Article

Assessing the Suitability of Future Multi- and Hyperspectral Satellite Systems for Mapping the Spatial Distribution of Norway Spruce Timber Volume

Sascha Nink [1],*, Joachim Hill [1], Henning Buddenbaum [1], Johannes Stoffels [1], Thomas Sachtleber [1] and Joachim Langshausen [2]

[1] Environmental Remote Sensing and Geoinformatics, University of Trier, Behringstraße 21, 54286 Trier, Germany; E-Mails: hillj@uni-trier.de (J.H.); buddenbaum@uni-trier.de (H.B.); stoffels@uni-trier.de (J.S.); t.sachtleber@gmx.de (T.S.)

[2] Landesforsten Rheinland-Pfalz, Zentralstelle der Forstverwaltung, Südallee 15-19, 56068 Koblenz, Germany; E-Mail: joachim.langshausen@wald-rlp.de

* Author to whom correspondence should be addressed; E-Mail: ninks@uni-trier.de; Tel.: +49-651-201-4620; Fax: +49-651-201-3815.

Academic Editors: Josef Kellndorfer and Prasad S. Thenkabail

Abstract: The availability of accurate and timely information on timber volume is important for supporting operational forest management. One option is to combine statistical concepts (e.g., small area estimates) with specifically designed terrestrial sampling strategies to provide estimations also on the level of administrative units such as forest districts. This may suffice for economic assessments, but still fails to provide spatially explicit information on the distribution of timber volume within these management units. This type of information, however, is needed for decision-makers to design and implement appropriate management operations. The German federal state of Rhineland-Palatinate is currently implementing an object-oriented database that will also allow the direct integration of Earth observation data products. This work analyzes the suitability of forthcoming multi- and hyperspectral satellite imaging systems for producing local distribution maps for timber volume of Norway spruce, one of the most economically important tree species. In combination with site-specific inventory data, fully processed hyperspectral data sets (HyMap) were used to simulate datasets of the forthcoming EnMAP and Sentinel-2 systems to establish adequate models for estimating timber volume

maps. The analysis included PLS regression and the *k*-NN method. Root Mean Square Errors between 21.6% and 26.5% were obtained, where *k*-NN performed slightly better than PLSR. It was concluded that the datasets of both simulated sensor systems fulfill accuracy requirements to support local forest management operations and could be used in synergy. Sentinel-2 can provide meaningful volume distribution maps in higher geometric resolution, while EnMAP, due to its hyperspectral coverage, can contribute complementary information, e.g., on biophysical conditions.

1. Introduction

Forests, with large storage of biomass and their function as an important terrestrial carbon dioxide sink, cover about 4 billion hectares worldwide, which is almost one third of the Earth's land surface [1–3]. Forest ecosystems also cover almost one third of the territory of Germany and provide a wide range of economic benefits and ecosystem services [4]. In post-industrial regions like Europe, the priority in forest management is sustaining multi-functional forest systems. Accordingly, the strategic objectives of the German forest policy for the coming decades are focused on developing a balance between the diverging demands on ecological, economic, and socioeconomic functions of forests and their sustainable performance [5]. These strategic objectives, in combination with national and international commitments for reporting on forest resources, are triggering an increasing demand for expanded information on forest resources [6,7].

In the federal state of Rhineland-Palatinate, forested areas cover more than 42% of the land and, notwithstanding the ecological, social, and cultural services they are providing, they also represent important economic value. As part of a centralized inventory action at the federal level (Bundeswaldinventur), exhaustive state forest inventories are based on a regular sampling grid and carried out at approximately 10-year intervals. In addition, information on the level of local forest districts is collected through expert assessments at 5–10 year intervals [8]. The latter is staff-intensive, time-consuming, and therefore unable to comply with short-term information demands. Furthermore, terrestrial assessments are not capable of capturing the full spatial diversity of forest attributes. Since remote sensing data products may compensate for such shortcomings [9], the federal state of Rhineland-Palatinate has begun to explore pathways for a direct integration of Earth observation data products into an innovative object-oriented data base concept, which is currently being implemented.

Satellite observations have already qualified as an important tool for the assessment of forest resources and forest health on global scale. Applications include forest mapping [10], forest cover change detection [11–13], and accurate estimates of forest biomass and its dynamics, which contribute to global carbon flux analyses [14]. Remotely sensed assessment of past and present forest carbon stock is important for global carbon emissions trade and for international report duties (e.g. the Kyoto Protocol) [15]. Some studies have also focused on how remote sensing may support operational forest management and contribute to forest ecosystem research on the regional to local scale [16–18].

On a national and regional scale, the main foci of forest resource assessment are the estimation of tree species distribution mapping, forest health, and the distribution of timber volume [19]. Studies on implementing remote sensing methods in forest inventories have already been carried out in numerous countries. Finland was the first to successfully derive operational results in the early 1990s, introducing the non-parametric *k*-Nearest Neighbor method (*k*-NN) in forestry applications [20]. Promising results were also achieved by combining remotely sensed image data and ground truth data in the National Forest Inventory of Sweden [21], as well as for tax assessment of forest properties [22]. In Norway, maps were derived by combining data from the national forest inventory and Landsat-5 TM data [23]. McInerney and Nieuwenhuis [24] obtained good results when predicting timber volume and basal area in Ireland based on non-parametrical estimation techniques with SPOT 4 imagery and terrestrial inventory data. In Austria, satisfying agreement between estimates and terrestrial inventories was found [25]. The use of Landsat 7 ETM+ data for estimating basal area and leaf area index in a Mediterranean forest was less successful; this was attributed to more complex environmental and spectral conditions [26]. In Germany, studies based on Landsat-5 TM data in North Rhine-Westphalia produced useful maps for large areas, but the results on the level of local management units (forest stands) were not satisfactory [27,28]. The potential for regionalizing forest inventory data was demonstrated in a study from South Germany, presenting a relatively small estimation error for the mapping of timber volume using *k*-NN, but moderate accuracy for the classification of six main tree species [29]. A comprehensive summary of related studies with sensors and methods used for the estimation of aboveground biomass or timber volume is given in Lu [30] and Gleason [31].

Regular implementation of satellite-based timber volume distribution maps, however, still suffers from temporal data availability and/or data quality [27,32]. Currently, the availability of observations depends on satellites, where low frequency revisit times, limited spatial coverage, and, in some cases, high costs are the limiting factors, particularly in areas with relatively short vegetation periods and frequent cloud cover. With respect to these limitations of current satellites, the forthcoming Sentinel-2 mission [33,34] has the potential for a real breakthrough. Besides the spectral and spatial resolution characteristics (13 spectral bands in up to 10 m spatial resolution), its large area coverage and the increased temporal repetition rates qualify Sentinel-2 as a prime candidate for operational forest mapping. This will be particularly true when two identical systems will be operating in parallel. These operational capabilities of Sentinel-2 will be complemented by the planned hyperspectral satellite mission EnMAP, expected to be launched in early 2018. EnMAP will be equally qualified as a tool for forest resource assessment and for monitoring concepts [35]. Conceptual thoughts on the synergistic use of both earth observation systems have raised much interest: while Sentinel-2 provides a wide spatial coverage with increased observation frequencies, EnMAP might be expected to act as a spectral "magnifying glass" for detailed and more profound analyses of bio-physical canopy conditions in hot spot areas.

Against this background, this study focuses on assessing the potential of Sentinel-2 and EnMAP for spatially explicit mapping of timber volume as a key variable in sustainable forest management at stand level. The following objectives will be addressed:

1. Integration of routinely acquired forest resource assessments and spectral information for estimating the spatial distribution of timber volume in managed Norway spruce (*Picea abies*) forests.

2. Identification of sensitive wavelengths for volume estimation in intensively managed Norway spruce forests.

3. Assessment of concepts (PLSR *vs.* *k*-NN) for producing spatially explicit maps of Norway spruce timber volume within administrative forest units.

4. Evaluation of the spectral and spatial resolution characteristics of Sentinel-2 and EnMAP for mapping Norway spruce timber volume.

2. Material and Methods

2.1. Study Area

The study area includes three locations (Idarwald, Gerolstein, and Daun) in Rhineland-Palatinate, Germany, with a total size of 10,868 hectares, for which hyperspectral images have been acquired. The Idarwald is part of the Hunsrück low mountain range (49°49′N, 7°10′E), while the Gerolstein (50°12′N, 6°41′E) and Daun sites (50°9′N, 6°52′E) are located in the Eifel mountains, west of the Moselle valley (Figure 1).
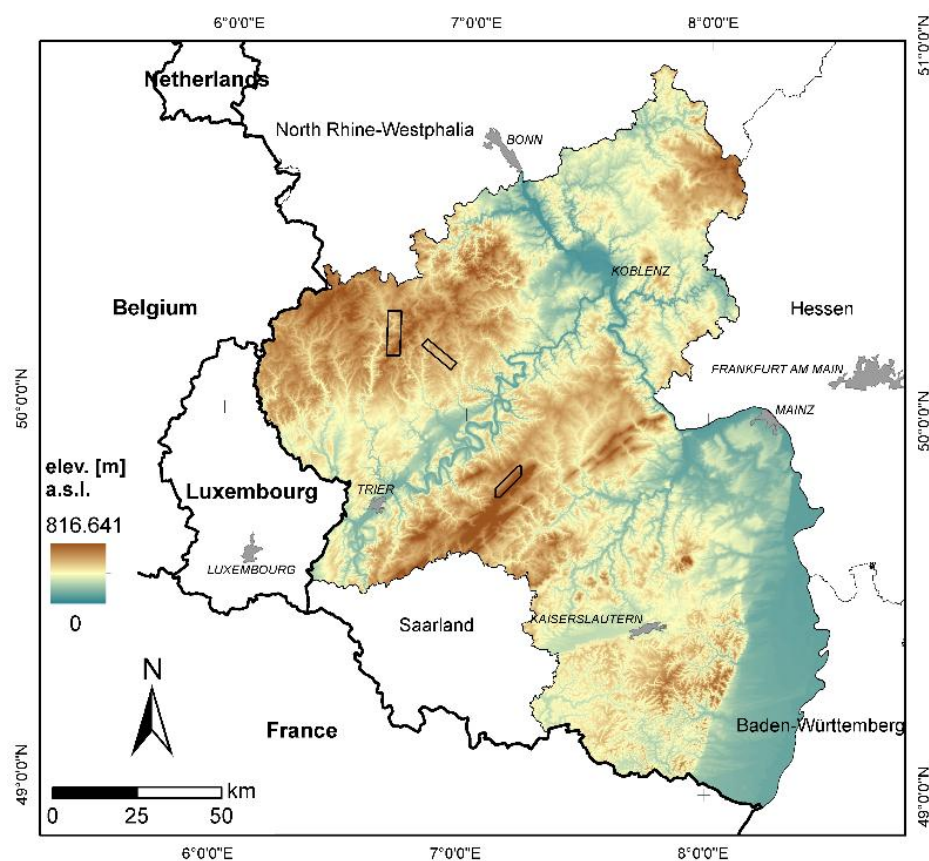

Figure 1. Study areas in the German federal state of Rhineland-Palatinate (black rectangles).

The geologic setting of the Idarwald area (Devonian slate and quartzite) produced soils with limited nutrient availability, which are mostly used for forestry. The soils in Gerolstein and Daun developed on volcanic material; as a result, agricultural land use is more abundant there. The annual temperature variability at all study sites is small due to maritime influence [36]. Within the study areas, approximately 56% of the land surface is covered by forest. Although European beech is the natural vegetation, spruce forests have become dominant, due to large-scale plantation programs during the 19th and 20th centuries. Norway spruce (*Picea abies*) is the main species with a coverage proportion of 44.8%, followed by European beech (*Fagus sylvatica*) with 31.2%. Additional tree species include oak (*Quercus petraea* and *Quercus robur*) with 7.8%, Douglas fir (*Pseudotsuga menziesii*) with 5.3% and Scots pine (*Pinus sylvestris*) with 1.8% proportion. Norway spruce stands are present in all development stages and aged between 4 and 99 years. The average and maximum timber volumes amount to 283.69 m³/ha and 586 m³/ha, respectively. Norway spruce represents a highly important economic factor; approximately 40 percent of timber logged in Rhineland-Palatinate is Norway spruce, and the demand continues to be high.

2.2. Data

2.2.1. Airborne HyMap imagery

Three high-resolution airborne hyperspectral datasets were acquired with the HyMap instrument during a European airborne campaign in 2003 (Figure 1). One hundred twenty-eight spectral bands ranging between 450 and 2480 nanometers with bandwidths between 10 and 20 nanometers were recorded [37]. All datasets were acquired on 15 July 2003 around solar noon. The tracks are 2.5 to 4 km wide and have a length of 10–13 km. According to a sensor altitude above ground, the ground sampling distance varies between 4 and 7 meters. Illumination effects, which are due to differences in reflectance according to the view angle, were corrected with an across-track illumination correction [17]. Geometric correction to sub-pixel accuracy was accomplished with PARGE software [38,39]. The data was projected to the local Gauss-Krüger coordinate system [17,40]. Radiometric correction involved sensor calibration and correction of atmospheric and topographic effects with the AtCPro© radiative transfer code [41,42], which is based on the 5S concept [43]. A limited number of spectral bands with too low signal-to-noise-ratio (1330–1490 nm and 1770–1990 nm) were eliminated [44].

