

Article

# Designing a Validation Protocol for Remote Sensing Based Operational Forest Masks Applications. Comparison of Products Across Europe

Angel Fernandez-Carrillo \*, Antonio Franco-Nieto, Erika Pinto-Bañuls, Miguel Basarte-Mena and Beatriz Revilla-Romero

Remote Sensing and Geospatial Analytics Division, GMV, Isaac Newton 11, P.T.M. Tres Cantos, E-28760 Madrid, Spain; afranco@gmv.com (A.F.-N.); erika.pinto.b@gmv.com (E.P.-B.); mbasarte@gmv.com (M.B.-M.); brevilla@gmv.com (B.R.-R.)

\* Correspondence: aafernandez@gmv.com; Tel.: +34 660-136-356

Received: 1 September 2020; Accepted: 24 September 2020; Published: 26 September 2020



**Abstract:** The spatial and temporal dynamics of the forest cover can be captured using remote sensing data. Forest masks are a valuable tool to monitor forest characteristics, such as biomass, deforestation, health condition and disturbances. This study was carried out under the umbrella of the EC H2020 MySustainableForest (MSF) project. A key achievement has been the development of supervised classification methods for delineating forest cover. The forest masks presented here are binary forest/non-forest classification maps obtained using Sentinel-2 data for 16 study areas across Europe with different forest types. Performance metrics can be selected to measure accuracy of forest mask. However, large-scale reference datasets are scarce and typically cannot be considered as ground truth. In this study, we implemented a stratified random sampling system and the generation of a reference dataset based on visual interpretation of satellite images. This dataset was used for validation of the forest masks, MSF and two other similar products: HRL by Copernicus and FNF by the DLR. MSF forest masks showed a good performance ( $OA_{MSF} = 96.3\%$ ;  $DC_{MSF} = 96.5$ ), with high overall accuracy (88.7–99.5%) across all the areas, and omission and commission errors were low and balanced ( $OE_{MSF} = 2.4\%$ ;  $CE_{MSF} = 4.5\%$ ;  $relB_{MSF} = 2\%$ ), while the other products showed on average lower accuracies ( $OA_{HRL} = 89.2\%$ ;  $OA_{FNF} = 76\%$ ). However, for all three products, the Mediterranean areas were challenging to model, where the complexity of forest structure led to relatively high omission errors ( $OE_{MSF} = 9.5\%$ ;  $OE_{HRL} = 59.5\%$ ;  $OE_{FNF} = 71.4\%$ ). Comparing these results with the vision from external local stakeholders highlighted the need of establishing clear large-scale validation datasets and protocols for remote sensing-based forest products. Future research will be done to test the MSF mask in forest types not present in Europe and compare new outputs to available reference datasets.

**Keywords:** forest mask; validation; probability sampling; remote sensing; earth observations; forestry; accuracy assessment; forest classification

## 1. Introduction

One of the main data challenges forest managers face is the lack of accurate and up-to-date data to support specific activities in the silvicultural cycle, such as reporting and planning for commercial forestry or recreational activities [1,2]. Forests are dynamic environments and information related to its characterization and health condition is key [3–5]. In addition, some natural dynamics such as pest and droughts are being exacerbated by recent changes in climate, causing more frequent damages and costs [3,6–9]. Therefore, there is a need for reliable, precise and quality data which can cover large

areas and with frequent updates adapted to meet forest managers' needs. Remote sensing, such as satellite [10–12] and LiDAR [13–15] sensors, offers data and added-value products to operational solutions for silviculture. In this framework, the European Union Horizon-2020 MySustainableForest (MSF) project provides a portfolio of geo-information products to support forest activities and sustainable management from the afforestation to the forest products transformation markets [16]. These products are based on satellite data, LiDAR, non-invasive sonic measurements and statistical data. The MSF project has demonstrated the advantage of incorporating remote sensing products into the daily decision making, protocols and operations of the different stakeholders across the silvicultural chain.

