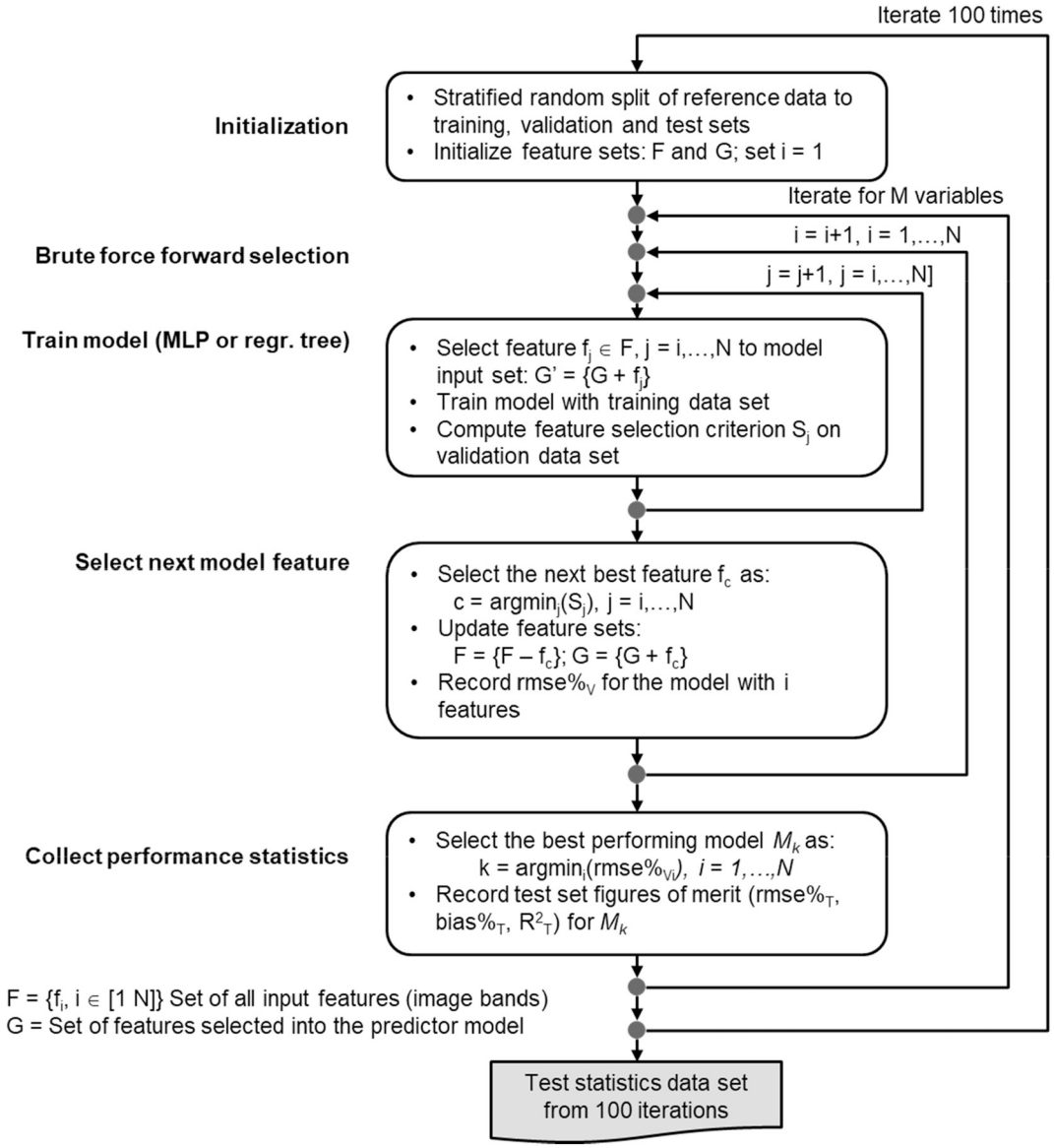


Fig. 2. Overall flow chart of the modelling program.

$j \neq l : S_l \geq S_j.\}$
$RMSE\%_V$ = relative RMS-error on validation set
$abs(BIAS\%_V)$ = absolute value of validation set relative BIAS (or systematic error)
R^2_V = validation set coefficient of determination
i = loop counter $i = 1, \ldots, N$ for the brute force forward feature selection loop
j = loop counter for the still unselected features after feature selection loop i ($j = i, \ldots, N$)
c = index of the best feature on iteration i of the feature selection loop

The feature with minimum $\boldsymbol{S_j}$ was always selected as the next feature of the model (feature set G) in the end of each feature selection loop. The objective of the criterion $\boldsymbol{S_j}$ was that the selected input features would simultaneously minimize both the relative root-mean-squared error ($RMSE\%_V$) and the absolute relative systematic error ($BIAS\%_V$) and maximize the coefficient of determination (R^2_V) for the validation set.

After selecting the next feature, the relative RMS error for the validation data set ($RMSE\%_V$) was recorded for the obtained model having i input features. The model M_k that produced the minimum $RMSE\%_{V\,i}$ ($i = 1, \ldots, N$) was selected as the best performing model after the completion of the feature selection.

The performance measures for the test set were then computed with the obtained model in the end of each run of the wrapper program. The accuracy measures included the coefficient of determination R^2, absolute and relative root mean square error (RMSE and RMSE%) and absolute and relative bias (BIAS and BIAS%):

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{2}$$

$$RMSE = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{n}}, \quad RMSE\% = \frac{RMSE}{\bar{y}} \times 100 \tag{3}$$

$$BIAS = \frac{\sum_i (y_i - \hat{y}_i)}{n}, \quad BIAS\% = \frac{BIAS}{\bar{y}} \times 100 \tag{4}$$

With y_i as the observed values from forest inventory and $\bar{y}$ their mean. $\hat{y}_i$ were the predicted values and n was the total number of field plots. All sums are for $i = 1, \ldots, n$.

The wrapper was repeated 100 times for each separate test setup.

The arithmetic mean values of the performance measures and their lower and upper confidence levels were computed from the results of these 100 iterations. Also, the order in which the input features (image channels) were selected to the model in the forward selection scheme and the number of features selected to the best performing model were recorded. The band selection order data from the 100 iterations were used to compute statistics on the most important S2 and L8 bands.

3.3. Implementation of the modelling methods

An individual model was trained for each of the forest variables using the Matlab 'trainlm' function of the Neural Network Toolbox for the neural network model and 'fitrtree' function of the Statistics and Machine Learning Toolbox for the regression tree model. The multilayer perceptron (MLP) neural network model contained one hidden layer with sigmoidal transfer functions, and one linear output node. The network output was recalculated to cover only positive values. The number of hidden layer neurons was set to $n_f + 1$ (n_f = number of MLP input features). The 'trainlm' network training function updated the network weights and bias values according to Levenberg-Marquardt optimization (Gill and Murray, 1978).

The regression tree models were created using the 'fitrtree' function with default parameters. The function output is a binary tree where each branching node is split based on the values of a column of an input feature table, i.e. on the values of multispectral channel data. The number of tree branches varied roughly between 35 and 50 during the feature selection iterations reflecting the number of input features presented to the function.

4. Results

4.1. Best predictive image bands

The order in which the spectral bands were selected as model input features in the forward selection modelling reflected the importance of the band in the prediction of a certain forest variable. The results from the feature selection indicated clear differences between the spectral bands with the models for variable totals including all species. However, with the models for species-wise variables there was large variation in band order both between different variables and species, and consequently common conclusions for those models could not be made. The results reported here consider thus only the models for variable totals (D, G, H and V).

The first S2 red edge band (B05-RE1) was clearly the best predictor for all forest variables as shown in Table 4. Its performance was particularly pronounced for the growing stock volume V. The green band (B03-Green) was the best predictor of the visible bands for three of the variables: it was the second-best band for stem basal area G and growing stock volume V, and seventh for tree height H. The visible bands were not among the best six predictors for stem diameter D. One or both short-wave infrared bands (B11-SWIR1, B12-SWIR2) were included in the best-performing models (Table 4, greyed cells) for all variables. The near-infrared bands (B8a-nNIR or B08_NIR) were among the best predictors for stem diameter and tree height, but did not rank high for stem basal area and stem volume models. The 10-meter resolution of Sentinel-2 did not seem to provide a special advantage to the 20-meter bands but rather vice versa, which is not surprising since the average field plot size was closer to 20 m.

The best predictor for L8 models was the green band (B3-Green) that was ranked first for stem basal area, tree height and growing stock volume and second for stem diameter (Table 4). The near-infrared band (B5-NIR) was the best predictor for stem diameter, and second best for tree height, but was ranked last with stem basal area and stem volume. The blue bands (B1-VD Blue & B2-Blue) were generally ranked high (second or third). The median number of bands in the best models was 4 or 6 with S2 data and 4 or 5 with L8 data.