2.2.2. EnMAP and Sentinel-2 Data Simulation

EnMAP is designed to acquire 242 spectral bands with two independent detector systems. The visible and near-infrared (VNIR) spectral range from 420 to 1000 nm will be covered with a mean bandwidth of 6.5 nm and the shortwave-infrared (SWIR) spectral range from 900 to 2450 nm with a bandwidth of approximately 10 nm [35]. Due to the lack of sufficient data in the spectral regions around the wavelength of 1.3 and 1.9 μm, our simulated EnMAP images consist of only 187 spectral bands. EnMAP's geometric resolution has been achieved by aggregating imaging spectrometer pixels to the size of 30 m, using a Gaussian low pass kernel to approximate EnMAP's point spread function [45]. For simulating the EnMAP spectral bands, Gaussian response functions with an equivalent feature width at half maximum (FWHM) have been used (Figure 2).
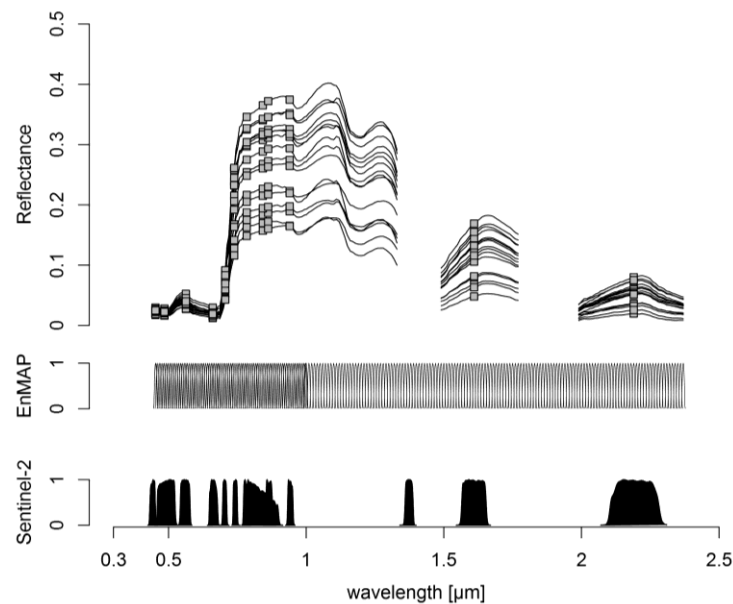
Figure 2. Simulated EnMAP (lines) and Sentinel-2 spectra (squares) of various tree species and stand structures in comparison to the spectral response functions used for generating the datasets based on interpolated HyMap reflectance spectra.

Sentinel-2 will provide 13 spectral bands in the range between 443 and 2190 nm (Figure 2), covering the VNIR and SWIR spectral range with varying spatial resolution (10, 20, and 60 m) [46]. In comparison to current operational sensors, Sentinel-2 will feature four spectral bands in the near-infrared (NIR) spectral range: three bands with a narrow width of 15–20 nm at 740 nm, 783 nm, and 865 nm and a geometric resolution of 20 m; and one spectral band at 842 nm with a bandwidth of 115 nm and 10 m geometric resolution. Furthermore, the Red Edge spectral range (705 nm) will be covered with a bandwidth of 15 nm and a geometric resolution of 20 m [33]. Our simulated Sentinel-2 images contain 12 spectral bands that were generated by convolving the continuously interpolated HyMap spectra with the Sentinel-2 spectral response functions (Figure 2). All simulated Sentinel-2 bands have been generated with a pixel size of 10 meters, assuming that efficient fusion and sharpening algorithms will allow for scaling all Sentinel-2 bands to this common spatial resolution.

2.2.3. Timber Volume Reference Data

For publicly managed forests, stand information is available in a tabular forest information database (FID), which is part of a geo-information system developed and maintained by the state forest administration of Rhineland-Palatinate [47]. All information is related to forest management units and includes economic and ecological attributes. The information on timber volume for each management unit stems from field assessments carried out by specifically trained staff ("expert assessments") within five- to 10-year repetition cycles; only a few parameters are measured directly. The data used in this study were acquired between 1998 and 2009.

A second set of reference information was collected in 2002 based on the systematic sampling grid (spacing = 2 × 2 km) used during the Federal State Forest Inventory (FSFI), which is carried out in 10-year-intervals. A range of forest parameters is measured according to specifically designed

formalized methods (e.g., "Bitterlich sampling"). Measured parameters include tree parameters (tree-type, height, age, diameter at breast-height) and the exact position of single trees within the plot. However, owing to the limited extension of our study sites, and the exclusion of inventory plots with mixed tree species, the number of available plots from the FSFI grid is limited.

Both information sources cannot be combined due to fundamental differences between the survey methods (yield-table based expert surveys *vs.* measurements at a systematic sampling grid), the spatial dimension of assessment units (complete management units *vs.* point measurements), and the time of acquisition (irregular intervals *vs.* regular 10-year-intervals). Yet, a comparison between both information levels provided important information for understanding the quality of the expert assessments of timber volume which are available in the forest information database. The study sites are maintained by the staff of four different forest administration offices; in total, 194 FSFI plots are available. A subset of 137 FSFI plots, which are either pure or dominated by Norway spruce, could be matched to the corresponding information from the FID. After careful inspection of aerial photographs and both data sets, 55 plots had to be excluded for different reasons: for example, management interventions (clearing or thinning) had significantly changed the stand in comparison to the time at which FID data were collected. A number of FSFI plots are located at stand borders or at the border of administrative forest units and some FSFI plots are within very heterogeneous FID units. In both cases, the data could not be linked. Finally, in some cases it was doubtful that the values of the dataset were representative for forest structure or age. The comparison of the remaining 82 sampling units exhibits a reasonably high correlation (R = 0.79), and thus confirms our expectation that the FID data provide consistent information (Figure 3).
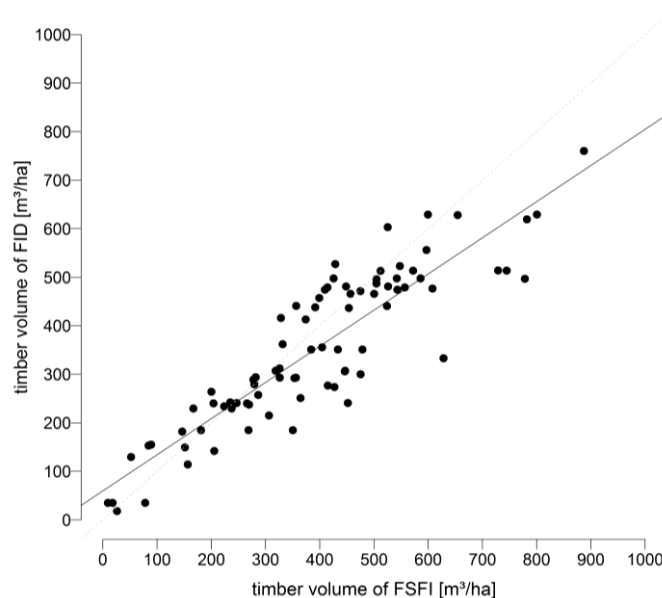

Figure 3. Comparison of timber volume from the plot-based Federal State Forest Inventory (FSFI 2003) and the corresponding expert assessments on stand level (FID, 1998–2009).

With respect to the relatively small number of FSFI plots within pure Norway spruce stands, it was decided to use only reference data on timber volume from the Rhineland-Palatinate forest administration database. This data set includes 56 forest units with pure stands of Norway spruce in

total, of which 25 units are located in the Eifel and 31 in the Hunsrück study site. Timber volume values are normally distributed (Shapiro-Wilk test, $p < 0.05$) and are considered representative for the study region.

2.3. Methods

2.3.1. Extraction of Spectral References

The mapping approach presented in this paper builds on establishing a set of spectral references that corresponds to the biophysical and structural characteristics of spruce stands with known timber volume. Since individual pixels may not be representative [48], we decided to extract these reference signatures from pixel aggregates at the geometric center of homogeneous stands. While other studies [20,49] used circular plots, in this work we chose a kernel-based sampling scheme to ensure consistency between the different spatial resolution of the simulated EnMAP and Sentinel-2 scenes. A sampling area of 900 m²was considered sufficiently large to represent the spectral characteristics of the structurally different forest canopies. Spectral references for the simulated EnMAP data were derived from a single pixel (30 × 30 m), while a 3 × 3 sampling kernel was used to extract the corresponding information from the simulated Sentinel-2 data (resampled to 10 × 10 m geometric resolution); the sampling areas were thus identical for both simulated datasets.

2.3.2. Reference Data and Identification of Sensitive Spectral Regions

The terrestrial inventories (FSFI) within the usable reference stands for this study had been updated within a mean time window of +/− 2 years with respect to the hyperspectral image acquisition in 2003. While the average annual growth rate of spruce (15.6 m^3/ha) as communicated by the latest German Federal Forest Inventory [50] is not expected to substantially degrade the relationship between spectral characteristics and stand-specific timber volume, it was considered important to identify spruce stands where management interventions (thinning, harvesting) have fundamentally changed stand properties compared to the time of HyMap data collection.

In addition to consulting the available management records, it was decided to conduct an analysis of the sensitivity of the wavebands to timber volume. Linear regression analysis of the relationship between simulated EnMAP and Sentinel-2 spectral reflectance measurements and stand-specific timber volume were computed. It was expected that the coefficient of determination (R^2) would be indicative for identifying the spectral ranges that are particularly sensitive to varying timber volume (*i.e.*, support band selection), but would also serve to identify outliers that should be removed from the analysis (owing to fundamentally altered stand conditions). The outlier analysis was based on a multiple random sampling procedure, where for each run 75% of all data pairs served as input for a linear regression and the subsequent analysis of its residuals; a specific data-pair was flagged as an outlier if its residual value was repeatedly higher than 1.5 standard deviations in the respective sample. The reference areas identified as outliers were additionally inspected on aerial photographs, which in all cases confirmed substantially modified stand properties.

2.3.3. Predictive Modeling

Generating spatially explicit estimates of timber volume based on indicators of canopy properties derived from hyper- or multi-spectral data faces the fundamental challenge of constructing good predictive models. If the factors are few in number, are not significantly redundant (collinear), and have a well-understood relationship to the responses, then multiple linear regression may provide an efficient solution. However, when using spectral measurements these conditions usually are not fulfilled because the factors (spectral band responses) are numerous and typically highly collinear.

For generating spatially explicit timber volume maps, we decided to test one parametrical (Partial-Least-Squares Regression, PLSR) and one non-parametrical model (*k*-Nearest-Neighbors, *k*-NN). If the emphasis is on predicting the responses (timber volume) and not necessarily on trying to understand the underlying relationships between the spectral variables, PLSR is an appropriate choice. Developed in the 1960s by the Swedish statistician Herman Wold, PLSR has qualified as a standard tool in chemistry and engineering [51], and has been applied in numerous spectrometric studies at the laboratory, field, and canopy scale [52–55]. It is based on the assumption that the response variable is explained by only a few underlying or latent factors (PLS components) that account for most of the variation in the response. PLS regressions in the context of forestry remote sensing have recently been applied for the estimation of multiple forestry parameters [56], or to derive aboveground biomass [57,58]. If the number of spectral variables is limited (as is the case with a multi-spectral imaging system), PLSR loses some of its specific advantages over linear multiple regression approaches, but can still be successfully applied.

Another approach for mapping timber volume is the *k*-NN algorithm, a non-parametric method widely used for classification and regression. The method has been widely used in forestry applications, starting from the pioneering work done in Finland and Sweden as a support for the National Forest Inventory [59]. In all cases, the input consists of the *k* closest training examples in the feature space (defined by the spectral variables). The retrieved value of the object (*i.e.*, timber volume) is derived from a linear combination of known response values from *k* selected data-pairs ("*k* nearest neighbors") in the reference database. Most widespread for selecting the training examples from the reference dataset is the Euclidean distance:

$$d_{p,p_i} = \sqrt{\sum_{j=1}^{nc} a_j^2 \cdot (x_{p,j} - x_{p_i,j})^2} \qquad (1)$$

where *nc* is the number of auxiliary variables, or spectral bands; $x_{p_i,j}$ is the spectral information of the image pixel with unknown response value (estimation pixel) in the *j*-th spectral band; and $x_{p,j}$ is the spectral information of the reference-data-pair *p* in the *j*-th spectral band. The factor a_j^2 steers the influence of the spectral variables.