One of the most important variables to monitor regularly is the forest extension regularly. This can be achieved using datasets commonly called forest masks. A forest mask is a spatial binary forest/non-forest land classification which can be derived with different methods and datasets [17,18]. The information provided by forest masks helps to track changes related to forest characterization [19,20], wood quality (e.g., forest types, site index and wood strength), forest monitoring (e.g., restoration, degradation, deforestation [21] and biomass [22,23]) or vegetation stress monitoring (e.g., damage due to pest [24], drought events, etc.). One of the key products of MSF is a forest mask, which is used as baseline to produce other outputs related to forest characterization and health condition.

Maps derived from remote sensing data contain errors from different sources [25]. Validation is a key requirement and step of the production of remote sensing datasets [25,26]. Products reliability and accuracy is highly demanded by the broad range of forest stakeholders, which often claim the quality of remote sensing-derived maps to be too low for operational use [27]. Feedback from forestry end-users has been key to develop and improve MSF products during the project. Forestry experts from seven European institutions, namely the Centre Nationale de la Propriété Forestière (CNPF), the Croatian Forest Research Institute (CFRI), the Forest Owners Association of Lithuania (FOAL), the Forest Owners Association of Navarra (FORESNA), Madera Plus Calidad Forestal (MADERA+), the Instituto de Investigação da Floresta e Papel (RAIZ) and the University Forest Enterprise Masaryk Forest Křtiny of Mendel University in Brno (UFE), have carried out a validation based on a combined qualitative and quantitative analysis of the products. The qualitative analysis was based on the expert knowledge from local stakeholders, while, for the quantitative analysis, local experts used their own field measurements to compare them against MSF outputs. Nevertheless, this external validation has several problems related to the sampling strategy and performance metrics for the quantitative validation and to the subjective view of the operator in case of the qualitative validation. Sample size was often too low to yield robust accuracy estimates. Hence, very diverse results were obtained for the same product (e.g., forest mask) in areas with similar forest characteristics. This experience highlighted the need of carrying out independent validation processes with a clear and homogeneous protocol. The use of a probabilistic sampling strategy cancels out the above-mentioned problems as it produces unbiased performance metrics with known associated errors [26,28–30].

Independent ground truth datasets must be used to obtain unbiased validation metrics [26,31]. There are several large scale, global or continental, forest mask datasets which may be compared to MSF forest mask [32–34]. However, these existing datasets have associated levels of uncertainty and thus should not be considered as ground truth. Validating new products with these existing datasets would provide useful information but results would show a comparison between products (i.e., each product might have errors from different sources) rather than a validation with real ground-truth data [35]. Therefore, the performance metrics obtained would show the agreement/disagreement between different products without any precise statistical meaning about product quality. Considering the lack of accurate ground-truth data for forest/non-forest classification in Europe, another option is the generation of a validation dataset based in visual interpretation by an independent expert [30,36]