Table 5 with near-equivalent S2 and L8 bands shows very good consistency among the stem basal area, tree height and growing stock volume models for both satellite data with the green band as the best predictor, and as second and third best for stem diameter. The near-infrared (B8a-nNIR) band was ranked generally high in the S2 models, except for stem basal area. The blue band (B2-Blue) was the second best predictor for all the variables for L8 satellite data.

The results from the test with the combined S2 and L8 image bands (Table 6) confirm the strength of S2 data as model predictors. The four best predictors for all variables were S2 bands, the red-edge band 1 (S2_B05_RE1) being the best for all variables but stem diameter, for which it was third. There were only 1 or 2 L8 bands in the best performing models on average, while the number of predictors in them varied from 6 to 10.

4.2. Prediction performance

Examples of the progress of performance figures of MLP modelling for two forest variables, total stem volume (V) and tree height (H), are shown in Figs. 3 and 4 respectively. The graph plots show the median values of performance measures RMSE%, BIAS% and R^2 for the training set, validation set and test set from 100 iterations using Sentinel-2 data (a), and Landsat 8 data (b). The top graphs show the typical overall decreasing development of the training set RMSE%. The minimum locations of the validation set and test set RMSE% indicate the point where overfitting to the training set starts to occur. The figures also show the ranking of input image bands from the modelling process. Scatterplots of test data set prediction for one modelling iteration are shown in Figs. 5 and 6 for the four total variables including all species (D, G, H and V).

The averages of test set performance measures relative RMS-value, relative BIAS and the coefficient of determination and their 99% lower and upper confidence intervals from 100 MLP modelling iterations, are shown in Table 7 (RMSE%, BIAS%) and Table 8 (R^2) for the four forest variables and their species-wise components. All image bands (feature set A) were used for both S2 and L8 for the data in these two tables (test setups S2_All and L8_All). The mean RMSE% and R^2 values for the variable totals (D, G, H and V) in all test setups and are shown in Fig. 7. For a complete set of estimated variables the RMSE% and R^2 values are shown in Appendix A, Tables A1 and A2, respectively.

4.2.1. Models using all spectral bands

When all the spectral bands of both S2 and L8 were included in the modelling (test setups S2_All and L8_All), the test set RMS errors of S2 predictions were 6.3–13.0% lower than with L8 data when the total variable (D, G, H and V) predictions were considered. The RMSE values for the species-wise variables for S2 were 4.3–23.5% lower than for L8. The BIAS values were very low for both S2 and L8 predictions. The average of the BIAS% absolute values over all the forest variables was 0.8% and 1.2% for S2 and L8, respectively (Table 7). The coefficient of determination (R^2) for S2 predictions varied between 0.56 and 0.65 for total variables including all species and 0.23 and 0.57 for species-wise variables, as for L8 the corresponding figures were 0.38–0.51 and 0.17–0.38 (Table 8).

4.2.2. Comparison of the different test setups

Figs. 7 and 8 show the RMSE% and R^2 results for the variable totals from the defined test setups. The bar groups S2_All and L8_All are replicas of the results shown already in Tables 7 and 8. The bar group S2_30_All shows the average performance of the models produced with S2 data with all bands (feature set A) that was down-sampled to 30 m pixel resolution. The RMSE values increased on average 2.2% and 2.7% for total variables and species-wise variables, respectively, when compared to the S2 results with original pixel resolution (test setups S2_All vs. S2_30_All).

The down-sampled S2 data was compared to L8 data using the near-

Table 4
The number of cases of 100 iterations in which a band was selected as input to the best model sorted in decreasing order. MLP model using feature set A. The greyed cells indicate the median number (of 100 iterations) of bands that were selected into the best-performing model.

Sentinel-2/feature set A								
No	D	No	G	No	H	No	V	No
	Band	cases	Band	cases	Band	cases	Band	cases
1	B05_RE1	84	B05_RE1	87	B05_RE1	89	B05_RE1	88
2	B08_NIR	75	B03_Green	65	B06_RE2	68	B03_Green	62
3	B12_SWIR2	66	B12_SWIR2	45	B11_SWIR1	66	B06_RE2	49
4	B06_RE2	60	B02_Blue	43	B08_NIR	63	B11_SWIR1	48
5	B07_RE3	58	B07_RE3	40	B8a_nNIR	62	B08_NIR	43
6	B8a_nNIR	56	B06_RE2	38	B12_SWIR2	59	B8a_nNIR	43
7	B02_Blue	53	B04_Red	37	B03_Green	58	B12_SWIR2	41
8	B04_Red	53	B11_SWIR1	36	B04_Red	48	B04_Red	39
9	B03_Green	50	B8a_nNIR	35	B07_RE3	48	B07_RE3	38
10	B11_SWIR1	48	B08_NIR	33	B02_Blue	43	B02_Blue	33

Landsat 8/feature set A								
No	D	No	G	No	H	No	V	No
	Band	cases	Band	cases	Band	cases	Band	cases
1	B5-NIR	94	B3-Green	91	B3-Green	86	B3-Green	88
2	B3-Green	81	B1-VD Blue	73	B5-NIR	79	B2-Blue	73
3	B1-VD Blue	70	B6-SWIR1	68	B2-Blue	78	B4-Red	63
4	B4-Red	67	B4-Red	67	B4-Red	67	B7-SWIR2	58
5	B6-SWIR1	63	B7-SWIR2	64	B1-VD Blue	62	B1-VD Blue	57
6	B2-Blue	60	B2-Blue	61	B6-SWIR1	56	B6-SWIR1	56
7	B7-SWIR2	56	B5-NIR	61	B7-SWIR2	56	B5-NIR	52

equivalent bands of S2 and L8 (feature set B). The differences were smaller than with S2 original resolution data but still clear for the benefit of S2: The RMSE of L8-based predictions was 3.8% and 5.8% higher on average than the S2 predictions for total and species-wise variables, respectively (setup L8_Eqv vs. S2_30_Eqv in Fig. 7). The R^2 values of S2 for this test case were 0.52, 0.50, 0.54 and 0.49 for stem diameter, stem basal area, tree height and growing stock volume, respectively.

The S2 total variable models without the red-edge bands (feature set C) performed on average 1.6% worse than the S2 models using all image bands (setup S2_All vs. S2_wo_RE in Fig. 7). With species-wise variables there was no clear difference between the test setups. The MLP models with only the S2 10 m resolution bands (feature set D) produced 2.2% and 1.4% higher RMSE% values than the models with all image bands for total and species-wise variables, respectively (setup S2_All vs. S2_10). The corresponding figures for the models with only 20 m resolution bands (feature set E) were 0.4% and 2.3%, i.e. with the total forest variables there was no reduction in performance (setup S2_All vs. S2_20).

There was no added value of the L8 band usage to the prediction performance in terms of RMSE when using combined S2 and L8 bands for modelling (setup S2_All vs. S2_L8_All in Fig. 7). The four or more best predictive features were S2 bands (Table 6).

The regression tree models did not perform as well as the multilayer perceptron models (test setups S2_All & L8_All vs. setups S2_All_rTree & L8_All_rTree, respectively, in Fig. 7). The RMSE values for the

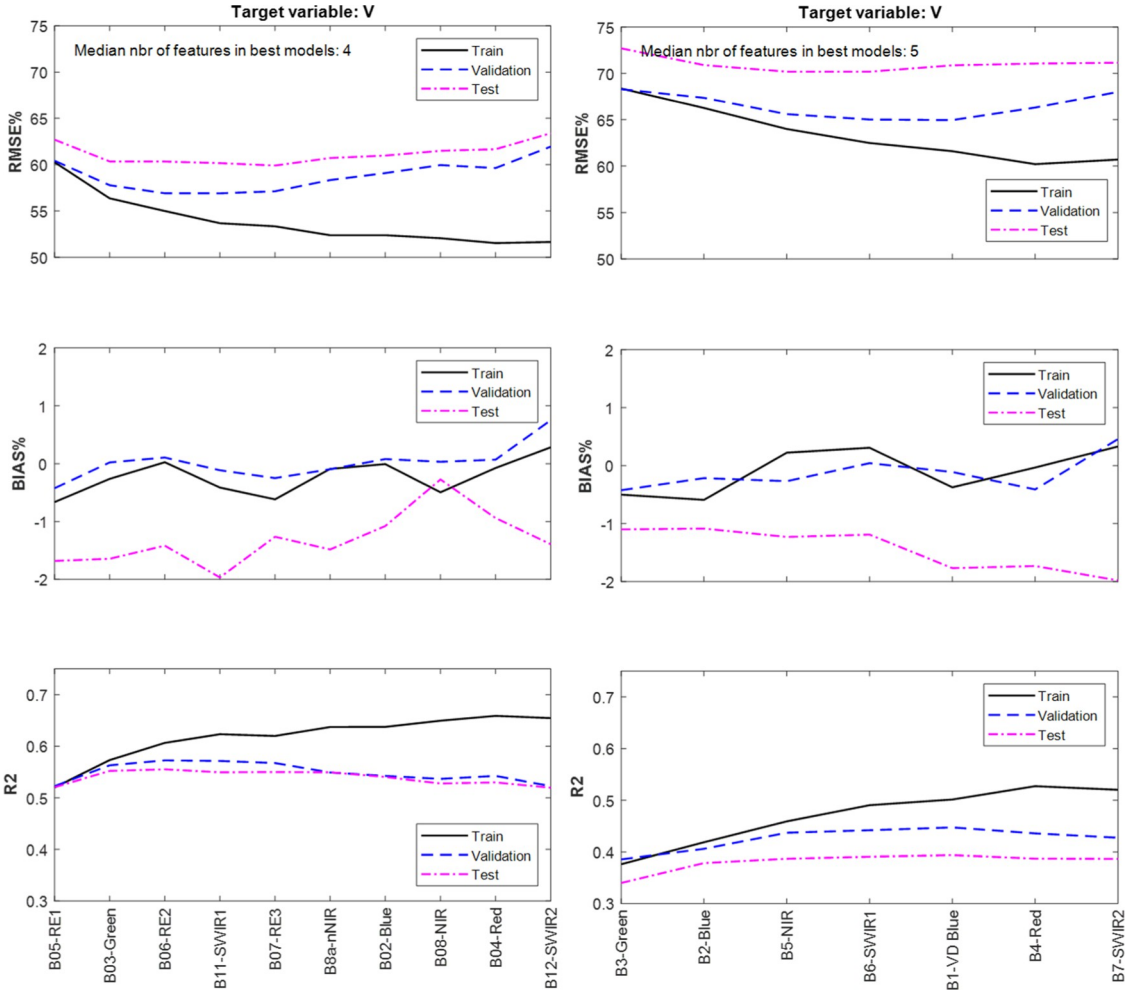
Table 5
The number of cases of 100 iterations in which a band was selected as input to the best model sorted in decreasing order. MLP model using feature set B. The greyed cells indicate the median number (of 100 iterations) of bands that were selected into the best-performing model.