Weighting factors are derived for the *k* selected training examples, based on the distance to the estimation pixel in feature space. These can be either equally weighted, or differentially weighted by taking into consideration the higher influence of the derived distances, which increases the preference of training examples in closer distance:

$$w_{p_i,\mathrm{p}} = \frac{1}{d_{p_i,p}^t} \Bigg/ \sum_{i=1}^{k} \frac{1}{d_{p_i,p}^t} \tag{2}$$

$d_{p_i,p}^t$ is the distance calculated in equation 1. The exponent t, $t = \{0,1,2\}$, further influences the weighting factor. For $t > 0$, training examples with closer distances to the estimation pixel are assigned higher weighting factors and *vice versa*. The sum of all weighting factors is always 1.

The response value $\widehat{y_p}$ of the estimation pixel is then calculated as a linear combination between the weighting factors $w_{p_i,\mathrm{p}}$ and response values y_{p_i} of the k nearest training examples from the reference data:

$$\widehat{y_p} = \sum_{i=1}^{k} w_{p_i,p} \cdot y_{p_i} \tag{3}$$

Since the k-NN algorithm uses the k closest data points for estimation, it is able to take full advantage of local information and provide nonlinear, adaptive estimates for each data point. The k-NN algorithm can be optimized based on the following parameters: number of chosen neighbors, *i.e.* training examples from reference data, weighting factors for the spectral variables, and weighting factor of distances.

However, the k-NN algorithm also has some disadvantages to be considered. Besides being computationally intensive and demanding in storage when applied to large datasets, the most relevant limitation in the context of this study is its susceptibility to data dimensionality. With respect to the excessive number of spectral bands of the simulated EnMAP data, and in support of the idea of also comparing the performance of the k-NN method for simulated EnMAP- and Sentinel-2-data, a principal component transformation was applied to reduce the dimensionality of both datasets, while still retaining as much of the variance in the dataset as possible. Transformation eigenvectors were defined exclusively from the forested areas in the three HyMap images used in this study. One single set of eigenvectors was used for all images. Correlation of each principal component to timber volume was calculated.

2.3.4. Model Validation

The best way to measure the predictive ability of a model is obviously to test it on a set of data not used in estimation. However, due to the limited area coverage of the HyMap data available for this study, there were not enough reference plots to allow some of them to be kept back for testing. Instead, leave-one-out cross validation (loo-cv) [60] has been used to find the optimal model parameterizations. The widely used validation metric is the absolute and relative Root-Mean-Squared-Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (y_{p,i} - y_{o,i})^2} \tag{4}$$

$$\%RMSE = \frac{RMSE}{\overline{y_o}} \cdot 100 \tag{5}$$

n is the number of observations (reference data pairs), $y_{p,i}$ are the estimated values, and $y_{o,i}$ are observed values of the *i*-th data-pair. For reasons of comparability, the relative RMSE (Equation 5) is more probate with $\overline{y_o}$ as the mean value of n observations.

To keep the models simple, different information criteria like the Akaike Information Criterion (AIC), the Schwartz–Bayes Criterion (SBC), and the adjusted R ²(Equations (6)–(8)) are more probate means of validation, as they give penalties on higher numbers of auxiliary variables (PLS components or spectral bands) used in the models [61,62].

$$AIC = \ln(\sigma^2_{Residuals}) + \frac{M}{n} \cdot 2 \tag{6}$$

$$SBC = \ln(\sigma^2_{residuals}) + \frac{M}{n} \cdot \ln(n) \tag{7}$$

$$R^2_{adj} = R^2 - \left(1 - R^2\right) \cdot \left(\frac{M}{(n-M-1)}\right). \tag{8}$$

$\sigma^2_{Residuals}$ is the variance of the residuals, M is the number of predictor variables used in the model, and n is the number of observations.

2.3.5. Model Application

In order to find the optimal number of PLS components, we calculated PLS regressions with up to 20 components and computed the accuracy measures described in the last section with loo-cv for each model. The model with the highest accuracy according to these measures was chosen for estimating timber volume.

A range of *k*-NN estimates of timber volume was computed with the following parameterizations: all combinations of principal components that showed a highly significant correlation ($p < 0.001$) to timber volume were used as spectral auxiliary data; up to 10 nearest neighbors from the reference dataset were used. Weighting factors were computed inversely proportional to the Euclidean distance. The optimal number of *k* was selected with the condition of a certain decrease of more than 1 percent between RMSEs derived with *k* and *k* + 1 nearest neighbors [63]. Further optimization was done by applying band weighting factors. In our case, factors between 0.25 and 2 with increment steps of 0.25 were applied. The optimal model in terms of accuracy and parsimony was chosen with the computed accuracy measures described in the previous section.

2.3.6. Prediction Maps

Since estimation models are only established and valid for Norway spruce, all other land cover had been masked in the image data, based on an existing classification map for five main tree species in Rhineland-Palatinate [18]. The regression coefficients of the loo-cv derived optimal models were thus applied on a pixel-by-pixel basis to the pre-stratified images. In addition to the loo-cv, the prediction maps were validated by comparing the mapping results (*i.e.*, the average volume estimate within the administrative forest units) with the official FID information. To minimize the influence of border effects (roads, infrastructure, and other tree species), a 5-m buffer was applied along the stand perimeter [64]. The pixels containing training information were also removed for validation.

3. Results

3.1. Identification of Sensitive Spectral Regions

Before setting up regression models, the reference data were carefully analyzed. The relationship between the spectral response in the sensitive spectral regions and the FID-derived timber volume for Norway spruce reveals that the three stands may qualify as outliers. As an example, this is displayed for the red edge wavelength region (705 nm) in Figure 4a. Visual inspection of aerial photographs confirmed that the corresponding stands had been either cleared or substantially thinned shortly before the HyMap data were acquired; the forest database, however, had not yet been updated after this treatment. After removing these three data pairs, a set of 51 validated reference data pairs remained for further analysis.
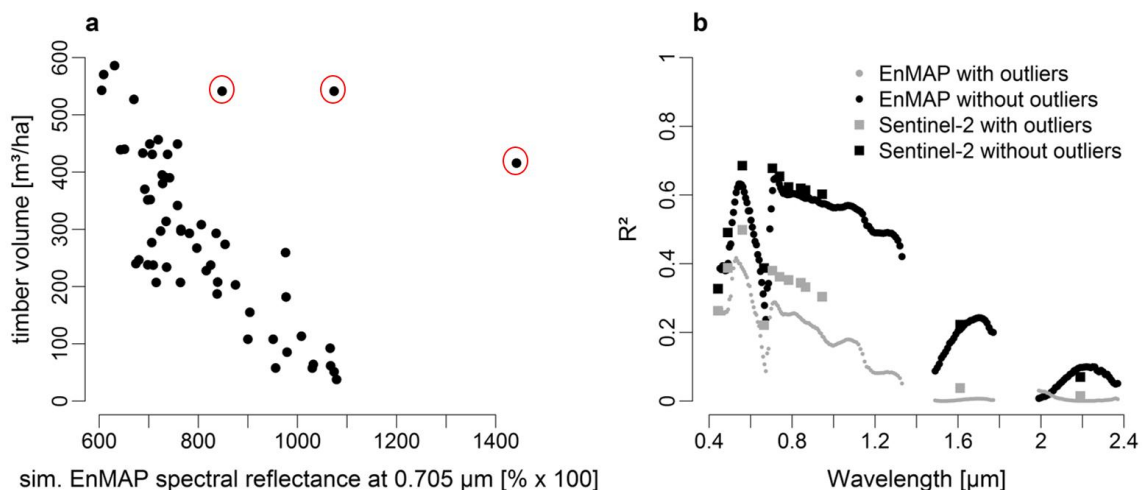
Figure 4. (**a**) Single band spectral response *vs.* timber volume, with identified outliers (red circles). (**b**) R²over the spectral range between 450 and 2500 nm for simulated EnMAP and Sentinel-2 data before (gray) and after (black) elimination of the outliers.

The elimination of these outliers substantially increased the coefficient of determination (R^2) for the relationship between stand reflectance and timber volume in almost all spectral bands; the highest R^2 values were found in the visible green, the red edge, and near-infrared spectral range (Figure 4b). From the near-infrared plateau onwards the coefficient of determination decreased continuously, such that in the SWIR spectral region only two local maxima with comparatively low R^2 are found.

3.2. Principal Component Analysis

For the simulated EnMAP and Sentinel-2 datasets, more than 99.5% of the overall variance is explained by the first three principal components (Figure 5). Five components from each dataset were found to have a high correlation with timber volume ($p < 0.001$); the maximum R^2 values were obtained for the first and third components of the simulated EnMAP spectra ($R^2_{PC1} = 0.69$, $R^2_{PC3} = 0.63$), while R^2 scored significantly lower for all remaining principal components ($R^2 < 0.37$). The Sentinel-2 simulation exhibits a similar pattern ($R^2_{PC1} = 0.68$, $R^2_{PC3} = 0.69$) while the other principal components with a significant correlation to timber volume had R^2 values smaller than 0.28.
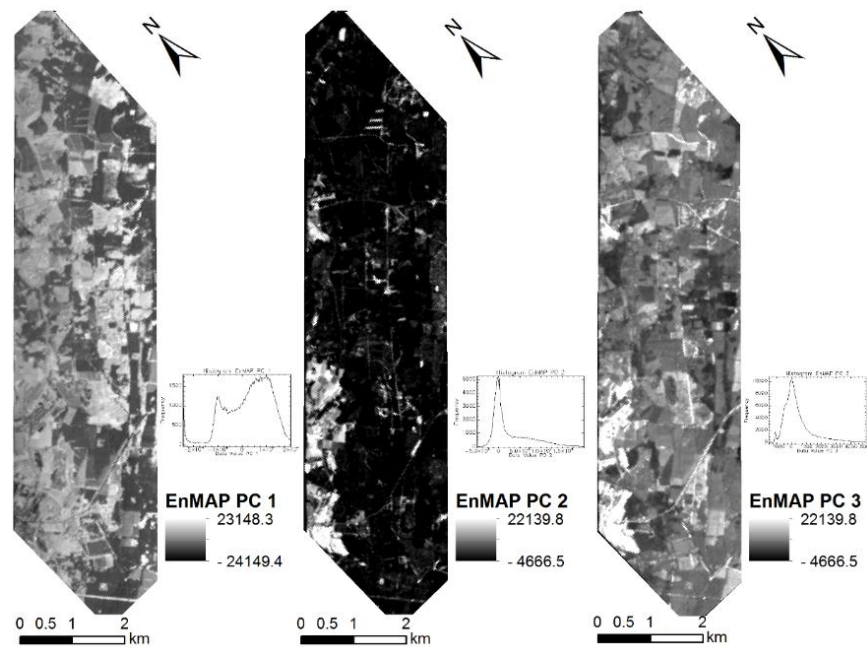
Figure 5. Images of the first three principal components of EnMAP data.

3.3. PLSR

RMSE$_{cv}$, AIC, and SBC plotted against the number of PLS components exhibit the same pattern; the best model is based on three PLS-components (EnMAP: RMSE$_{cv}$ = 28.19 %, equivalent to 79.97 m^3/ha; Sentinel-2: RMSE$_{cv}$ = 26.81 %, 80.88 m^3/ha). A higher number of components leads to a more complicated model with higher estimation errors (Figure 6). The Bias$_{cv}$ is negligible for EnMAP; for Sentinel-2 a small general underestimation was obtained (Table 1).
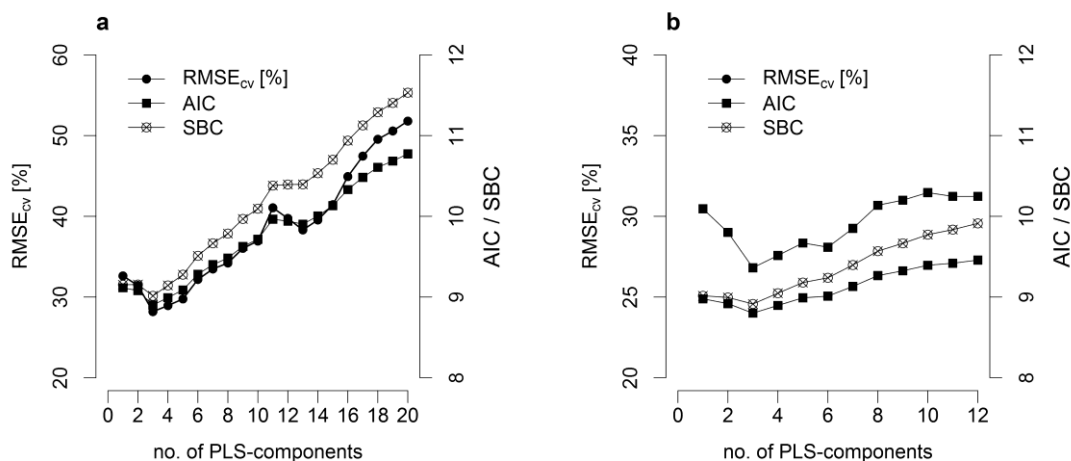
Figure 6. Loo-cv values of RMSE, AIC, and SBC for the PLSR with simulated EnMAP data (**a**) and Sentinel-2 data (**b**).