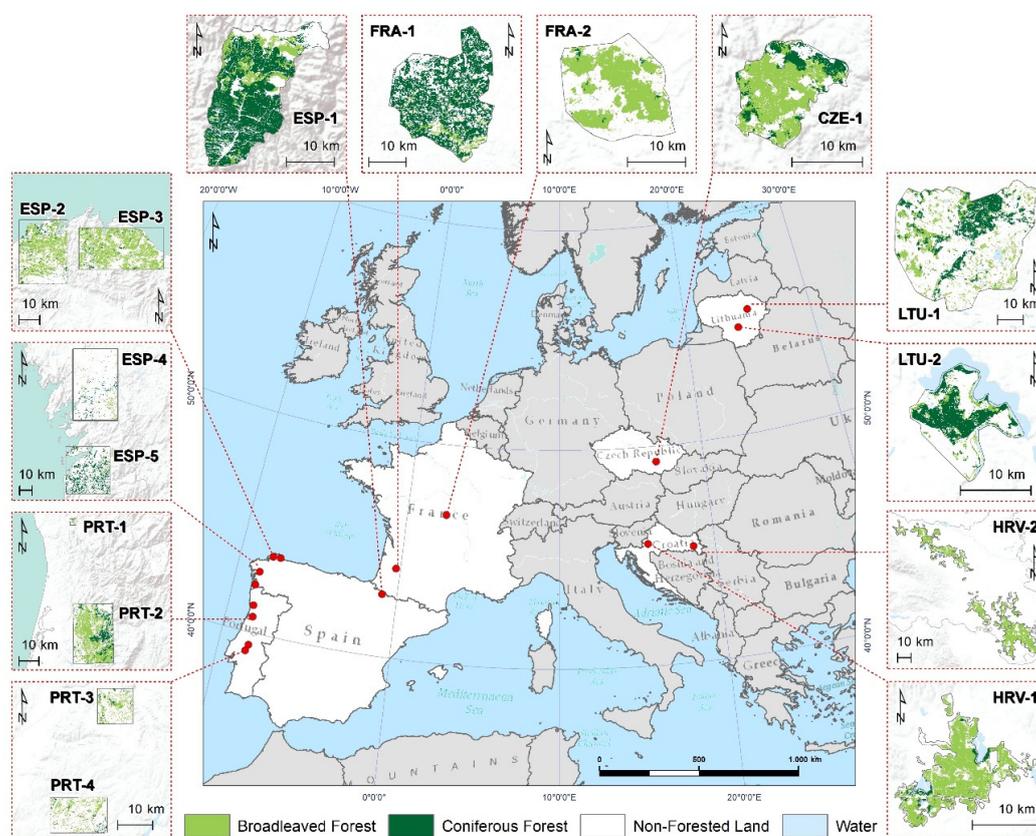
The aim of this study was to establish a validation methodology suitable for operational forest remote sensing products across different locations in Europe, assuming there is no homogeneous field dataset available. The specific objectives of this study were: (1) to design an operational validation

protocol for large-scale forestry products; (2) to validate our forest masks across 16 locations and European forest types; and (3) to assess the skill of MSF Forest Masks in comparison to other large-scale products available on those areas.

## 2. Study Areas and Data

### 2.1. Study Areas

The MSF forest mask layers used on this study were implemented on sixteen case demos in six countries and include the most representative forest types across Europe (Figure 1). These Areas of Interest (AOI) are managed by forest owners' associations and forestry research institutes which promote conservation and sustainable management values in over one million hectares in size. For simplicity, each AOI has an acronym related to the country (e.g., Spain is ESP) and the number of AOI present in each country.



**Figure 1.** Location of the 16 AOIs across Europe with their abbreviations. The forest type classes showed are taken from the Joint Research Centre Forest Type Map (2006) [20,37].

Further details of each AOI can be found on the MySustainableForest demo cases website section [16], including the details of the main forest systems present:

- Croatia: Continental lowland forests by heterogeneous stands and lowland Slavonian pedunculated oak forests (HRV-1 and HRV-2).
- Czech Republic: Temperate Pannonian mixed forests: (i) conifers, namely spruces, pines and larches; and (ii) broadleaf, namely beeches, oaks and hornbeams (CZE-1).
- France: Oceanic maritime pine forest (FRA-1) and temperate continental with pedunculated oak (FRA-2).

- Lithuania: Boreal forest with forests dominated by scots pines, birches and Norway spruces (LTU-1 and LTU-2).
- Portugal: Plantations of *Eucalyptus spp.* on two climate subtypes: Atlantic-Mediterranean (PRT-1 and PRT-2) and typical Mediterranean combined with agroforestry areas (PRT-3 and PRT-4).
- Spain: Alpine forest dominated by beeches, oaks and pines (ESP-1) and Atlantic plantations of *Eucalyptus spp.* where land ownership is highly fragmented (ESP-2, ESP-3, ESP-4 and ESP-5).