Sentinel-2/feature set B								
No	D	No	G	No	H	No	V	No
	Band	cases	Band	cases	Band	cases	Band	cases
1	B8a_nNIR	97	B03_Green	96	B03_Green	98	B03_Green	88
2	B03_Green	82	B12_SWIR2	68	B8a_nNIR	86	B8a_nNIR	84
3	B04_Red	60	B11_SWIR1	63	B04_Red	70	B04_Red	65
4	B11_SWIR1	56	B02_Blue	57	B02_Blue	60	B11_SWIR1	63
5	B12_SWIR2	55	B04_Red	54	B11_SWIR1	59	B12_SWIR2	60
6	B02_Blue	54	B8a_nNIR	51	B12_SWIR2	57	B02_Blue	46

Landsat 8/feature set B								
No	D	No	G	No	H	No	V	No
	Band	cases	Band	cases	Band	cases	Band	cases
1	B5-NIR	98	B3-Green	90	B3-Green	90	B3-Green	90
2	B2-Blue	79	B2-Blue	73	B2-Blue	87	B2-Blue	84
3	B3-Green	78	B6-SWIR1	70	B5-NIR	81	B4-Red	69
4	B6-SWIR1	65	B4-Red	66	B6-SWIR1	70	B5-NIR	68
5	B4-Red	63	B5-NIR	61	B4-Red	68	B6-SWIR1	66
6	B7-SWIR2	54	B7-SWIR2	60	B7-SWIR2	59	B7-SWIR2	63

Table 6
The number of cases of 100 iterations in which a band was selected as input to the best model sorted in decreasing order. MLP model using both Sentinel-2 and Landsat 8 image bands of feature set A. The greyed cells indicate the median number (of 100 iterations) of bands that were selected into the best performing model.

No	D	No	G	No	H	No	V	No
	Band	cases	Band	cases	Band	cases	Band	cases
1	S2_B06_RE2	70	S2_B05_RE1	79	S2_B05_RE1	80	S2_B05_RE1	90
2	S2_B08_NIR	63	S2_B03_Green	64	S2_B06_RE2	67	S2_B03_Green	51
3	S2_B05_RE1	61	S2_B11_SWIR1	54	S2_B03_Green	65	S2_B06_RE2	49
4	S2_B03_Green	54	S2_B04_Red	44	S2_B11_SWIR1	65	S2_B04_Red	46
5	S2_B07_RE3	54	L8_B1_VD_Blue	44	S2_B08_NIR	62	L8_B4_Red	45
6	L8_B1_VD_Blue	54	S2_B06_RE2	42	S2_B07_RE3	61	S2_B02_Blue	43
7	L8_B3_Green	53	L8_B3_Green	39	S2_B8a_nNIR	59	S2_B08_NIR	40
8	S2_B8a_nNIR	51	S2_B07_RE3	38	S2_B12_SWIR2	59	L8_B2_Blue	39
9	L8_B4_Red	46	S2_B8a_nNIR	38	L8_B3_Green	59	L8_B3_Green	39
10	L8_B7_SWIR2	46	S2_B12_SWIR2	38	L8_B6_SWIR1	59	S2_B07_RE3	38
11	S2_B04_Red	45	L8_B4_Red	35	L8_B1_VD_Blue	58	S2_B12_SWIR2	38
12	S2_B12_SWIR2	44	S2_B02_Blue	34	S2_B04_Red	56	L8_B5_NIR	38
13	L8_B2_Blue	44	S2_B08_NIR	34	L8_B2_Blue	54	L8_B6_SWIR1	38
14	L8_B5_NIR	44	L8_B5_NIR	34	L8_B7_SWIR2	54	S2_B11_SWIR1	37
15	S2_B11_SWIR1	41	L8_B7_SWIR2	33	L8_B4_Red	50	L8_B1_VD_Blue	37
16	S2_B02_Blue	37	L8_B2_Blue	31	S2_B02_Blue	48	L8_B7_SWIR2	36
17	L8_B6_SWIR1	37	L8_B6_SWIR1	23	L8_B5_NIR	41	S2_B8a_nNIR	35



a) Total Stem Volume (V) - Sentinel-2 b) Total Stem Volume (V) - Landsat 8

Fig. 3. Examples of median values of relative RMS-error (RMSE%), relative BIAS (BIAS%) and coefficient of determination (R^2) for total stem volume (V) obtained from 100 MLP modelling iterations using a) Sentinel-2 and b) Landsat 8 data and feature set A. The horizontal axis of the bottom graph shows the ranking of the features from the 100 iterations, with the best predictors on the left. The median number of features in the 100 best performing models (providing the minimum validation set RMSE% error) has been indicated in the top graphs.