The PLS regression coefficients (Figure 7) confirm the results of the spectral sensitivity analysis and demonstrate that the most important wavelengths are in the red edge, the NIR, and green spectral regions. In the SWIR, the bands between 1500 and 1700 and around 2000 nm get the highest weights.
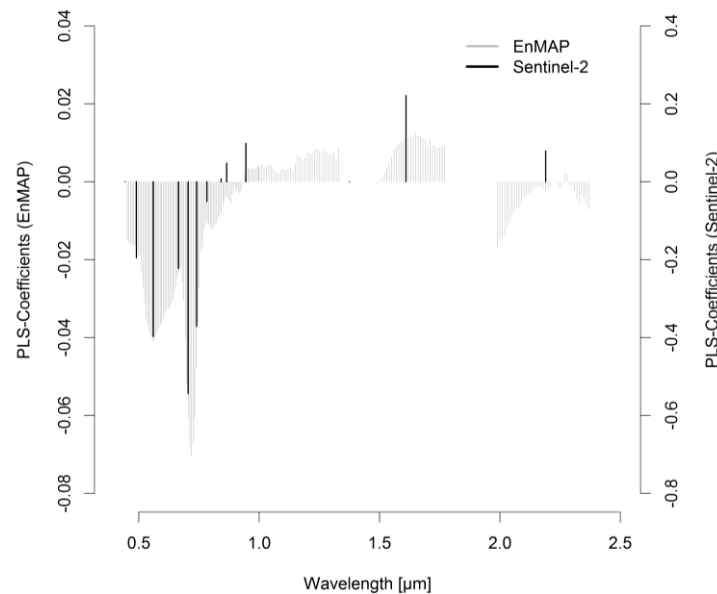
Figure 7. PLS regression coefficients for all EnMAP and Sentinel-2 spectral bands.

The validation of the mapping product produced somewhat smaller errors than loo-cv (EnMAP: RMSE = 26.5%, equivalent to 75.11 m³/ha; Sentinel-2: RMSE = 25.6%, 72.62 m³/ha). For both validations (loo-cv and map product validation), the RMSE for Sentinel-2 is slightly smaller than for EnMAP (Table 1), but not significantly. Figure 8 shows the comparison between observed and estimated timber volume for the available reference stands. Lower volume classes exhibit a marginal overestimation, while higher volume classes are underestimated. The highest scattering was observed in the volume ranges between 200 and 300 m³/ha. The largest mean deviation between observed and estimated values was found in the timber volume class of more than 500 m³/ha. This is true for both sensors. Sentinel-2 based predictions, however, show a smaller scattering.
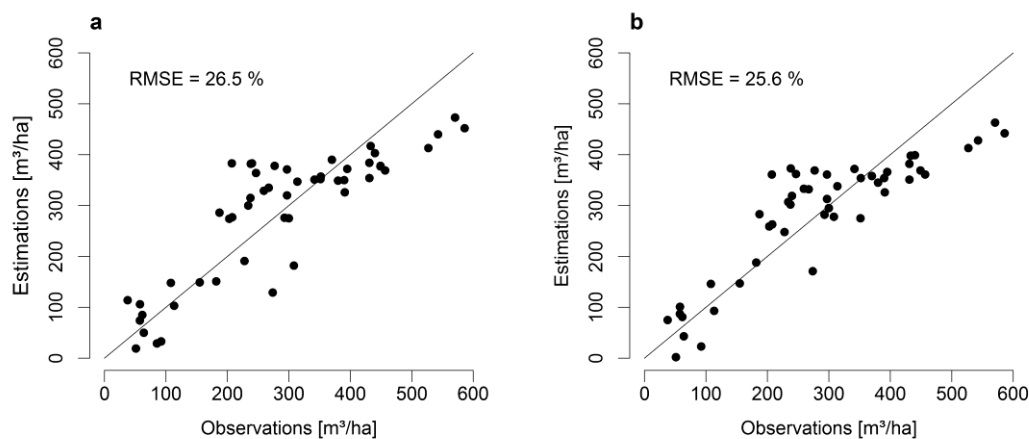

Figure 8. Estimated timber volume for EnMAP (**a**) and Sentinel-2 (**b**) in comparison to observed values on stand level, based on PLS-regression with all spectral bands.

Table 1. RMSE$_{cv}$, and Bias$_{cv}$ for three PLS-components. Timber volume estimation was based on all available spectral bands. Validation was done using loo-cv (training examples) and on administrative forest units (image-derived maps).

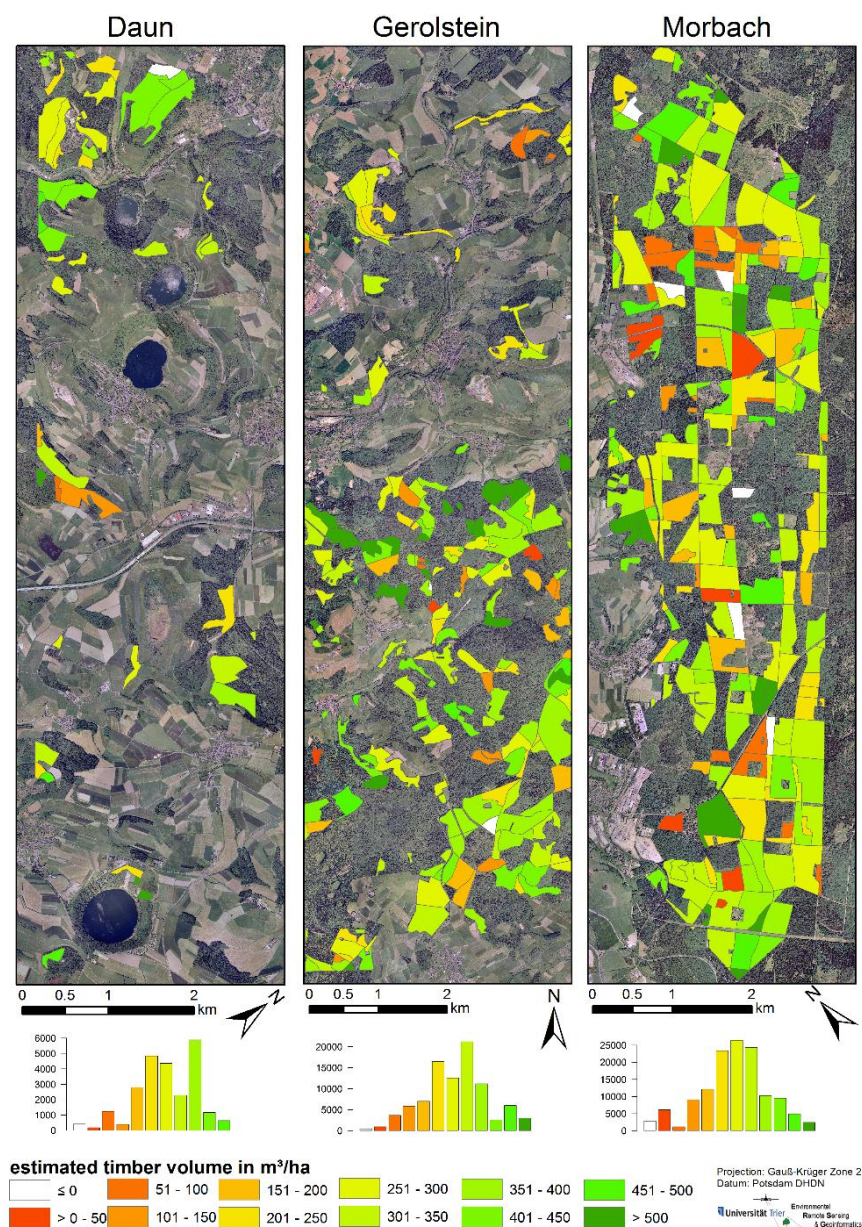|	Estimation						Reference
	loo-cv			Administrative Forest Unit			
	Mean Timber Volume (m³/ha)	Bias$_{cv}$ (m³/ha)	RMSE$_{cv}$ (%)	Mean Timber Volume (m³/ha)	Bias$_{cv}$ (m³/ha)	RMSE$_{cv}$ (%)	Mean Timber Volume (m³/ha)
PLSR (EnMAP)	283.61	−0.08	28.19	283.63	−0.06	26.48	283.69
PLSR (Sentinel-2)	282.39	−1.29	26.81	278.08	−5.61	25.60	

Figure 9. Timber volume distribution map based on FID data.

With respect to the spatialized FID and aerial photographs, it becomes obvious that the timber volume distribution in the image-derived maps (Figure 9) is largely consistent. In addition, the image-derived maps reveal various kinds of spatial heterogeneities, such as clear-cuts, storm damage, the access road network, and thinning-dependent volume gradients. Minor mapping errors occur in shadow areas along forest borders; moderately negative predictions only occur for pixels with exceptionally high reflectance (clear-cuts). Not surprisingly, predictions based on Sentinel-2 simulated data reveal more spatial detail, owing to the improved geometric resolution (Figure 10, Figure A1).
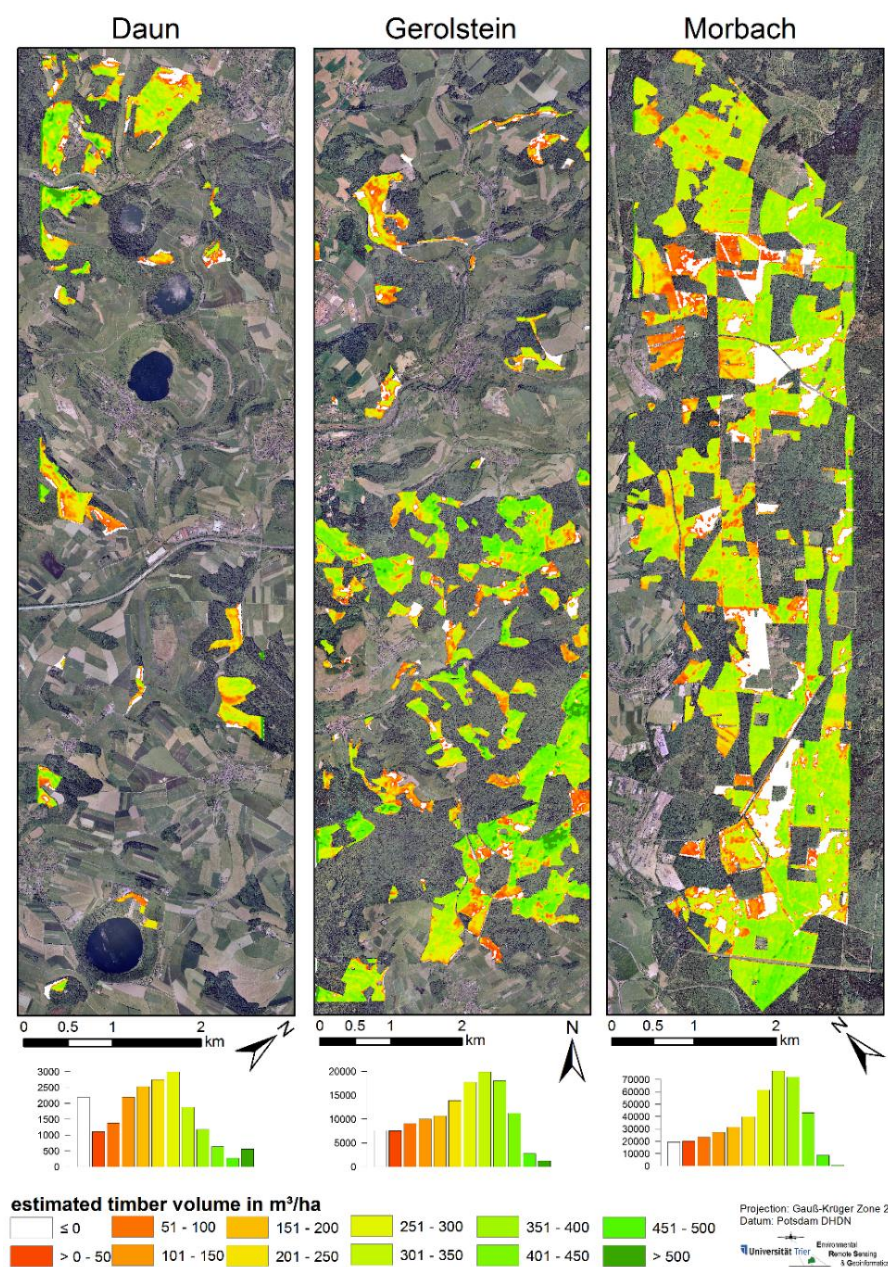

Figure 10. Timber volume prediction map based on PLSR with simulated Sentinel-2 data.