## 2.2. High Resolution Forest/Non-Forest MSF Classification Dataset

According to Food and Agriculture Organization (FAO), forest land is any surface of more than 0.5 ha with more than 10% of tree canopy and trees higher than 5 m, excluding land predominated by agricultural or urban uses [38]. To calculate the MSF forest mask used in this study, we considered as forest the area with more than 50% of tree stratum coverage (i.e., when more that 50% of the pixel appears covered by tree canopy) [33]. We are aware that we are excluding areas with lower than 50% tree coverage which per definition may cause some disagreement with ground truth data and other forest mask datasets.

Within the MSF project, the forest mask products were derived using three different input datasets: satellite high resolution and very high resolution imagery and LiDAR data. We present here the results of developing forest masks layers based on Sentinel imagery for 2018, as well as satellite data from the autumn/winter period of 2017 for some Spanish AOIs (Table 1). A first version of this product was previously published [18], being this one an improved version based on the experiences learned from the first one. Two images (i.e., summer and winter) were used and processed to capture different phenological conditions, especially for deciduous/mixed forest discrimination. Sentinel-2 images were acquired in format Level-2A (i.e., surface reflectance). Only bands in the visible NIR and SWIR wavelengths were selected. Visible and NIR bands have a pixel size of 10 m. SWIR bands, originally at 20 m spatial resolution, were resampled to 10 m. The obtained forest masks had a spatial resolution of 10 m and Minimum Mapping Unit (MMU) of 0.1 ha (equivalent to 10 pixels/10 m pixel). A Random Forest classifier [39] was trained using as input the original Sentinel-2 bands, several vegetation and texture indices. Normalized Difference Vegetation Index (NDVI) [40], Enhanced Vegetation Index (EVI) [41], Transformed Chlorophyll Absorption Reflectance Index (TCARI) [42], Bare Soil Index [43] and Tasseled Cap (TC) Wetness [44] were added to the input data together with NDVI homogeneity and entropy [45] The model was trained using 2500 forest/non-forest random points distributed across Europe.

**Table 1.** Tile and acquisition date of Sentinel-2 imagery used to generate the validation dataset.

AOI Name	Sentinel-2 Tile	Date 1	Date 2
CZE-1	T33UXQ	22-03-2018	29-08-2018
ESP-1	T30TXN	26-10-2017	04-08-2018
ESP-2	T29TNJ	22-02-2018	11-08-2018
ESP-3	T29TPJ	21-12-2017	11-08-2018
ESP-4	T29TNH	22-02-2018	11-08-2018
ESP-5	T29TNG	24-02-2018	11-08-2018
FRA-1	T30TYP/T30TYQ	24-01-2018	19-08-2018
FRA-2	T31TDM	25-02-2018	19-08-2018
HRV-1	T33TWL	08-04-2018	01-08-2018
HRV-2	T33TYL/T34TCR/T34TCQ	11-03-2018	13-08-2018
LTU-1	T35ULB	18-03-2018	23-08-2018
LTU-2	T35ULA	18-03-2018	23-08-2018
PRT-1	T29TNF	26-03-2018	18-08-2018
PRT-2	T29TNF	26-03-2018	18-08-2018
PRT-3	T29SND	26-03-2018	18-08-2018
PRT-4	T29SND	26-03-2018	18-08-2018

### 2.3. High Resolution Forest/Non-Forest External Classification Datasets

There are many methodologies and providers of forest mask layers. However, the number of forest/non-forest products available at large scale as in this case, across Europe, is limited. Two freely available forest/non-forest classification datasets were selected to compare them against MSF forest mask. The products chosen are comparable to MSF forest mask as they were produced close in time, within 3 years of the Sentinel images used for the development of our layers, and have lower but still comparable spatial resolution (20–50 m):