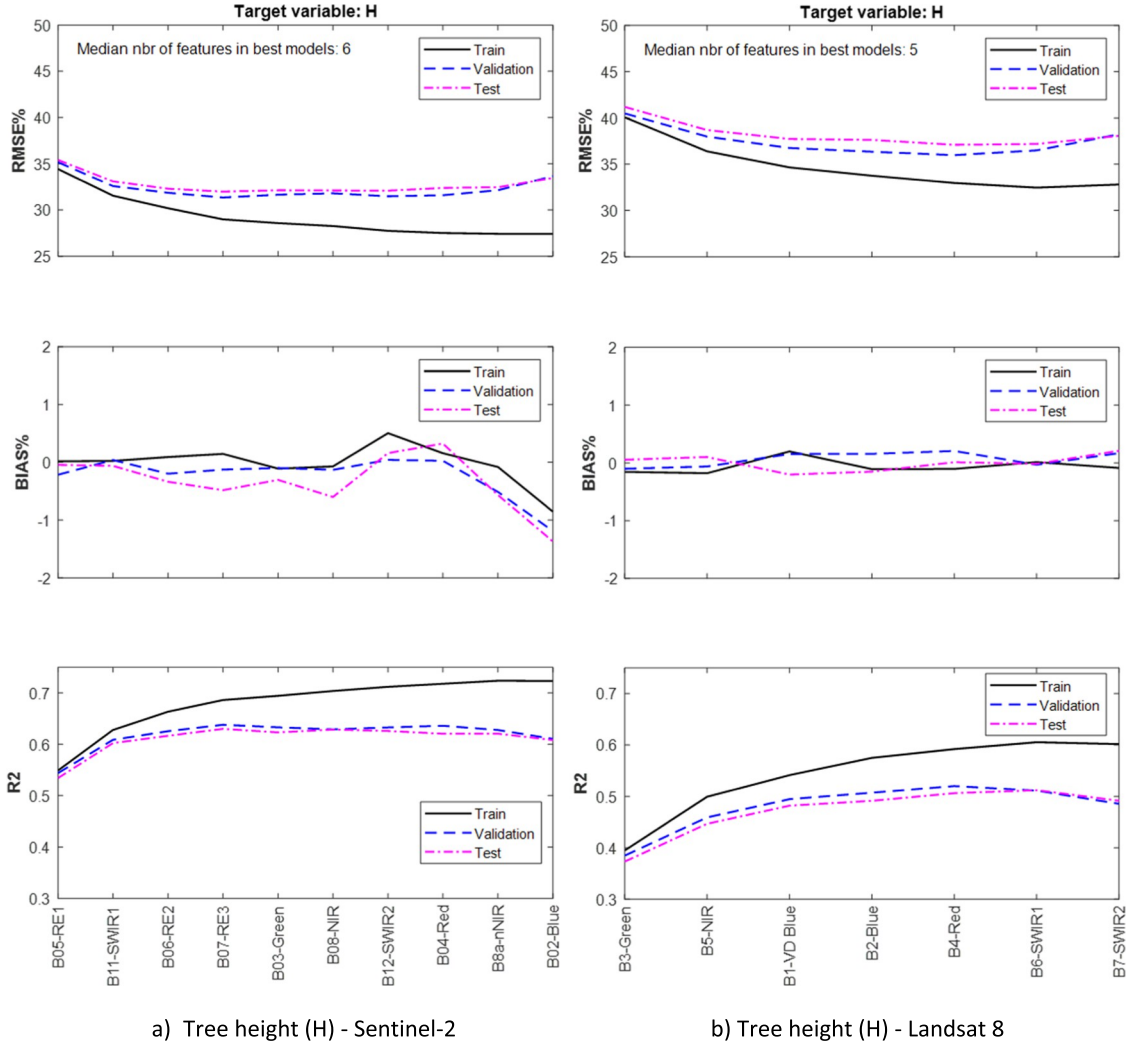


Fig. 4. Examples of median values of relative RMS-error (RMSE%), relative BIAS (BIAS%) and coefficient of determination (R^2) for tree height (H) obtained from 100 MLP modelling iterations using a) Sentinel-2 and b) Landsat 8 data and feature set A. The horizontal axis of the bottom graph shows the ranking of the features from the 100 iterations, with the best predictors on the left. The median number of features in the 100 best performing models (providing the minimum validation set RMSE% error) has been indicated in the top graphs.

regression tree predictions were on the average 15.9% or 12.3% higher for S2 and L8, respectively, than with the MLP models. The only exception to this was obtained with the L8 model for broadleaved trees, for which the regression tree model produced lower RMSE% values. However, the R^2 values for these predictions remained modest (≤ 0.27) (see Appendix A. Tables A1 & A2). The reason for this remained unclear. The performance differences between L8 and S2 are in line with the results from the tests with MLP models.

In the systematic error there was no significant difference between the two instruments. The BIAS% values were very low in all the tests for both instruments and all forest variable models.

In our tests there was no difference in the prediction performance between the models produced with the S2 narrow NIR (B8a_nNIR) band when compared with models with the NIR (B08_NIR) band.

5. Discussion

The obtained results show better predictive capability of the models using Sentinel-2 MSI data when compared to Landsat 8 OLI data. The explanatory factors for the better S2 performance are the additional image channels, especially the red-edge band 1, and the better pixel resolution. In our tests, down-sampling the original spatial resolution of S2 data had an approximately equal decreasing effect on the prediction performance as leaving out the red-edge bands. With near-equivalent bands and with reduced spatial resolution, the models based on S2 data performed better than the L8 models. One possible explanation for this might be that as the field plot data were measured with radii R = 5.6 m, R = 9 m and R = 12.6 m it was supposed to favor the smaller pixel resolution S2 data with ground pixel area matching better the field plot area. However, this could not be verified as the reference data set was not representative enough, and thus the factors for the remaining difference continue to be unclear. An interesting result was that the S2 performance for predicting the total forest variables was almost equal with the 20 m image bands only and with all S2 bands. This may be significant for mapping projects with limited critical processing power or data storage capacity.

The results from the feature selection reflect the high correlations between the forest variables and the Sentinel-2 B05-RE1, B03-Green and B11-SWIR1/B12-SWIR2 image bands, and the Landsat 8 B3-Green band (see Appendix B. Tables B1 & B2). An interesting result was that the Landsat 8 blue bands (B1-VD Blue, B2-Blue) were among the best predictors for all the forest variables.

Our results agree to large extent with the results from other studies. The predictive power of S2 red-edge bands, and especially red-edge 1, has been reported in recent studies of crop and tree species classification (Immitzer et al., 2016) and biophysical variable prediction

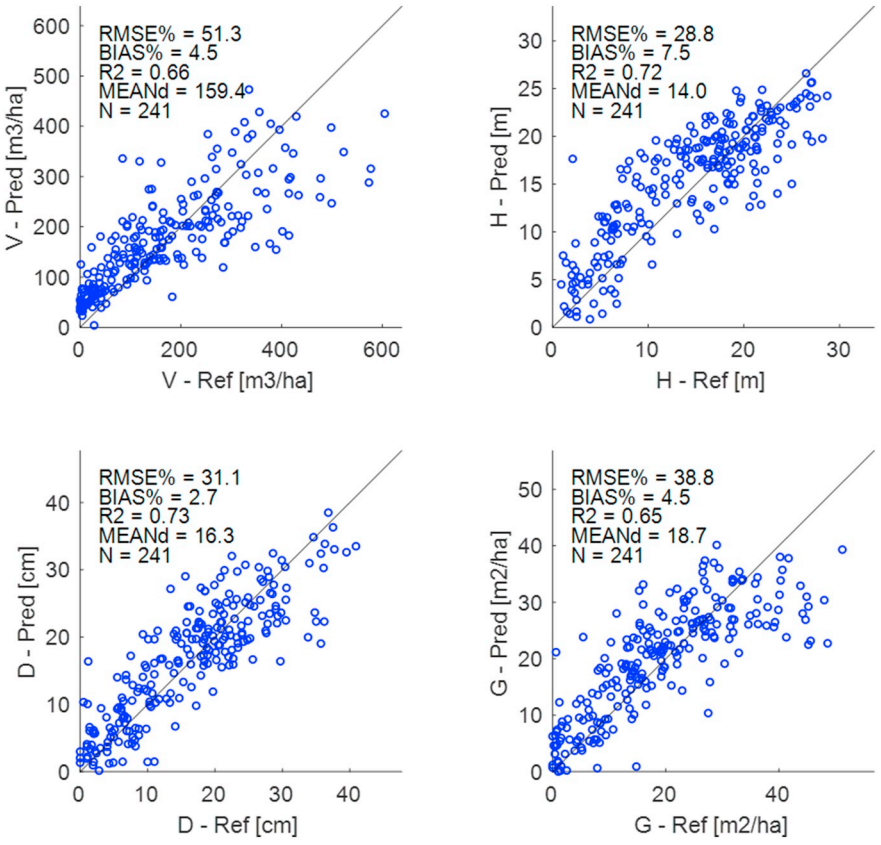


Fig. 5. Scatterplots of test data set predictions for Sentinel-2 example models. MLP prediction, feature set A.

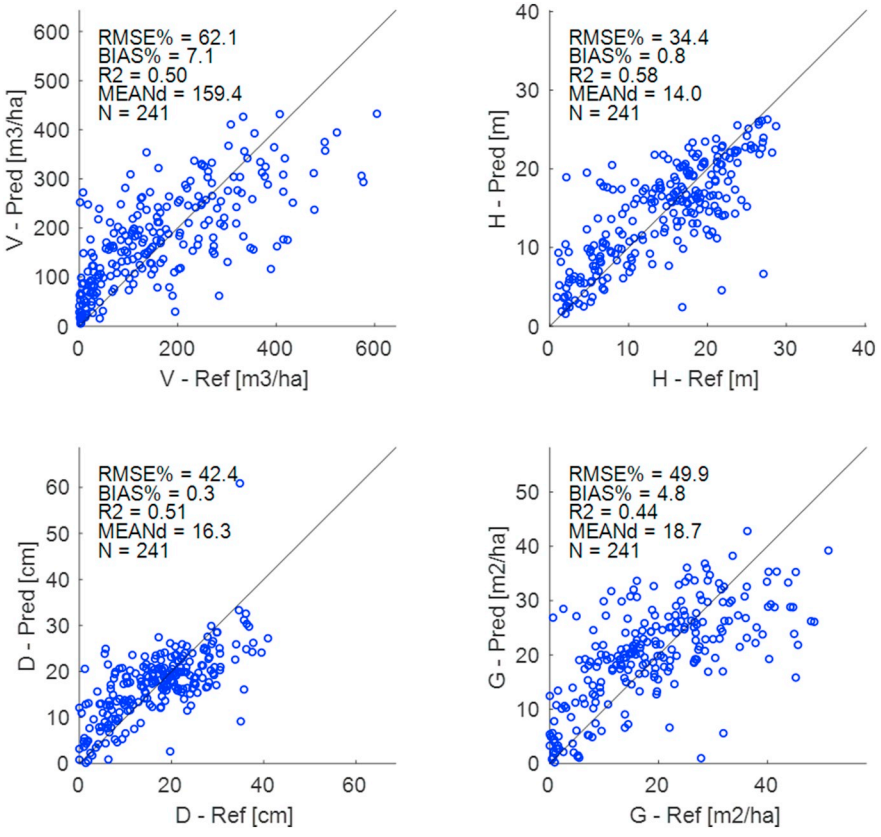


Fig. 6. Scatterplots of test data set predictions for Landsat 8 example models. MLP prediction, feature set A.