3.4. k-NN

A selection of feature combinations with the optimal number of *k* and resulting errors is presented in Table 2. Acceptable results were already obtained when using the two most sensitive principal

components (PC1 and PC3). Including additional PC features did not result in significantly improved results. Somewhat improved results were achieved when applying higher weighting factors for PC3.

Table 2. Parameterizations of EnMAP and Sentinel-2 *k*-NN model.

		Principal Component				No. of Nearest Neighbours (*k*)	RMSE$_{loo-cv}$ (%)	RMSE$_{loo-cv}$ (m³/ha)
EnMAP		PC 1	PC 2	PC 3	PC 7			
	R²	0.69	0.35	0.63	0.37			
		1				4	30.22	85.72
				1		4	41.41	117.49
	feature band weighting	1		1		3	28.49	80.81
		1	1	1		3	28.48	80.80
		1		1	1	3	28.43	80.65
		1	1	1	1	3	28.44	80.67
	optimized feature band weighting	0.25		2		3	25.70	72.93

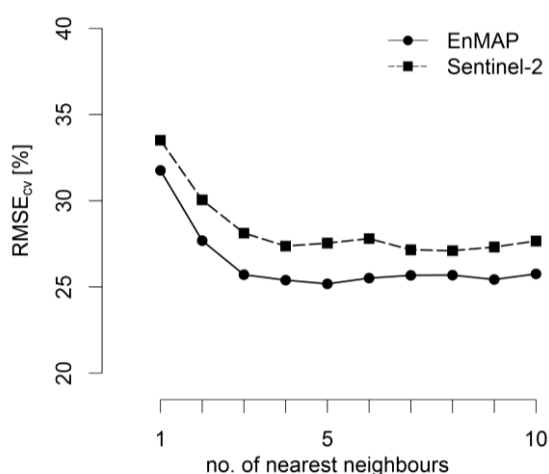|		Principal Component				No. of Nearest Neighbours (*k*)	RMSE$_{loo-cv}$ [%]	RMSE$_{loo-cv}$ [m³/ha]
Sentinel-2		PC 1	PC 3	PC 9	PC 10			
	R²	0.68	0.69	0.27	0.28			
		1				5	32.66	92.65
			1			3	37.27	105.73
	feature band weighting	1	1			4	28.40	80.56
		1	1	1		4	28.32	80.34
		1	1	1	1	4	28.18	79.93
	optimized feature band weighting	0.5	1.75			4	27.37	77.44

Figure 11. RMSEcv as a function of the number of neighbors for EnMAP and Sentinel-2 based optimized *k*-NN estimation, using PC1 and PC3.

Increasing the number of neighbors reduced the output dynamic range, leading to a constant, and with increasing *k* higher RMSE. It was therefore decided not to increase the number of neighbors beyond three (EnMAP) or four (Sentinel-2), respectively (Figure 11). This setup produced an RMSEcv of 25.7% (equivalent to 72.9 m³/ha) for EnMAP; the best model for the simulated Sentinel-2 data had a

slightly inferior performance (RMSE$_{cv}$ = 27.3%, 77.4 m³/ha). In both cases, the estimates exhibit only a very small bias (Table 2 and Table 3).

The validation of the mapping products (*i.e.*, obtained by averaging the estimated timber volume for each pixel included inside the management units) produced lower errors than the loo-cv (RMSE of 23.1% (equivalent to 65.5 m³/ha) for EnMAP and 21.6% (61.3 m³/ha) for Sentinel-2). In comparison to the PLSR results, very high timber volumes are better estimated by the *k*-NN approach, where estimations are less affected by saturation effects in high timber volume classes. Substantial scattering occurs in the medium volume range, especially for the EnMAP-based map (Figure 12).
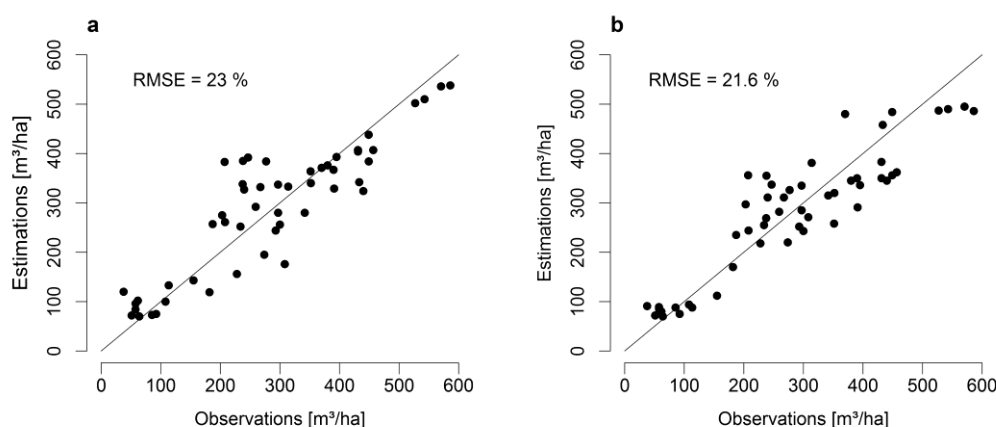

Figure 12. *k*-NN-based estimations in comparison to observed values for EnMAP (**a**) and Sentinel-2 (**b**) on the level of administrative forest units.

Table 3. RMSE$_{cv}$, and bias for the optimized k-NN model, based on PC1 and PC3. Validation with loo-cv and for administrative forest units.

	LOO-cv			Administrative Forest Unit			Reference
	Mean Timber Volume (estim.) (m³/ha)	**Bias$_{cv}$ (m³/ha)**	**RMSE$_{cv}$ (%)**	**Mean Timber Volume (estim.) (m³/ha)**	**Bias$_{cv}$ (m³/ha)**	**RMSE$_{cv}$ (%)**	**Mean Timber Volume (m³/ha)**
k-NN (EnMAP); *k* = 3	280.16	−3.53	25.7	287.33	3.64	23.11	283.69
k-NN (Sentinel-2); *k* = 4	279.47	−4.22	27.37	280.16	−3.53	21.58	

Again, the difference between the mapping results for the two simulated sensors is statistically not significant and becomes evident rather on the level of spatial detail.

Differences in forest structure and in timber volume are well represented in the prediction maps derived from the *k*-NN approach. While PLSR is able to extrapolate, the *k*-NN-derived maps by definition do not include any estimates outside the value range represented in the reference. Like the PLS mapping results, timber volume distribution maps are largely consistent with ground information, but also exhibit important within-stand spatial heterogeneities. The data range of the *k*-NN derived mapping product cannot exceed the data range of the training examples. This creates problems for very

dark areas (e.g., shadow at borders of forest stands) and very bright areas (e.g., clear-cuts), where volumes are assigned although no timber is present (Figures A2 and A3).

4. Discussion

4.1. Sensitive Spectral Ranges

Regression analysis of the relationship between site-specific timber volume and spectral reflectance has revealed the existence of particularly strong (negative) correlations in the spectral range between 520 and 1300 nm; this relationship only breaks down within the red section of the electromagnetic spectrum (~650–680 nm), where the sensitivity to structural differences in spruce canopies is diminished, owing to the high degree of opacity within the chlorophyll absorption well [65]. A weaker, but nevertheless significant, relationship between canopy reflectance (as a proxy for stand density and, consequently, timber volume) also exists in the SWIR around 1600 nm and, though much weaker, around 2300 nm. While Ardö [66] had identified the SWIR region around 1600 nm as the most sensitive spectral range when trying to map timber volume based on Landsat-TM data in Scandinavian conifer forests, in our work the highest sensitivity occurs in the NIR region, which conforms to other studies (e.g., [67,68]). The maximum correlation is in fact observed along the red edge, which will of course be covered by the EnMAP hyperspectral imager, but is also observed by Sentinel-2. It is thus not surprising that the results in mapping spruce timber volume are not substantially different with respect to both systems, apart from the spatial detail.

The fact that the correlation between timber volume and spectral reflectance is consistently negative (*i.e.*, decreases with increasing volume) suggests two points as driving factor. On the one side, (multiple) scattering effects exist inside the vegetated crowns (between and within the shoots and the branches of the vegetated crown), which affect the reflectance signal [69]. Consequently, more voluminous mature trees involving a higher number of needle-layers lead to more absorption processes and less overall canopy reflectance [70]. Furthermore, under the conditions of managed forests (as in the study region), structure-dependent shadow effects in line with development stages (establishment, qualification, dimensioning, and maturity), mainly driven by selective thinning and logging, affect the reflectance signal [67]. During establishment, young spruce trees are closely packed and tree height is quite uniform. Owing to this and to a very high degree of canopy closure, no significant shadow fraction will influence the spectral signal in this early growth stage. However, due to competitive growth dynamics, the crown layer will increase in roughness and heterogeneity and thereby produce more shaded parts. Management interventions, such as selective thinning, will create small gaps in the canopy to provide better growing conditions for the qualified trees; the higher amount of shadow, together with increasingly opaque needles, will contribute to further decreasing reflectance. In line with continued management activities (logging of mature trees), canopy closure further decreases while the amount of shaded areas increases. During the final stage of stand development and management, the proportion of shaded gaps and canopy section will saturate, such that the correlation to timber volume is expected to become insensitive (Figure 13).
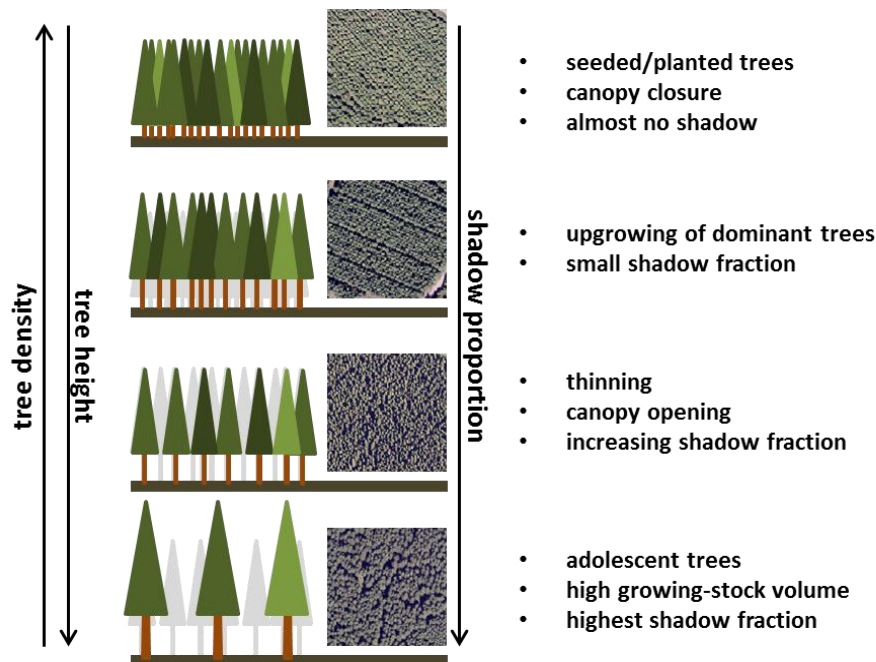
Figure 13. Structural differences and their appearance in aerial photos during the development stages of Norway spruce forest stands.

Intensive management activities in combination with slow updating cycles of the forest information data base are a major reason for the weaker relationship between timber volume and spectral response in the dimensioning stage of development [71]. Some forest stands in this particular volume class are already thinned, presenting a higher proportion of shadow and lower reflectance values; less thinned forest stands within the same development stage but comparable timber volume present a higher spectral signal due to more canopy closure. This effect weakens the correlation between timber volume and spectral information, as already described in an earlier study [48].

4.2. Estimation Models and Sensors

Designed for dealing with high dimensional, collinear data, PLSR has once more proven its suitability for estimating timber volume based on hyperspectral data [61]. The choice of an adequately low number of PLS components keeps the models simple and sufficiently parsimonious, and thereby avoids the problem of overfitting, which is a core requirement for maintaining sufficient generalization capacity to map timber volume over extended image datasets [51,62,72]. Different parameterization strategies with pre-selected spectral bands (e.g., only those with a high sensitivity to the target variable) did not result in better results; this underlines the strength of the PLSR to deal with high dimensional, collinear, and partially noisy spectral data [73], without the necessity of applying data pre-processing and conversion steps.