- Forests High-Resolution Layer (HRL) is provided by Copernicus Land Monitoring Service (CLMS) and coordinated by the European Environment Agency (EEA) [32]. The Tree Cover Density (TCD) product provides the level of tree coverage per pixel in a range of 0–100%. It is obtained through a semi-automatic classification of multitemporal satellite images (Sentinel-2 and Landsat-8) for the year 2015 ( $\pm 1$  year) [32], using a combination of supervised and unsupervised classification techniques. The resulting product has 20 m spatial resolution. To make it comparable with the MSF forest masks, the TCD product was pre-processed to obtain a binary forest/non-forest classification, considering as forest all those pixels with a cover density of 50% or higher. A future update of this dataset for 2018 reference year was announced in August 2022 by Copernicus, but it was not available prior submission of this paper.
- TanDEM-X Forest/Non-Forest (FNF) is provided by Microwaves and Radar Institute of the German Aerospace Center (DLR) [33,34]. It is a global forest/non-forest map generated from interferometric synthetic aperture radar (InSAR) data from the TanDEM-X mission. The bistatic stripmap single polarization (HH) InSAR data were acquired between 2011 and 2016, being 2015 the reference year for the final product. FNF maps are produced with a final pixel size of 50 m.

### 2.4. Satellite Data Required to Build the Independent Validation Dataset

Two sources with different bands spatial resolutions were used to create the validation dataset: (1) High Resolution imagery: Sentinel-2 L2A provided by Copernicus (Table 1); and (2) Very High Resolution imagery: satellite data (GoogleEarth™ and Bing Aerial imagery), taking into consideration the spatial and temporal resolution of the forest/non-forest classification datasets.

## 3. Methods

Validation and intercomparison of the results across different areas may not be straightforward if not following a standard methodology. Considering the non-existence of homogenized forest/non-forest ground truth datasets, there are different strategies that may be implemented to perform a quantitative validation of forest masks:

- Cross validation or dataset split [46]. Commonly used for products developed using any supervised classification approach. A training dataset is needed, either provided by the user or built by the producer. The use of cross-validation techniques is widely accepted when there are not independent training and testing sets. However, the statistical distribution of the training and test samples are not independent, leading to an optimistic bias in the resulting metrics [46]. Note that the final values of the metrics will be strongly determined by the quality of the input dataset.
- Using data from national forest inventories (NFI). Many methodologies used to perform forest inventories as the selected method by each nation will depend of the purpose and the scale of the inventory, leading to significant efforts to perform data search, normalization and data engineering. In Europe, the European National Forest Inventory Network (ENFIN) [47] promotes NFIs and harmonizes forest information. However, there are still different methods and plot sizes used. Moreover, inventories contain errors from different sources, as they rely on sampling strategies and the measuring protocols are not always clear, thus adding uncertainty to the metrics generated through the validation process [4,48,49]. Hence, validation metrics might differ significantly

across areas where the product has identical quality, due to the different inventory methods used, which are in general heterogeneous.

- Building independent datasets based in visual interpretation of images [50]. This method is costly, and the validation must be carried out by independent interpreters in order to build an unbiased dataset trying to represent the variability of forest types within a determined area. The statistical metrics obtained with this method may be close to the real quality of the product if sampling and interpretation processes are carefully designed.

Considering the difficulties of using cross-validation and national forest inventories for our objectives, a new reference forest/non-forest dataset was generated by an independent operator to validate MSF forest masks and compare it with the other external products (e.g., HRL and FNF). The validation strategy was divided into four phases: (1) sampling design; (2) generation of the reference dataset through visual interpretation; (3) definition of performance metrics; and (4) intercomparison of products. The methods explained in subsequent sections are summarized in the flowchart (Figure 2).