Table 7
Average relative test set RMS-error and relative BIAS for N = 100 iterations of the MLP prediction using feature set A with 99% confidence intervals (CI_L = confidence interval lower limit, CI_U = confidence interval upper limit) for total and species-wise variables (BL: broad-leaved, Spr: spruce). Test setups S2_All and L8_All.

Variable	RMSE%						BIAS%					
	Sentinel-2			Landsat 8			Sentinel-2			Landsat 8		
	Mean	CI_L	CI_U	Mean	CI_L	CI_U	Mean	CI_L	CI_U	Mean	CI_L	CI_U
D	38.4	37.6	39.1	44.6	43.8	45.5	−0.13	−1.01	0.76	−1.36	−2.49	−0.22
G	42.5	41.9	43.1	50.2	49.0	51.5	−0.94	−1.78	−0.10	−1.62	−2.60	−0.65
H	30.4	30.0	30.9	36.6	36.0	37.2	−0.35	−1.09	0.40	−0.81	−1.64	0.02
V	59.3	58.4	60.2	72.2	69.3	75.2	−1.42	−2.54	−0.29	−2.50	−3.75	−1.24
D_BL	85.5	84.3	86.7	89.8	88.4	91.1	1.02	−0.89	2.93	0.18	−1.90	2.27
D_Pine	101.4	99.7	103.1	108.2	105.2	111.2	−1.18	−3.39	1.02	1.53	−1.11	4.17
D_Spr	64.5	63.5	65.5	74.1	72.7	75.4	0.47	−1.14	2.07	0.67	−0.92	2.26
G_BL	119.8	116.4	123.2	136.3	125.3	147.3	0.71	−2.04	3.46	−1.73	−5.25	1.79
G_Pine	110.6	108.4	112.8	124.0	121.1	126.8	0.93	−1.47	3.32	−1.18	−4.30	1.94
G_Spr	92.8	91.4	94.2	111.7	109.3	114.1	−1.94	−3.85	−0.03	−0.61	−2.80	1.59
H_BL	69.0	68.1	70.0	73.7	72.5	75.0	0.64	−1.02	2.30	0.67	−1.04	2.38
H_Pine	93.6	92.0	95.1	99.5	97.9	101.0	−0.78	−2.80	1.25	1.14	−1.15	3.44
H_Spr	58.2	57.3	59.0	67.2	66.1	68.3	0.15	−1.21	1.51	1.00	−0.51	2.50
V_BL	160.0	154.7	165.3	191.1	142.6	239.6	−0.17	−4.48	4.14	0.25	−5.27	5.77
V_Pine	122.3	119.9	124.7	140.3	135.9	144.7	−0.59	−3.26	2.07	−1.42	−4.59	1.74
V_Spr	111.7	110.0	113.4	135.1	132.9	137.3	−1.04	−3.36	1.27	−1.97	−3.89	−0.04

Table 8
Average coefficient of determination (R^2) for N = 100 iterations of the MLP prediction using feature set A with 99% confidence intervals (CI_L = confidence interval lower limit, CI_U = confidence interval upper limit) for total and species-wise variables (BL: broad-leaved, Spr: spruce). Test setups S2_All and L8_All.

Variable	R2					
	Sentinel-2			Landsat 8		
	Mean	CI_L	CI_U	Mean	CI_L	CI_U
D	0.60	0.59	0.61	0.47	0.46	0.49
G	0.57	0.56	0.58	0.42	0.40	0.44
H	0.65	0.64	0.66	0.51	0.50	0.53
V	0.56	0.55	0.57	0.38	0.36	0.40
D_BL	0.23	0.22	0.25	0.17	0.15	0.18
D_Pine	0.29	0.28	0.31	0.22	0.20	0.24
D_Spr	0.52	0.51	0.53	0.38	0.37	0.40
G_BL	0.36	0.35	0.38	0.30	0.27	0.32
G_Pine	0.46	0.45	0.47	0.33	0.32	0.35
G_Spr	0.56	0.55	0.57	0.38	0.37	0.40
H_BL	0.29	0.28	0.31	0.21	0.19	0.22
H_Pine	0.32	0.31	0.34	0.24	0.23	0.25
H_Spr	0.54	0.53	0.55	0.40	0.39	0.42
V_BL	0.23	0.21	0.25	0.24	0.21	0.27
V_Pine	0.44	0.42	0.45	0.29	0.27	0.31
V_Spr	0.57	0.56	0.58	0.38	0.36	0.39

(Delegido et al., 2011; Korhonen et al., 2017). With a similar type of test arrangement using L8-equivalent bands from S2 re-sampled to 30 m resolution (Topaloğlu et al., 2016) a better classification accuracy was reported for S2 in land cover/land use mapping.

Chrysafis et al. (2017) reported a similar overall order of predictor importance for the growing stock volume (GSV) random forest regression model. The most interesting difference was that they found the short-wave infrared band 1 (SWIR1) to be the most important band for both S2 and L8 models, as the red-edge 1 (RE-1) and green bands obtained the highest ranks in our tests for S2 and L8 models, respectively. For the GSV prediction performance the authors reported only a slight difference between S2 and L8, while our results showed clearly better results for S2. The different biome with respect to our study may explain the differences in the results.

6. Conclusions

The models based on Sentinel-2 data outperformed Landsat 8 models for all forest variables in terms of root mean square error and coefficient of determination (R2). In the systematic error there was no significant difference between the two instruments. In general, the systematic error of the test data set was very low for all forest variable models and for both satellite data sets. The selected MLP modelling method resulted in low bias in most of the models. A somewhat surprising result was that using Landsat 8 image bands in combination with Sentinel-2 data did not improve the prediction accuracy with respect to the best model with Sentinel-2 bands only. Another interesting result was that the prediction accuracy of the models for variable totals was practically the same when using all image bands or only the 20 m resolution bands of Sentinel-2.

The factors explaining the better performance of S2 data with respect to L8 data were the additional red-edge bands and the better spatial resolution, which had a slightly stronger effect. However, our tests did not explain all the observed performance difference between S2- and L8-based predictions.

According to our tests, the best predictive S2 image band was the red-edge 1 (B05_RE1) for the estimation models of variable totals including all species. The short-wave infrared bands (B11_SWIR1 or B12_SWIR2) and the visible green band (B03_Green) were also among the best predictors. General conclusions for the predictor importance for the species-wise variable models could not be made. The median number of predictors in the best performing models was 4–6 for S2 models and 4–5 for L8 models.

In Finland and Scandinavia in general, forest management systems are very advanced and they receive plenty of public financial support. National forest inventories in Finland and Sweden use satellite images to augment the principal information from the field plots[4,5] (Tomppo et al., 2014; Fridman et al., 2014). The operational management planning has moved from using aerial photography and fieldwork to airborne laser scanning and digital airborne imaging. Free and timely Sentinel-2 data has raised growing interest within past two years. The

[4] http://kartta.luke.fi/index-en.html, accessed December 7, 2018.

[5] https://www.slu.se/en/Collaborative-Centres-and-Projects/the-swedish-national-forest-inventory/forest-statistics/slu-forest-map/, accessed December 7, 2018.

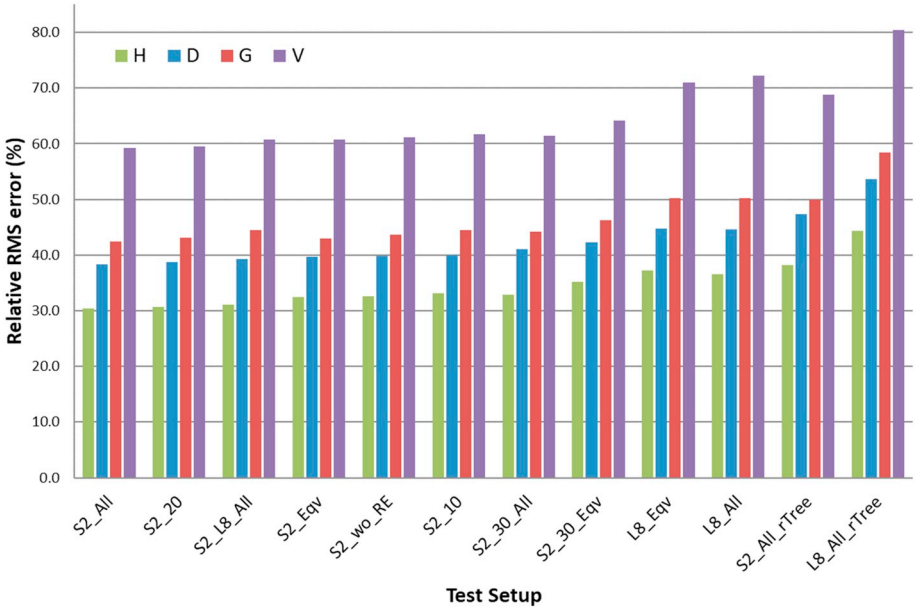


Fig. 7. The mean relative RMSE values from the different modelling test setups for four forest variables. Columns for test setups S2_All and L8_All repeat the mean RMSE% results from Table 7. Test setups ordered according to increasing sum of RMSE% of variables D, G, H & V.

ongoing or near-future operational applications of Sentinel-2 are on monitoring of forest cuttings and development and condition of the seedling stands.