In contrast to the PLSR, the non-parametric *k*-NN method requires certain data pre-processing strategies prior to the analysis. Principal component analysis has been employed as a probate means for reducing the data volume and eliminating redundant information [74]. The results concerning the model parameterization are consistent with other studies, which showed that including too many (partially correlated) information layers may even lead to inferior results [75]. It could thus also be

confirmed in this study that overly complicated models with a high number of predictor variables were less efficient and did not provide significantly better estimations. Taking only the loo-cv based RMSE or the bias for a larger area into consideration, the resulting optimal number of nearest neighbors may be quite high. Generally, the full range of reference data can only be covered with $k = 1$, and the data range in the map result will decrease with a higher number of k. Moreover, too high values of k lead to a leverage effect in low and high ranges of timber volume, where estimations are assigned a value closer to the mean of the reference data [24]. However, it is difficult to generalize these findings as k-NN may be expected to improve when applied over larger areas with a higher number of reference samples (which might also favor using a higher number of neighbors). In consequence, especially for higher numbers of k, a large and continuous reference database is essential in order to nearly get the full data range in the output. Regarding the limited number of training examples in our study area, this number should be adapted to the number of training examples in the reference dataset.

4.3. Prediction Maps

With respect to the requirements of operational forest management activities, the achieved mapping results are considered acceptable. Their accuracy is within the range of terrestrial inventories on stand level. Severe deviations between both methods as well as both sensors have not been identified. A comparison of the prediction maps indicates that most differences are less than +/− 100 m³/ha. The data ranges presented in the maps are different between the two methods. A property of the PLSR is extrapolation, which might be advantageous in the case of larger gaps in training examples. One drawback is the possibility of timber volume estimations beyond the reference data's range. However, the occurrence of such cases is limited because the stratification ensures that the models are applied to Norway spruce only. In contrast, k-NN based estimation values by definition stay in the range of the reference data [76], but the model precision may be weaker in the case of larger gaps in the training examples, especially if the number of k is small. Substantial differences between the PLSR and k-NN maps only occur in those areas where the PLSR-based estimations are extrapolated beyond the reference data range. This can be seen for pixels with high reflectance values, e.g. clear cut areas, sunlit road network, or even parts of recently established young Norway spruce stands, while high overestimation occurs for areas with a very low reflectance, like aggregations of pixels with a high proportion of shadow.

The very high spectral resolution of EnMAP did not result in significantly better prediction maps, while EnMAP's 30 × 30 m² pixel size is less capable of capturing small spatial structures and associated differences in timber volume.

Because the Sentinel-2 spectral reference information has been sampled in a 3 × 3 kernel, it may be advisable to apply a corresponding low pass filter before calculating the prediction map. Border effects, which preferentially occur near forest unit boundaries or infrastructure, leading to very high or very low estimation values, can be minimized by applying larger buffers around the forest units [64].

4.4. Forest Reference Information

Feasibility studies such as the one we present here require sufficiently reliable reference information. This is problematic because appropriate data on timber volumes are either collected over

small plots in the context of large comprehensive statistical sampling frames such as the FSFI (which may limit the availability of data due to the spatial coverage of experimental airborne data), or they represent expert assessments for complete stands or management units, made available through the continuously updated Forest Inventory Database (FID). The latter may increase the number of reference data, but are usually considered less precise. However, a comparison between corresponding locations (where both types of inventory information were available) suggests that the FID is consistent with the corresponding FSFI plot measurements (Figure 3). It was thus justifiable to base this study on the FID data, which provided sufficient reference data.

It was confirmed that the FID-derived inventory information can in principle be combined with remote sensing imagery to produce useful timber volume distribution maps for Norway spruce. However, it should be considered that in our study area the majority of forest units (85.25%) had been inventoried within two years before or after image acquisition. In less favorable cases, it may occur that the inventory data are substantially outdated, owing to the relatively slow updating cycles of five to 10 years.

In addition, because FID volume values are not measured but estimated in the field and updated based on yield tables, due to substantially changed growth conditions (*i.e.*, increased nutrient deposition and climate change effects [77,78]) they tend to underestimate today's stronger forest growth. Additional variability may be introduced because of the individual perceptions of the inventory specialists carrying out the surveys.

5. Conclusions

To some extent unexpectedly (as earlier research on coniferous trees had emphasized the importance of the SWIR spectral range), our work has revealed that the NIR spectral bands, particularly those along the red edge, were most sensitive to the structural characteristics of Norway spruce stands. The fact that Sentinel-2 also provides good coverage of this spectral range is obviously one of the major reasons why both systems demonstrated almost identical timber volume mapping capacities, independently of the total number of bands.

The achieved mapping accuracies (RMSE $<$ 80 m³/ha) appear moderate, but are nevertheless considered acceptable for planning and supporting the implementation of management interventions. In comparison, FSFI data are not useful on the local level (due to the data acquisition method and a small number of training examples), and the FID holds information that is based on expert assessments instead of systematic measurements, usually up-to-date within a time span of 5–10 years only. It is difficult to provide an unbiased characterization of their accuracy; based on the comparison between timber volume information from the FSFI (Bitterlich sampling at plot level) and FID (stand-wise expert assessment), it appears that an error in the range of 90 m³/ha is inherent to the field data collection methods. In comparison to this finding, it is concluded that the accuracy of the image-derived maps is acceptable, especially because these mapping results can be provided at high spatial detail and over extended areas.

The differences in estimation and mapping accuracies provided by PLSR and *k*-NN are negligible and do not suggest that one of the methods is truly superior to the other. PLSR has the substantial advantage that it will not require any pre-selection of spectral bands and is more robust towards data

gaps, while efficient data reduction is important before applying the k-NN approach. Additionally, the number of k should be carefully chosen by considering the number of training examples and possible gaps this data. However, the fact that k-NN does not extrapolate and thus does not lead to negative or excessively high estimates can be seen as an advantage when the training data comprise the full data range of the study area. In this study, k-NN seems to have an advantage in dealing with saturation effects at high volume levels while PLSR shows some saturation effects. Moreover, it is expected to develop its full strength (*i.e.*, adaptive capacity) when larger areas with more reference data can be processed. Regardless of the chosen approach, it is evident that the intermediate volume range (*i.e.*, 200–350 m³) is affected by larger estimation errors than low and high values of timber volume. This is due to the fact that management interventions with a strong impact on canopy reflectance properties (*i.e.*, selective thinning operations) are conducted with varying intensity and across a relatively large period of stand development, which will increase the spectral variability of stands with similar timber volumes.

Owing to its improved spatial resolution, larger area coverage, and the more frequent acquisition opportunities, we expect that Sentinel-2 might be to some extent preferred for producing timber volume maps in areas with small-structured management units. However, the compatibility between timber volume maps produced by EnMAP and Sentinel-2 largely facilitates an integration of both products without the need to develop individual processing concepts. Due to its hyperspectral coverage, EnMAP is expected to contribute complementary information on the biophysical condition of Norway spruce, such as indicators of increasing drought stress and insect calamities (e.g., bark beetle attacks) [79], which are expected to increase because of regional warming trends [80,81]. Since both information levels are important for optimizing economically viable and ecologically balanced management strategies, we expect that Sentinel-2 and EnMAP, once both are operational, can be used in synergy.

Acknowledgments

Author Contributions

All authors contributed significantly to the work presented in this paper. Sascha Nink was the primary analyst and main author of this work. Joachim Hill did the simulation of datasets and supervised this research and the writing of the paper; Henning Buddenbaum and Johannes Stoffels analyzed the results and contributed significantly to the discussion; Thomas Sachtleber and Joachim Langshausen contributed auxiliary data (classification and forest information data).
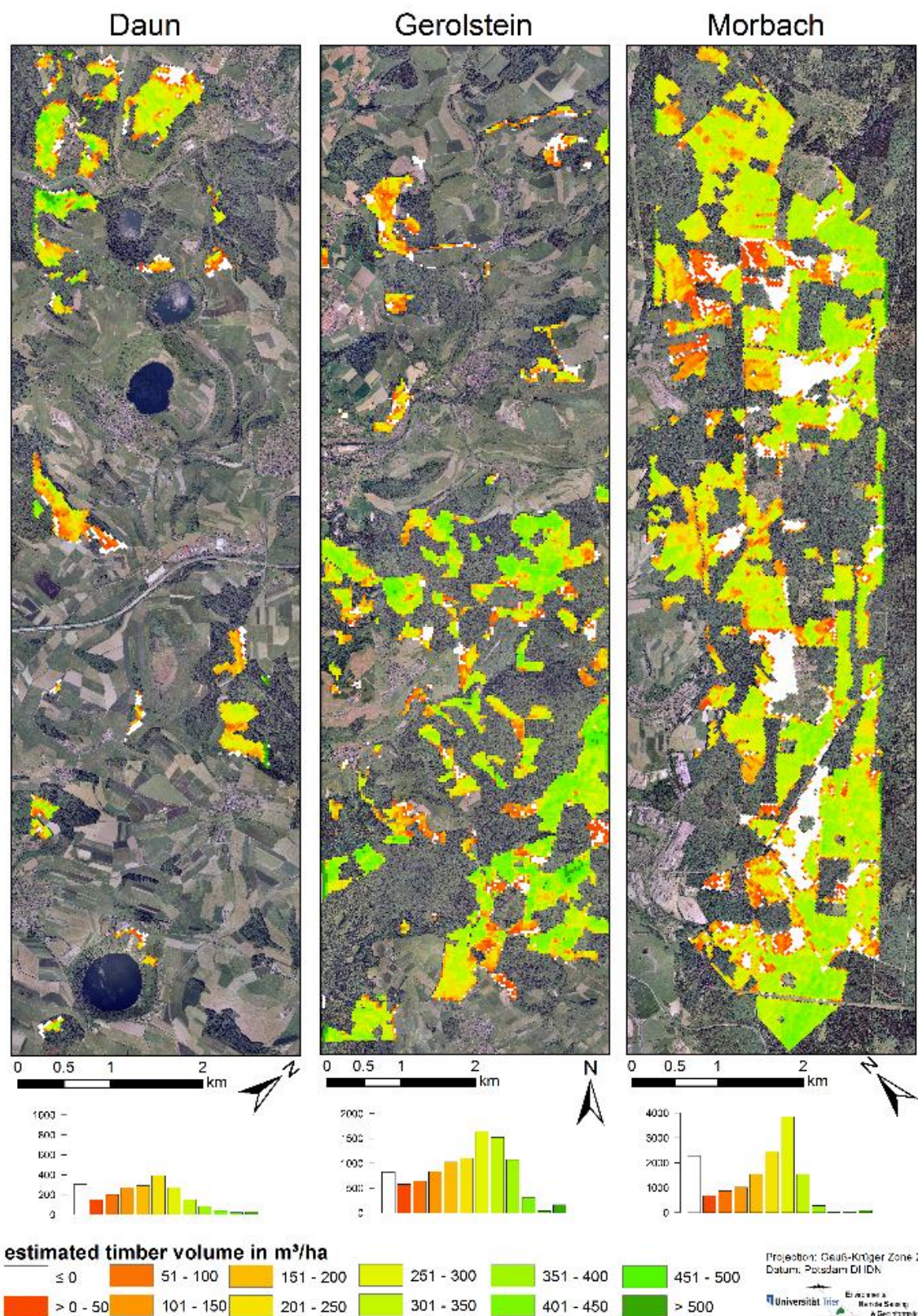
Appendix