The next phase in the utilization of Sentinel-2 data to support forest management will be to extend the services to growing forest stands. Airborne laser scanning can reach approximately to RMS-errors of 20% of the mean of the total volume at the level of a single sample plot (Kankare et al., 2013) whereas in this study the errors were close to 60%. A problem with optical and radar satellite data has been saturation of the estimate approximately at 250 m^3/ha or 150 tons/ha (Zhao et al., 2016). In this study, the saturation was less pronounced than what was observed earlier. Later studies will focus on investigating the accuracy of Sentinel-2 based estimation at forest stand level, i.e. for units of a couple of hectares. The stands are the basic units in the operational forest management. The errors are expected to be clearly smaller than what was reported here. Methods to combine less frequent and expensive but more accurate airborne laser scanner data with frequent and inexpensive Sentinel-2 based estimation will be investigated.

In the tropical forest with strong anthropogenic influence and modest biomass levels the performance of optical (and radar) satellite data has been similar to the boreal forest (Häme et al., 2013). We hypothesize that the difference of the performance between Sentinel-2 and Landsat 8 is also in the same order of magnitude in the tropics. Laser scanning is not a realistic option in most tropical areas but not in most boreal forests either, due to expenses or administrative reasons. Rainforest with closed canopy is a challenge for remote sensing. The

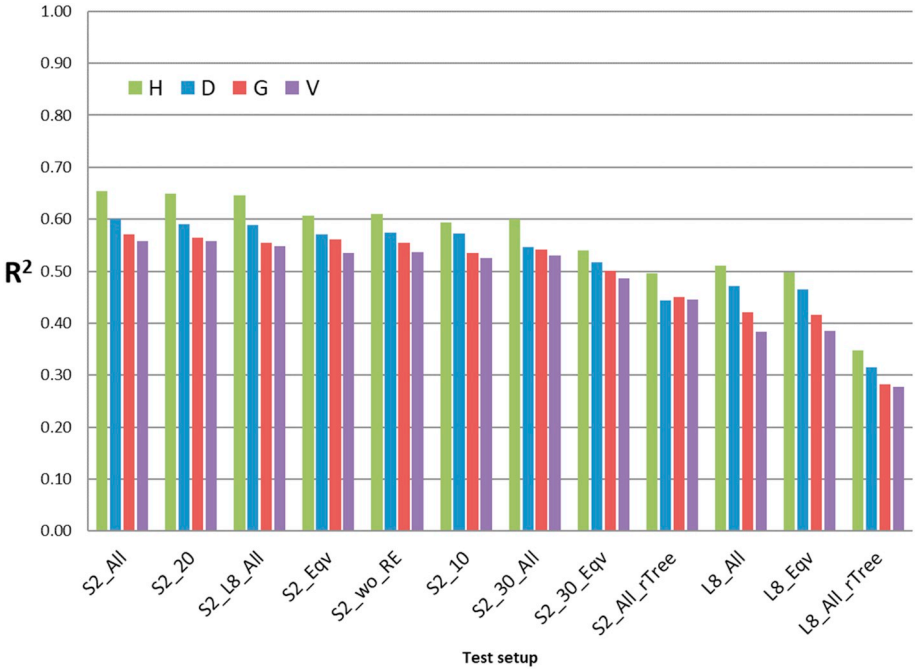


Fig. 8. The mean coefficient of determination R^2 values rom the different modelling test setups for four forest variables. Test setups ordered according to decreasing sum of R^2 of variables D, G, H & V.

most promising alternatives there include SAR interferometry (Antropov et al., 2018; Solberg et al., 2017) and the low frequency P-band SAR that will be on board the Biomass mission of the European Space Agency (Le Toan et al., 2011).

Acknowledgements

We thank the anonymous reviewers for their helpful comments. This study was funded by the Ministry of Agriculture and Forestry of Finland (project OH300-S42100-03) and by VTT Technical Research Centre of Finland Ltd. The field reference data was provided by the Finnish Forest Centre.

Appendix A. The RMSE% and R^2 values of the different modelling test setups for sixteen forest variables

Table A1
The mean relative RMS errors of the different modelling test setups for sixteen forest variables.

Test setup	S2_All	L8_All	S2_Eqv	L8_Eqv	S2_30_Eqv	S2_30_All	S2_wo_RE	S2_10	S2_20	S2_L8_All	S2_All_rTree	L8_All_rTree
Instrument (L8/S2)	S2	L8	S2	L8	S2	S2	S2	S2	S2	L8 & S2	S2	L8
Pixel resolution	10/20 m	30 m	10/20 m	30 m	30 m	30 m	10/20 m	10 m	20 m	10/20/30 m	10/20 m	30 m
Feature set	A	A	B	B	B	A	C	D	E	A	A	A
D	38.4	44.6	39.7	44.7	42.3	41.1	39.8	39.9	38.8	39.2	47.4	53.7
G	42.5	50.2	43.0	50.2	46.2	44.1	43.6	44.5	43.1	44.5	49.9	58.4
H	30.4	36.6	32.5	37.3	35.2	32.9	32.5	33.1	30.7	31.1	38.2	44.3
V	59.3	72.2	60.8	71.0	64.1	61.5	61.2	61.7	59.4	60.7	68.8	80.4
D_BL	85.5	89.8	87.9	90.1	89.7	87.8	86.8	88.0	85.6	85.6	103.1	108.3
D_Pine	101.4	108.2	101.8	106.3	102.4	103.0	100.9	102.3	101.7	103.0	120.3	124.2
D_Spr	64.5	74.1	66.2	74.9	69.8	68.2	65.8	66.3	66.9	65.5	78.0	88.8
G_BL	119.8	136.3	123.7	126.5	132.0	121.7	116.5	122.5	118.9	121.9	147.4	144.3
G_Pine	110.6	124.0	111.8	125.9	119.0	116.8	109.9	108.2	119.1	116.5	128.7	143.9
G_Spr	92.8	111.7	92.0	111.0	97.4	97.7	93.0	95.2	96.7	94.9	107.1	127.8
H_BL	69.0	73.7	72.1	73.6	74.3	71.7	70.6	72.2	69.2	69.1	85.3	88.7
H_Pine	93.6	99.5	95.3	100.3	96.0	96.1	93.8	94.6	94.3	96.1	112.4	116.5
H_Spr	58.2	67.2	59.3	67.5	62.1	61.2	59.1	59.7	60.2	58.6	70.8	80.1
V_BL	160.0	191.1	163.2	166.6	163.8	153.8	154.0	162.2	157.7	179.9	189.7	180.4
V_Pine	122.3	140.3	122.5	140.8	131.0	130.1	120.3	120.7	130.3	126.7	139.1	158.1
V_Spr	111.7	135.1	112.7	140.5	116.9	114.3	112.8	114.1	115.9	112.9	127.4	152.8

Table A2
The mean R2 values of the different modelling test setups for sixteen forest variables.

Test setup	S2_All	L8_All	S2_Eqv	L8_Eqv	S2_30_Eqv	S2_30_All	S2_wo_RE	S2_10	S2_20	S2_L8_All	S2_All_rTree	L8_All_rTree
Instrument (L8/S2)	S2	L8	S2	L8	S2	S2	S2	S2	S2	L8 & S2	S2	L8
Pixel resolution	10/20 m	30 m	10/20 m	30 m	30 m	30 m	10/20 m	10 m	20 m	10/20/30 m	10/20 m	30 m
Feature set	A	A	B	B	B	A	C	D	E	A	A	A
D	0.60	0.47	0.57	0.47	0.52	0.55	0.57	0.57	0.59	0.59	0.44	0.32
G	0.57	0.42	0.56	0.42	0.50	0.54	0.55	0.54	0.56	0.55	0.45	0.28
H	0.65	0.51	0.61	0.50	0.54	0.60	0.61	0.59	0.65	0.65	0.50	0.35
V	0.56	0.38	0.54	0.39	0.49	0.53	0.54	0.52	0.56	0.55	0.45	0.28
D_BL	0.23	0.17	0.19	0.16	0.15	0.19	0.21	0.19	0.22	0.23	0.11	0.05
D_Pine	0.29	0.22	0.28	0.23	0.28	0.28	0.30	0.28	0.28	0.28	0.17	0.13
D_Spr	0.52	0.38	0.50	0.37	0.45	0.47	0.50	0.50	0.49	0.51	0.37	0.23
G_BL	0.36	0.30	0.33	0.31	0.24	0.32	0.36	0.33	0.37	0.35	0.19	0.17
G_Pine	0.46	0.33	0.44	0.31	0.38	0.40	0.46	0.48	0.36	0.42	0.34	0.21
G_Spr	0.56	0.38	0.57	0.38	0.52	0.52	0.57	0.54	0.53	0.55	0.46	0.27
H_BL	0.29	0.21	0.23	0.20	0.17	0.24	0.26	0.23	0.28	0.29	0.13	0.08
H_Pine	0.32	0.24	0.30	0.24	0.29	0.29	0.32	0.31	0.30	0.29	0.18	0.14
H_Spr	0.54	0.40	0.53	0.39	0.48	0.50	0.53	0.52	0.51	0.54	0.39	0.26
V_BL	0.23	0.24	0.21	0.24	0.16	0.24	0.22	0.20	0.26	0.22	0.09	0.11
V_Pine	0.44	0.29	0.43	0.27	0.35	0.36	0.45	0.44	0.34	0.40	0.34	0.19
V_Spr	0.57	0.38	0.57	0.35	0.53	0.54	0.57	0.55	0.54	0.56	0.46	0.27

Appendix B. Correlation matrix between satellite image channels and forest variables

Table B1
Correlation coefficients between Sentinel-2 bands and forest variables V, H, D, G. The image channels correlating maximally with each forest variable indicated with bold figure.