Figure A1. Timber volume prediction map based on PLSR with simulated EnMAP data.
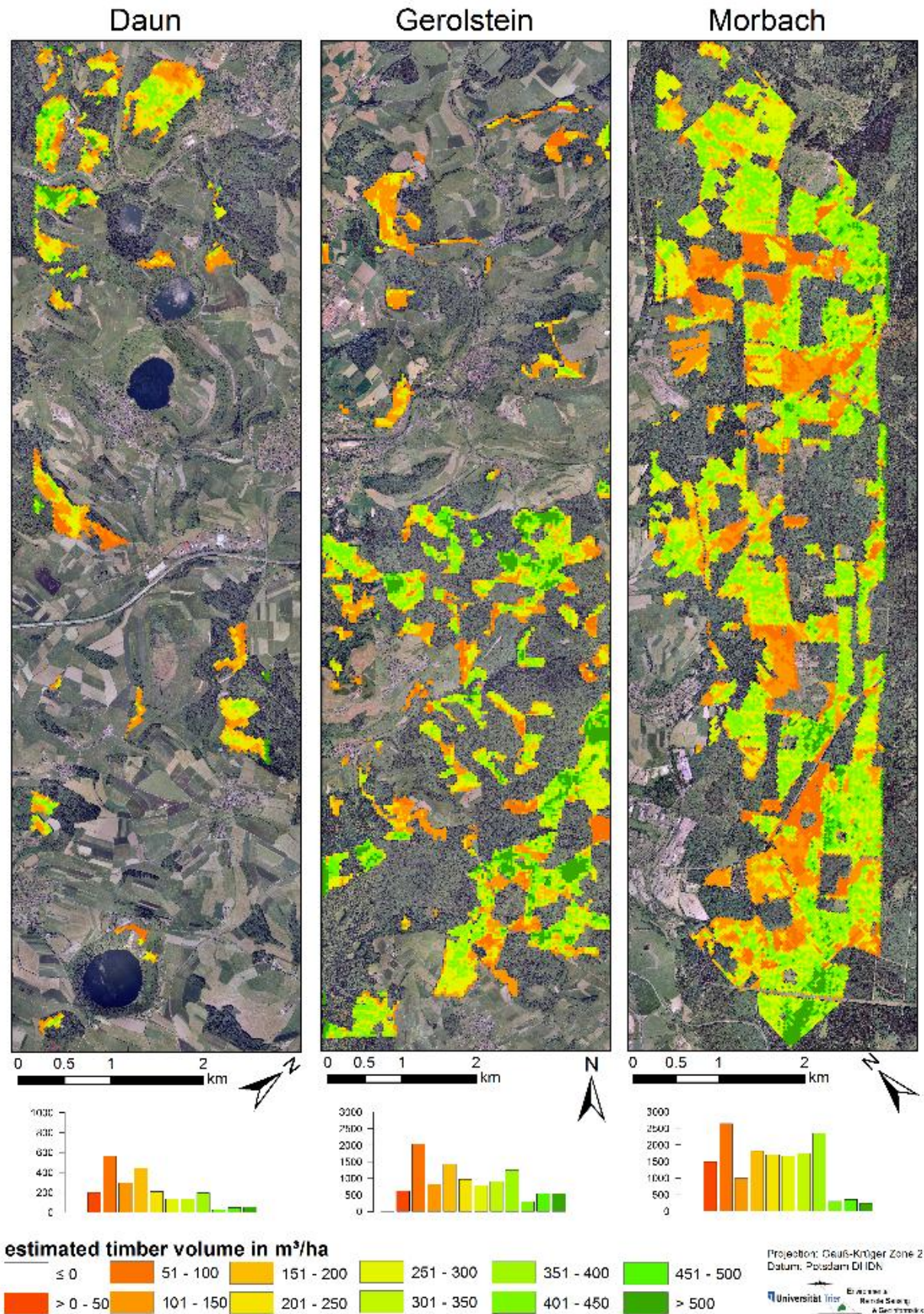
Figure A2. Timber volume prediction map based on *k*-NN with simulated EnMAP data.
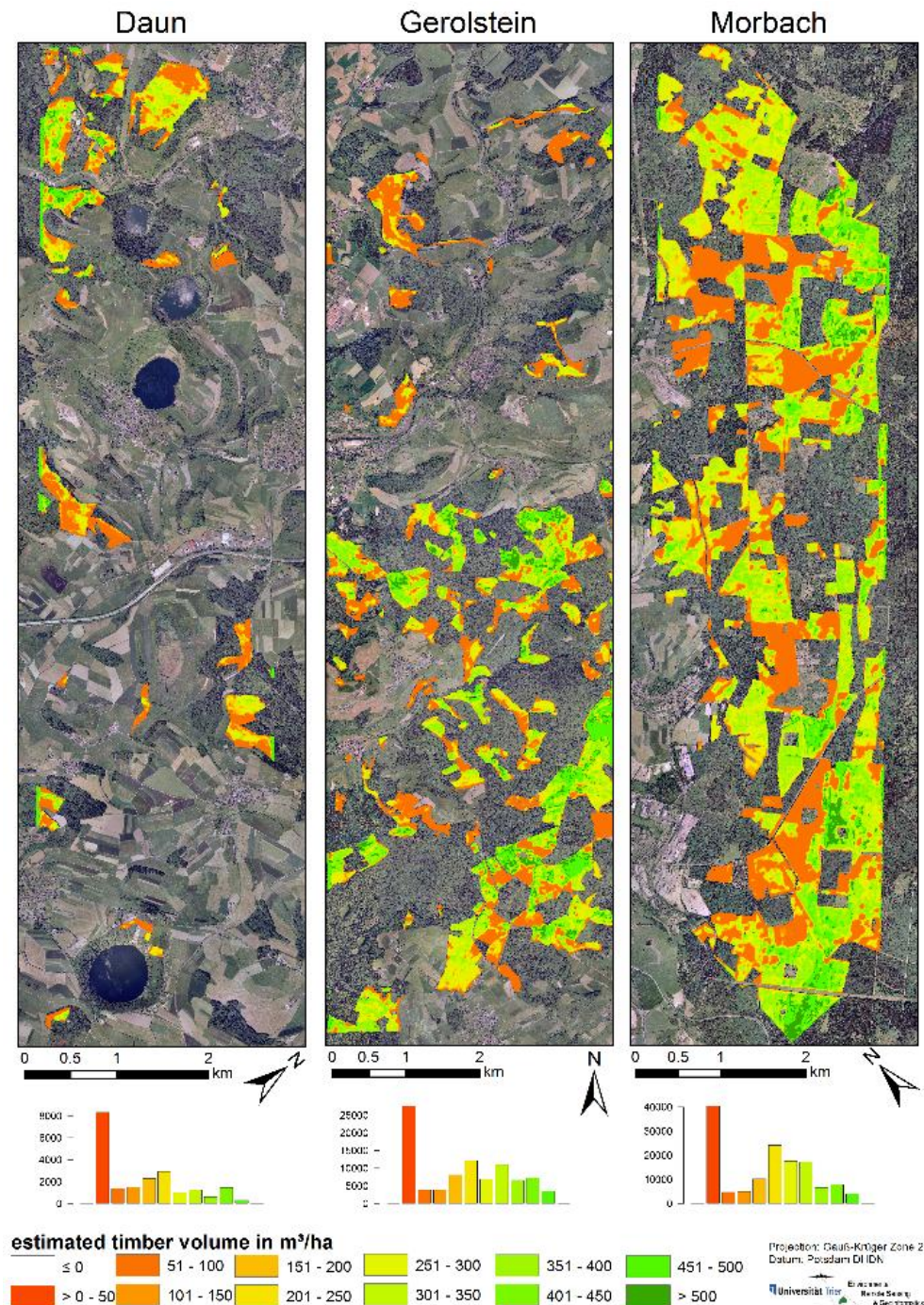
Figure A3. Timber volume prediction map based on *k*-NN with simulated Sentinel-2 data.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Foley, J.A.; DeFries, R.; Asner, G.P.; Barford, C.; Bonan, G.; Carpenter, S.R.; Chapin, F.S.; Coe, M.T.; Daily, G.C.; Gibbs, H.K.; *et al.* Global consequences of land use. *Science* **2005**, 570–574.

2. Bond, I.; Grieg-Gran, M.; Wertz-Kanounnikoff, S.; Hazlewood, P.; Wunder, S.; Angelsen, A. *Incentives to Sustain Forest Ecosystem Services*: *A Review and Lessons for REDD*; International Institute for Environment and Development: London, UK, 2009.

3. Food and Agriculture Organization of the United Nations. *State of the World's Forests*; FAO: Rome, France, 2012.

4. Federal Ministry of Food, Agriculture and Consumer Protection. German Forests—Nature and Economic Factor. Available online: http://www.bmel.de/SharedDocs/Downloads/EN/ Publications/GermanForests.pdf?_blob=publicationFile (accessed on 13 June 2014).

5. Federal Ministry of Food, Agriculture and Consumer Protection. Forest Strategy 2020: Sustainable Forest Management—An Opportunity and a Challenge for Society. Available online: http://www.bmel.de/SharedDocs/Downloads/EN/Publications/ForestStrategy2020.pdf?__blob=pu blicationFile (accessed on 13 June 2014).

6. Bolte, A.; Eisenhauer, D.R.; Erhart, H.P.; Grpß, J.; Hanewinkel, M.; Kölling, C.; Profft, I.; Rohde, M.; Röhe, P.; Amereller, K. Klimawandel und Forstwirtschaft—Übereinstimmungen und Unterschiede bei der Einschätzung der Anpassungsnotwendigkeiten und Anpassungsstrategien der Bundesländer. *Landbauforsch. vTI Agric. Forest Res.* **2009**, *59*, 267–278.

7. Standing Forestry Committee. Ad Hoc Working Group on Forest Information and Monitoring: Final Report. Available online: http://ec.europa.eu/environment/forests/pdf/Fin%20report%20 info%20monit%20wg.pdf (accessed on 13 June 2014).

8. Peerenboom, H.G.; Ontrup, G.; Böhmer, O. Weiterentwicklung der Forsteinrichtung in Rheinland-Pfalz. *Forst und Holz* **2003**, *58*, 728–731.

9. Köhl, M. New approaches for multi resource forest inventories. In *Advances in Forest Inventory for Sustainable Forest Management and Biodiversity Monitoring*; Corona, P., Köhl, M., Marchetti, M., Eds.; Springer Netherland: Dordrecht, The Netherlands, 2003.

10. Potapov, P.V.; Yaroshenko, A.; Turubanova, S.; Dubinin, M.; Laestadius, L.; Thies, C.; Aksenov, D.; Egorov, A.; Yesipova, Y.; Glushkov, I.; *et al.* Mapping the world's intact forest landscapes by remote sensing. Available online: https://dlc.dlib.indiana.edu/dlc/bitstream/handle/10535/2817/ ES-2008-2670.pdf?sequence=1&isAllowed=y (accessed on 28 July 2015).

11. Hansen, M.C.; Stehman, S.V.; Potapov, P.V. Quantification of global gross forest cover loss. *Proc. Natl. Acad. Sci.* **2010**, *107*, 8650–8655.

12. Hansen, M.C.; Potapov, P.V.; Moore, R.; Hancher, M.; Turubanova, S.A.; Tyukavina, A.; Thau, D.; Stehman, S.V.; Goetz, S.J.; Loveland, T.R.; *et al.* High-resolution global maps of 21st-century forest cover change. *Science* **2013**, *342*, 850–853.

13. Food and Agriculture Organisation of the United Nations & European Commission Joint Research Centre. *Global Forest Land-Use Change: 1990–2005*; FAO & JRC: Rome, France, 2012.

14. Le Toan, T.; Quegan, S.; Davidson, M.; Balzter, H.; Paillou, P.; Papathanassiou, K.; Plummer, S.; Rocca, F.; Saatchi, S.; Shugart, H.; *et al.* The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sens. Environ.* **2011**, *115*, 2850–2860.

15. Rosenqvist, Å.; Milne, A.; Lucas, R.; Imhoff, M.; Dobson, C. A review of remote sensing technology in support of the Kyoto Protocol. *Environ. Sci. Policy* **2003**, *6*, 441–455.

16. Martin, M.E.; Newman, S.D.; Aber, J.D.; Congalton, R.G. Determining forest species composition using high spectral resolution remote sensing data. *Remote Sens. Environ.* **1998**, *65*, 249–254.

17. Schlerf, M.; Atzberger, C.; Hill, J. Remote sensing of forest biophysical variables using HyMap imaging spectrometer data. *Remote Sens. Environ.* **2005**, *95*, 177–194.

18. Stoffels, J.; Mader, S.; Hill, J.; Werner, W.; Ontrup, G. Satellite-based stand-wise forest cover type mapping using a spatially adaptive classification approach. *Eur. J. Forest Res.* **2012**, *131*, 1071–1089.

19. Franklin, S.E. *Remote Sensing for Sustainable Forest Management*; CRC Press: Boca Raton, FL, USA, 2001.

20. Tomppo, E.; Katila, M.; Mäkisara, K.; Peräsaari, J. The multi-source national forest inventory of Finland—Methods and results 2007. Available online: http://www.metla.fi/julkaisut/ workingpapers/2012/mwp227.pdf (accessed on 28 July 2015).

21. Reese, H.; Nilsson, M.; Granqvist Pahlén, T.; Hagner, O.; Tingelöf, U.; Egberth, M.; Olsson, H. Countrywide estimates of forest variables using satellite data and field data from the National Forest Inventory. *AMBIO* **2003**, *32*, 542–548.

22. Nilsson, M.; Bohlin, J.; Olsson, H.; Svensson, S.A.; Haapaniemi, M. Operational use of remote sensing for regional level assessment of forest estate values. In *New Strategies for European Remote Sensing*, *Proceedings of the 24th Symposium of the European Association of Remote Sensing Laboratories, Dubrovnik, Croatia, 25–27 May 2004*; Oluić, M., Ed.; Mill Press: Dubrovnik, Croatia, 2005; pp. 263–268.

23. Gjertsen, A. Accuracy of forest mapping based on Landsat TM data and a kNN-based method. *Remote Sens. Environ.* **2007**, *110*, 420–430.

24. McInerney, D.O.; Nieuwenhuis, M. A comparative analysis of *k*NN and decision tree methods for the Irish National Forest Inventory. *Int. J. Remote Sens.* **2009**, *30*, 4937–4955.