	S2_B02_Blue	S2_B03_Green	S2_B04_Red	S2_B08_NIR	S2_B05_RE1	S2_B06_RE2	S2_B07_RE3	S2_B8a_nNIR	S2_B11_SWIR1	S2_B12_SWIR2	Tot. stem vol - V	Tree height - H	Stem Diam - D	Basal Area - G
S2_B02_Blue	1	0.877	0.929	0.491	0.830	0.545	0.470	0.478	0.806	0.867	−0.581	−0.490	−0.560	−0.536
S2_B03_Green		1	0.873	0.691	0.933	0.733	0.660	0.667	0.843	0.835	−0.718	−0.664	−0.731	−0.662
S2_B04_Red			1	0.427	0.857	0.498	0.419	0.431	0.816	0.898	−0.600	−0.480	−0.556	−0.538
S2_B08_NIR				1	0.698	0.961	0.966	0.966	0.742	0.567	−0.607	−0.722	−0.683	−0.614
S2_B05_RE1					1	0.786	0.717	0.727	0.922	0.902	**−0.743**	−0.686	**−0.750**	**−0.690**
S2_B06_RE2						1	0.986	0.989	0.814	0.650	−0.643	**−0.737**	−0.718	−0.644
S2_B07_RE3							1	0.995	0.765	0.587	−0.597	−0.706	−0.673	−0.607
S2_B8a_nNIR								1	0.778	0.598	−0.607	−0.709	−0.676	−0.613
S2_B11_SWIR1									1	0.955	−0.695	−0.610	−0.636	−0.635
S2_B12_SWIR2										1	−0.649	−0.515	−0.571	−0.581
Tot. stem vol - V											1	0.717	0.782	0.951
Tree height - H												1	0.955	0.798
Stem Diam - D													1	0.851
Basal Area - G														1

Table B2
Correlation coefficients between Landsat 8 bands and forest variables V, H, D, G. The image channels correlating maximally with each forest variable indicated with bold figure.

	B1 VD Blue	B2 Blue	B3 Green	B4 Red	B5 NIR	B6 SWIR1	B7 SWIR2	Tot. stem vol - V	Tree height - H	Stem Diam - D	Basal Area - G
B1 VD Blue	1	0.977	0.854	0.870	0.407	0.729	0.809	−0.345	−0.261	−0.298	−0.312
B2 Blue		1	0.902	0.932	0.402	0.765	0.863	−0.375	−0.303	−0.349	−0.348
B3 Green			1	0.950	0.611	0.885	0.921	−0.567	−0.527	−0.586	−0.532
B4 Red				1	0.421	0.828	0.932	−0.471	−0.403	−0.465	−0.436
B5 NIR					1	0.770	0.573	−0.543	**−0.644**	**−0.606**	−0.545
B6 SWIR1						1	0.948	**−0.620**	−0.553	−0.575	**−0.567**
B7 SWIR2							1	−0.552	−0.452	−0.499	−0.497
Tot. stem vol - V								1	0.717	0.782	0.951
Tree height - H									1	0.955	0.798
Stem Diam - D										1	0.851
Basal Area - G											1

References

Antropov, O., Rauste, Y., Ahola, H., Häme, T., 2013. Stand-level stem volume of boreal forests from spaceborne SAR imagery at L-band. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 6 (1). https://doi.org/10.1109/JSTARS.2013.2241018.

Antropov, O., Rauste, Y., Tegel, K., Baral, Y., Junttila, V., Kauranne, T., Häme, T., Praks, J., 2018. Tropical forest tree height and above ground biomass mapping in Nepal using Tandem-X and ALOS PALSAR Data. In: Proc. IEEE International Geoscience and Remote Sensing Symposium IGARSS, 2018, pp. 5334–5336.

Astola, H., Bounsaythip, C., Ahola, J., Häme, T., Parmes, E., Sirro, L., Veikkanen, B., 2004. Highforest - forest parameter estimation from high resolution remote sensing data. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives.

Binder, S., Haight, R.G., Polasky, S., Warziniack, T., Mockrin, M.H., Deal, R.L., Arthaud, G., 2017. Assessment and Valuation of Forest Ecosystem Services: State of the Science Review'. pp. 56 (May).

Blum, A.L., Langley, P., 1997. Selection of relevant features and examples in machine learning. Artif. Intell. 97 (1–2), 245–271. https://doi.org/10.1016/S0004-3702(97)00063-5.

Breiman, L., 1998. Classification and Regression Trees. Chapman & Hall/CRC.

Cajander, A.K., 1949. Forest types and their significance. Acta For. Fennica 1–71.

Chrysafis, I., Mallinis, G., Siachalou, S., Patias, P., 2017. Assessing the relationships between growing stock volume and Sentinel-2 imagery in a Mediterranean forest ecosystem. Remote Sens. Lett. 8 (6), 508–517. https://doi.org/10.1080/2150704X.2017.1295479. (Taylor & Francis).

Cohen, W.B., Goward, S.N., 2017. Landsat's role in ecological applications of remote sensing. 54 (6).

Delegido, J., Verrelst, J., Alonso, L., Moreno, J., 2011. Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. Sensors 11 (7), 7063–7081. https://doi.org/10.3390/s110707063.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sens. Environ. 120, 25–36. https://doi.org/10.1016/j.rse.2011.11.026.

FAO, 2000. Global Forest Survey, Concept Paper. 2000. Forest Resource Assessment Programme. Working Paper 28 FAO, Rome.

FAO, 2015. FAO Global Forest Resources Assessment 2015. UN Food and Agriculture Organization, Rome.

Forkuor, G., Dimobe, K., Serme, I., Tondoh, J.E., 2017. Landsat-8 vs. Sentinel-2: examining the added value of sentinel-2's red-edge bands to land-use and land-cover mapping in Burkina Faso. GISci. Remote Sens. 00 (00), 1–24. https://doi.org/10.1080/15481603.2017.1370169. (Taylor & Francis).

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 (3). https://doi.org/10.14214/sf.1095.

Gill, P.E., Murray, W., 1978. Algorithms for the solution of the nonlinear least-squares problem. SIAM J. Numer. Anal. 15 (5), 977–992. https://doi.org/10.1137/0715063.

Griffiths, P., Kuemmerle, T., Baumann, M., Radeloff, V.C., Abrudan, I.V., Lieskovsky, J., Munteanu, C., Ostapowicz, K., Hostert, P., 2014. Forest disturbances, forest recovery, and changes in forest types across the carpathian ecoregion from 1985 to 2010 based on landsat image composites. Remote Sens. Environ. 151, 72–88. https://doi.org/10.1016/j.rse.2013.04.022.

Häme, T., Rauste, Y., Antropov, O., Ahola, H.A., Kilpi, J., 2013. Improved mapping of tropical forests with optical and sar imagery, part ii: above ground biomass estimation. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 6 (1). https://doi.org/10.1109/JSTARS.2013.2241020.

Härkönen, S., Lehtonen, A., Eerikäinen, K., Peltoniemi, M., Mäkelä, A., 2011. Estimating forest carbon fluxes for large regions based on process-based modelling, NFI data and Landsat satellite images. For. Ecol. Manag. 262 (12), 2364–2377. https://doi.org/10.1016/j.foreco.2011.08.035.

Helliwell, I.S., Turega, M.A., Cottis, R.A., 1995. Accountability of neural networks trained with 'real world' data. IEE Conf. Publ. (409), 26–28.

Hinton, G.E., Osindero, S., Teh, Y.-W., 2006. A fast learning algorithm for deep belief nets. Neural Comput. 18 (7), 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527.

Hyyppä, H., Inkinen, M., Engdahl, M., 2000. Accuracy Comparison of Various Remote Sensing Data Sources in the Retrieval of Forest Stand Attributes. vol. 128. pp. 109–120.

Immitzer, M., Vuolo, F., Atzberger, C., 2016. First experience with Sentinel-2 data for crop and tree species classifications in central Europe. Remote Sens. 8 (3). https://doi.org/10.3390/rs8030166.