25. Koukal, T. Nonparametric Assessment of Forest Attributes by Combination of Field Data of the Austrian Forest Inventory and Remote Sensing Data. Available online: http://citeseerx.ist.psu.edu/ viewdoc/download?doi=10.1.1.88.5946&rep=rep1&type=pdf (accessed on 28 July 2015).

26. Maselli, F.; Chirici, G.; Bottai, L.; Corona, P.; Marchetti, M. Estimation of Mediterranean forest attributes by the application of k-NN procedures to multitemporal Landsat ETM+ images. *Int. J. Remote Sens.* **2005**, *26*, 3781–3796.

27. Dees, M.; Duvenhorst, J.; Gross, C.P.; Koch, B. Combining Remote Sensing Data Sources and Terrestrial Sample-based Inventory Data for the Use in Forest Management Inventories. Available online: http://www.isprs.org/proceedings/XXXIII/congress/part7/355_XXXIII-part7.pdf (accessed on 28 July 2015).

28. Diemer, C.; Lucaschewsky, I.; Spelsberg, G.; Tomppo, E.; Pekkarinen, A. Integration of terrestrial forest sample plot data, map information and satellite data: An operational multisource inventory concept. In *Fusion of Earth Data: Merging Point Measurements, Raster Maps and Remotely Sensed Images*; Ranchin, T., Wald, L., Eds.; SEE/URISCA: Nice, France, 2000, pp. 143–150.

29. Scheuber, M. Potentials and limits of the k-nearest-neighbour method for regionalising sample-based data in forestry. *Eur. J. Forest Res.* **2009**, *129*, 825–832.

30. Lu, D. The potential and challenge of remote sensing-based biomass estimation. *Int. J. Remote Sens.* **2006**, *27*, 1297–1328.

31. Gleason, C.J.; Im, J. A review of remote sensing of forest biomass and biofuel: Options for small-area applications. *GISci. Remote Sens.* **2011**, *48*, 141–170.

32. Koch, B. Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 581–590.

33. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; *et al.* Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* **2012**, *120*, 25–36.

34. Malenovský, Z.; Rott, H.; Cihlar, J.; Schaepman, M.E.; García-Santos, G.; Fernandes, R.; Berger, M. Sentinels for science: Potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land. *Remote Sens. Environ.* **2012**, *120*, 91–101.

35. Kaufmann, H.; Hill, J.; Hostert, P.; Krasemann, H.; Mauser, W.; Müller, A. Science Plan of the Environmental Mapping and Analysis Program (EnMAP). Available online: http://www.enmap. org/ (accessed on 28 July 2015).

36. Wolff, B.; Erhard, M.; Holzhausen, M.; Kuhlow, T. *Das Klima in den forstlichen Wuchsgebieten und Wuchsbezirken Deutschlands*; Wiedebusch: Hamburg, Germany, 2003.

37. Cocks, T.; Jenssen, R.; Stewart, A.; Wilson, I.; Shields, T. The HyMap TM airborne hyperspectral sensor: The system, calibration and performance. In Proceedings of 1st EARSeL Workshop on Imaging Spectroscopy, EARSeL Paris, France, 6–8 October 1998; pp. 37–42.

38. Schläpfer, D.; Schaepman, M.E.; Itten, K. PARGE: parametric geocoding based on GCP-calibrated auxiliary data. *Proc. SPIE* **1998**, doi:10.1117/12.328114.

39. Schläpfer, D.; Richter, R. Geo-atmospheric processing of airborne imaging spectrometry data: Part 1: parametric orthorectification. *Int. J. Remote Sens.* **2002**, *23*, 2609–2630.

40. Buddenbaum, H.; Schlerf, M.; Hill, J. Classification of coniferous tree species and age classes using hyperspectral data and geostatistical methods. *Int. J. Remote Sens.* **2005**, *26*, 5453–5465.

41. Hill, J.; Mehl, W.; Radeloff, V. Improved forest mapping by combining corrections of atmospheric and topographic effects in Landsat TM imagery. In Proceedings of the 14th Sensors and environmental applications of remote sensing EARSel Symposium, Goïeborg, Sweden, 6–8 June 1994.

42. Hill, J.; Mehl, W. Geo- und radiometrische Aufbereitung multi- und hyperspektraler Daten zur Erzeugung langjähriger kalibrierter Zeitreihen. *Photogrammetrie Fernerkundung Geoinf.* **2003**, *7*, 7–14.

43. Tanré, D. Description of a computer code to simulate signal in the solar spectrum: The 5S code. *Int. J. Remote Sens.* **1990**, doi:10.1080/01431169008955048.

44. Buddenbaum, H.; Seeling, S.; Hill, J. Fusion of full-waveform Lidar and imaging spectroscopy remote sensing data for the characterization of forest stands. *Int. J. Remote Sens.* **2013**, *34*, 4511–4524.

45. Guanter, L.; Segl, K.; Kaufmann, H. Simulation of optical remote-sensing scenes with application to the EnMAP hyperspectral mission. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 2340–2351.

46. Martimort, P.; Berger, M.; Carnicero, B.; Del Bello, U.; Fernandez, V.; Gascon, F.; Silvestrin, P.; Spoto, F.; Sy, O.; Arino, O.; *et al.* Sentinel-2: The optical high-resolution mission for GMES operational services. *Eur. Space Agency Bull.* **2007**, 18–24.

47. Landesforsten Rheinland-Pfalz. Available online: http://www.wald-rlp.de/ueber-uns/ nachhaltigkeit/sicherung-der-nachhaltigkeit/mittelfristigebetriebsplanungforsteinrichtung/ inventur.html (accessed on 23 May 2014).

48. Trotter, C.M.; Dymond, J.R.; Goulding, C.J. Estimation of timber volume in a coniferous plantation forest using Landsat TM. *Int. J. Remote Sens.* **1997**, *18*, 2209–2223.

49. Reese, H.; Nilsson, M.; Sandström, P.; Olsson, H. Applications using estimates of forest parameters derived from satellite and forest inventory data. *Comput. Electron. Agr.* **2002**, *37*, 37–55.

50. Thünen Institut. Dritte Bundeswaldinventur-Ergebnisdatenbank, Auftragskürzel 77Z1PB_L458mf_0212_biHb, Archivierungsdatum: 2014–12–22 20:2:36.107, Überschrift: Zuwachs des Vorrates [m³/ha*a] nach Land und Baumartengruppe (rechnerischer Reinbestand), Filter: Periode=2002–2012. Available online: https://bwi.info (accessed on 30 June 2015).

51. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemometr. Intell. Lab.* **2001**, *58*, 109–130.

52. Farifteh, J.; van der Meer, F.; Atzberger, C.; Carranza, E. Quantitative analysis of salt-affected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN). *Remote Sens. Environ.* **2007**, *110*, 59–78.

53. Asner, G.; Martin, R. Spectral and chemical analysis of tropical forests: Scaling from leaf to canopy levels. *Remote Sens. Environ.* **2008**, *112*, 3958–3970.

54. Zhang, J.; Wu, J.; Zhou, L. Deriving vegetation leaf water content from spectrophotometric data with orthogonal signal correction-partial least square regression. *Int. J. Remote Sens.* **2011**, *32*, 7557–7574.

55. Axelsson, C.; Skidmore, A.K.; Schlerf, M.; Fauzi, A.; Verhoef, W. Hyperspectral analysis of mangrove foliar chemistry using PLSR and support vector regression. *Int. J. Remote Sens.* **2013**, *34*, 1724–1743.

56. Wolter, P.T.; Townsend, P.A.; Sturtevant, B.R. Estimation of forest structural parameters using 5 and 10 meter SPOT-5 satellite data. *Remote Sens. Environ.* **2009**, *113*, 2019–2036.

57. Lei, C.; Ju, C.Y.; Cai, T.J.; Jing, X.; Wei, X.H.; Di, X.Y. Estimating canopy closure density and above-ground tree biomass using partial least square methods in Chinese boreal forests. *J. Forest Res.* **2012**, *23*, 191–196.

58. Vaglio Laurin, G.; Chen, Q.; Lindsell, J.A.; Coomes, D.A.; Frate, F.D.; Guerriero, L.; Pirotti, F.; Valentini, R. Above ground biomass estimation in an African tropical forest with Lidar and hyperspectral data. *ISPRS J. Photogramm. Remote Sens.* **2014**, *89*, 49–58.

59. Tomppo, E.; Olsson, H.; Ståhl, G.; Nilsson, M.; Hagner, O.; Katila, M. Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sens. Environ.* **2008**, *112*, 1982–1999.

60. Gong, G. Crossvalidation, The Jackknife, and the Bootstrap: Excess Error Estimation in Forward Logistic Regression. Ph.D. Thesis, Stanford University, Stanford, CA, USA, 1982.

61. Höskuldsson, A. PLS regression methods. *J. Chemometr.* **1988**, *2*, 211–228.

62. Buddenbaum, H.; Steffens, M. Mapping the distribution of chemical properties in soil profiles using laboratory imaging spectroscopy, SVM and PLS regression. *EARSeL eProc.* **2012**, *11*, 25–32.

63. Katila, M.; Tomppo, E. Selecting estimation parameters for the Finnish multisource National Forest Inventory. *Remote Sens. Environ.* **2001**, *76*, 16–32.

64. Kilpeläinen, P.; Tokola, T. TM image-based estimates of stand volume. *Forest Ecol. Manag.* **1999**, *124*, 105–111.

65. Heiskanen, J. Estimating aboveground tree biomass and leaf area index in a mountain birch forest using ASTER satellite data. *Int. J. Remote Sens.* **2006**, *27*, 1135–1158.

66. Ardö, J. Volume quantification of coniferous forest compartments using spectral radiance recorded by Landsat Thematic Mapper. *Int. J. Remote Sens.* **1992**, *13*, 1779–1786.

67. Gemmell, F. Effects of forest cover, terrain and scale on timber volume estimation with TM data in a Rocky Mountain site. *Remote Sens. Environ.* **1995**, *51*, 294–305.

68. Muukkonen, P.; Heiskanen, J. Estimating biomass for boreal forests using ASTER satellite data combined with standwise forest inventory data. *Remote Sens. Environ.* **2005**, *99*, 434–447.

69. Stenberg, P. Penumbra in within-shoot and between-shoot shading in conifers and its significance for photosynthesis. *Ecol. Model.* **1995**, *77*, 215–231.

70. Smolander, S.; Stenberg, P. A method to account for shoot scale clumping in coniferous canopy reflectance models. *Remote Sens. Environ.* **2003**, *88*, 363–373.

71. Vohland, M.; Stoffels, J.; Hau, C.; Schüler, G. Remote sensing techniques for forest parameter assessment: Multispectral classification and linear spectral mixture analysis. *Silva Fennica* **2007**, *41*, 441–456.

72. Hawkins, D.M. The problem of overfitting. *J. Chem. Inf. Model.* **2004**, *44*, 1–12.

73. Atzberger, C.; Guérif, M.; Baret, F.; Werner, W. Comparative analysis of three chemometric techniques for the spectroradiometric assessment of canopy chlorophyll content in winter wheat. *Comput. Electron. Agr.* **2010**, *73*, 165–173.

74. Raykov, T.; Marcoulides, G.A. *An Introduction to Applied Multivariate analysis*; Routledge: New York, NY, USA, 2008.

75. Hyyppä, J.; Hyyppä, H.; Inkinen, M.; Engdahl, M.; Linko, S.; Zhu, Y.-H. Accuracy comparison of various remote sensing data sources in the retrieval of forest stand attributes. *Forest Ecol. Manag.* **2000**, *128*, 109–120.

76. McRoberts, R.; Tomppo, E. Remote sensing support for national forest inventories. *Remote Sens. Environ.* **2007**, *110*, 412–419.

77. Bedison, J.E.; McNeil, B.E. Is the growth of temperate forest trees enhanced along an ambient nitrogen deposition gradient? *Ecology* **2009**, *90*, 1736–1742.

78. McMahon, S.M.; Parker, G.G.; Miller, D.R. Evidence for a recent increase in forest growth. *Proc. Nation. Acad. Sci.* **2010**, *107*, 3611–3615.

79. Lausch, A.; Heurich, M.; Gordalla, D.; Dobner, H.J.; Gwillym-Margianto, S.; Salbach, C. Forecasting potential bark beetle outbreaks based on spruce forest vitality using hyperspectral remote-sensing techniques at different scales. *Forest Ecol. Manag.* **2013**, *308*, 76–89.

80. Jönsson, A.; Appelberg, G.; Harding, S.; Bärring, L. Spatio-temporal impact of climate change on the activity and voltinism of the spruce bark beetle, lps typographus. *Glob. Change Biol.* **2009**, *15*, 486–499.

81. Marini, L.; Ayres, M.P.; Battisti, A.; Faccoli, M. Climate affects severity and altitudinal distribution of outbreaks in an eruptive bark beetle. *Clim. Change* **2012**, *115*, 327–341.