IPCC, 2006. IPCC Guidelines for National Greenhouse Gas Inventories. Institute for Global Environmental Strategies (IGES).

James, G., Witten, D., Hastie, T., Tibshirani, R., 2000. An introduction to statistical learning. Curr. Med. Chem. https://doi.org/10.1007/978-1-4614-7138-7.

Kankare, V., Vastaranta, M., Holopainen, M., Räty, M., Yu, X., Hyyppä, J., Hyyppä, H., Alho, P., Viitala, R., 2013. Retrieval of forest aboveground biomass and stem volume with airborne scanning LiDAR. Remote Sens. 5 (5), 2257–2274. https://doi.org/10.3390/rs5052257.

Kaufman, Y.J., Sendra, C., 1988. Algorithm for automatic atmospheric corrections to visible and near-IR satellite imagery. Int. J. Remote Sens. 9 (8), 1357–1381. https://doi.org/10.1080/01431168808954942.

Kohavi, R., John, G.H., 1997. Wrappers for feature subset selection. Artif. Intell. 97 (1–2), 273–324. https://doi.org/10.1016/S0004-3702(97)00043-X.

Korhonen, L., Hadi, Packalen, P., Rautiainen, M., 2017. Comparison of Sentinel-2 and Landsat 8 in the estimation of boreal forest canopy cover and leaf area index. Remote Sens. Environ. 195, 259–274. https://doi.org/10.1016/j.rse.2017.03.021. (Elsevier Inc.).

Le Toan, T., Quegan, S., Davidson, M.W.J., Balzter, H., Paillou, P., Papathanassiou, K., Plummer, S., Rocca, F., Saatchi, S., Shugart, H., Ulander, L., 2011. The BIOMASS mission: mapping global forest biomass to better understand the terrestrial carbon cycle. Remote Sens. Environ. 115 (11), 2850–2860. https://doi.org/10.1016/J.RSE.2011.03.020. (Elsevier).

Lindberg, E., Hollaus, M., 2012. Comparison of methods for estimation of stem volume, stem number and basal area from airborne laser scanning data in a hemi-boreal forest. Remote Sens. 4 (4), 1004–1023. https://doi.org/10.3390/rs40401004.

Mäkelä, H., Pekkarinen, A., 2004. Estimation of Forest Stand Volumes by Landsat TM Imagery and Stand-level Field-inventory Data. vol. 196. pp. 245–255. https://doi.org/10.1016/j.foreco.2004.02.049.

Mäkelä, A., Pulkkinen, M., Kolari, P., Lagergren, F., Berbigier, P., Lindroth, A., Loustu, D., Nikinmaa, E., Vesala, T., Hari, P., 2007. Developing an empirical model of stand GPP with the LUE approach: analysis of eddy covariance data at five contrasting conifer sites in Europe. Glob. Chang. Biol. 0 (0). https://doi.org/10.1111/j.1365-2486.2007.01463.x. (Blackwell Publishing Ltd, p. 071124112207003–???).

Mandanici, E., Bitelli, G., 2016. Preliminary comparison of sentinel-2 and landsat 8 imagery for a combined use. Remote Sens. 8 (12), 1014. https://doi.org/10.3390/rs8121014.

Molinier, M., López-Sánchez, C.A., Toivanen, T., Korpela, I., Corral-Rivas, J.J., Tergujeff, R., Häme, T., 2016. Relasphone-mobile and participative in situ forest biomass measurements supporting satellite image mapping. Remote Sens. 8 (10). https://doi.org/10.3390/rs8100869.

Persson, H.J., 2016. Estimation of Boreal Forest Attributes From Very High Resolution Pléiades Data. https://doi.org/10.3390/rs8090736.

Persson, H.J., Fransson, J.E.S., 2017. Comparison between TanDEM-X- and ALS-based estimation of aboveground biomass and tree height in boreal forests. Scand. J. For. Res. 32 (4), 306–319. https://doi.org/10.1080/02827581.2016.1220618. (Taylor & Francis).

Persson, H., Wallerman, J., Olsson, H., Fransson, J.E.S., 2013. Estimating forest biomass and height using optical stereo satellite data and a DTM from laser scanning data. Can. J. Remote. Sens. 39 (3), 251–262. https://doi.org/10.5589/m13-032. (Taylor & Francis).

Peuhkurinen, J., Maltamo, M., Vesa, L., Packalén, P., 2008. Estimation of forest stand characteristics using spectral histograms derived from an Ikonos satellite image. Photogramm. Eng. Remote Sens. 74, 1335–1341 (American Society for Photogrammetry and Remote Sensing).

Reese, H., Nilsson, M., Pahlén, T.G., Hagner, O., Joyce, S., Tingelöf, U., Egberth, M., Olsson, H., 2003. Countrywide estimates of forest variables using satellite data and field data from the National Forest Inventory. AMBIO J. Hum. Environ. 32 (8), 542–548. https://doi.org/10.1579/0044-7447-32.8.542.

Reif, M., Shafait, F., 2014. Efficient feature size reduction via predictive forward selection. Pattern Recogn. 47 (4), 1664–1673. https://doi.org/10.1016/j.patcog.2013.10.009. (Elsevier).

Rokach, L., Maimon, O., 2007. Data mining with decision trees. In: World Scientific. Series in Machine Perception and Artificial Intelligencehttps://doi.org/10.1142/6604.

Rumelhart, D.E., McClelland, J.L., 1986. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. vol. 1 MIT Press.

Santoro, M., Beer, C., Cartus, O., Schmullius, C., Shvidenko, A., McCallum, I., Wegmüller, U., Wiesmann, A., 2011. Retrieval of growing stock volume in boreal forest using hyper-temporal series of Envisat ASAR ScanSAR backscatter measurements. Remote Sens. Environ. 115 (2), 490–507. https://doi.org/10.1016/j.rse.2010.09.018. (Elsevier Inc.).

Solberg, S., Hansen, E.H., Gobakken, T., Næssset, E., Zahabu, E., 2017. Biomass and InSAR height relationship in a dense tropical forest. Remote Sens. Environ. 192, 166–175.

Tokola, T., LeToan, T., v. Poncet, F., Tuominen, S., Holopainen, M., 2007. Forest reconnaissance surveys: comparison of estimates based on simulated Terrasar, and optical data. Photogramm. J. Finland 20 (2), 64–79.

Tomppo, E., Katila, M., Mäkisara, K., Peräsaari, J., 2014. The multi-source National Forest Inventory of Finland — methods and results 2011. In: Working Papers of the Finnish Forest Research Institute. vol. 319. http://urn.fi/URN:ISBN:978-951-40-2428-3.

Topaloğlu, R.H., Sertel, E., Musaoğlu, N., 2016. Assessment of classification accuracies of Sentinel-2 and Landsat-8 data for land cover/use mapping. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives. vol. 41. pp. 1055–1059. https://doi.org/10.5194/isprsarchives-XLI-B8-1055-2016. (July).

UNFCCC, 2014. Key Decisions Relevant for Reducing Emissions from Deforestation and Forest Degradation in Developing Countries (REDD+). Decision booklet REDD+ UNFCCC secretariat (June 2014).

UNFCCC, 2016. Report of the Conference of the Parties on its twenty-first session, held in Paris from 30 November to 13 December 2015. Addendum. Part two: Action taken by the Conference of the Parties at its twenty-first session. http://unfccc.int/resource/docs/2015/cop21/eng/10a01.pdf.

Varvia, P., Lähivaara, T., Maltamo, M., Packalen, P., Tokola, T., Seppänen, A., 2017. Uncertainty quantification in ALS-based volume estimation. IEEE Trans. Geosci. Remote Sens. 55 (3), 1671–1681.

Vincent, Pascalvincent P., Larochelle, Larocheh H., 2010. Stacked denoising auto-encoders: learning useful representations in a deep network with a local denoising criterion Pierre-Antoine Manzagol. J. Mach. Learn. Res. 11, 3371–3408. https://doi.

org/10.1111/1467-8535.00290.
Wulder, M.A., Masek, J.G., Cohen, W.B., Loveland, T.R., Woodcock, C.E., 2012. Remote sensing of environment opening the archive: how free data has enabled the science and monitoring promise of Landsat. Remote Sens. Environ. 122, 2–10. https://doi.org/10.1016/j.rse.2012.01.010. (Elsevier B.V.).
Ylitalo, E. (Ed.), 2012. Finnish Statistical Yearbook of Forestry 2012. Finnish Forest Research Institute (Metla), pp. 454 s ISBN 978-951-40-2392-7, ISBN 978-951-40-2391-0.
Zhao, P., Lu, D., Wang, G., Wu, C., Huang, Y., Yu, S., 2016. Examining spectral reflectance saturation in Landsat imagery and corresponding solutions to improve forest aboveground biomass estimation. Remote Sens. 8 (6), 469.
Zhu, Z., Woodcock, C.E., 2014. Continuous change detection and classification of land cover using all available Landsat data. Remote Sens. Environ. 144, 152–171. https://doi.org/10.1016/J.RSE.2014.01.011. (Elsevier).