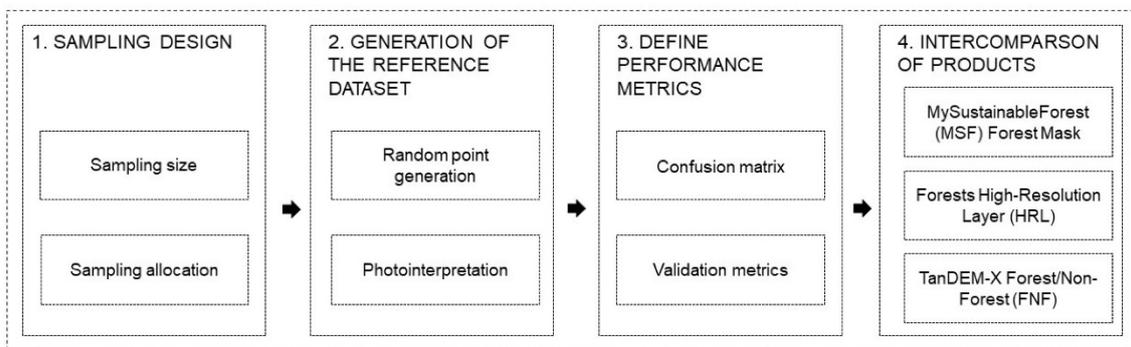


Figure 2. Methods used to validate the forest mask products.

### 3.1. Sampling Design

Pixels were the spatial unit selected as validation samples. Following the recommendations made by Olofsson et al. [26], a probability sampling strategy was applied to select the samples of the reference dataset. The two conditions defining a probability sample are: (1) the inclusion probability must be known for each unit selected in the sample; and (2) the inclusion probability must be greater than zero for all units in the AOI [29]. A stratified random sampling strategy was designed considering the two strata forest and non-forest. This design allows increasing the sample size in classes that occupy a small proportion of area to reduce the standard errors of the class-specific accuracy estimates for these rare classes [26].

To obtain the number of points to be included in the reference dataset, the following equation was used (Equation (1)) [51]:

$$n = \frac{(\sum W_i S_i)^2}{[S(\hat{\theta})]^2 + (1/N) \sum W_i S_i^2} \approx \left( \frac{\sum W_i S_i}{S(\hat{\theta})} \right)^2 \quad (1)$$

where  $N$  represents the number of units in the AOI,  $S(\hat{\theta})$  is the standard error of the estimated overall accuracy,  $W_i$  is the mapped proportion of area of stratum  $i$  and  $S_i$  is the standard deviation of stratum  $i$ .  $S_i = \sqrt{U_i(1 - U_i)}$ , where  $U_i$  is the expected user's accuracy.

With a pessimistic user's accuracy estimate of 0.7 and a standard error of 0.01 for the overall accuracy, the total number of points for the dataset was 2100.

To choose the optimal number of points per AOI, the recommendation [26] is to use an allocation method that should meet two conditions. Firstly, the number of points per AOI should lay between equal and proportional allocation (i.e., proportional to the area of the AOI relative to the total area of all

the AOIs). Secondly, the final allocation should push the proportional distribution to the equal, in order to have enough samples to robustly validate the poorer class, but without it becoming completely equal.

The specific equation (Equation (2)) used for sample allocation per AOI was:

$$n_{AOI} = \frac{p + 2e}{3} \quad (2)$$

where  $p$  is the distribution of  $n$  proportional to the area of each AOI and  $e$  is the equal distribution of  $n$  for each AOI.

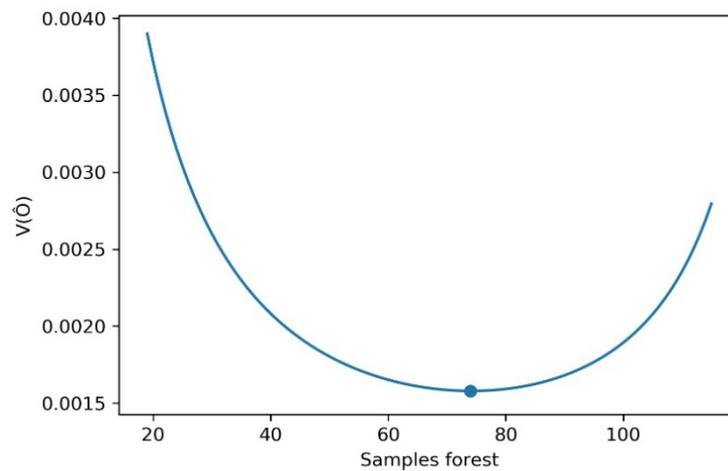
Once the sample size per AOI was computed, samples need to be allocated to strata (i.e., forest/non-forest). It is important that the sample allocation allows precise estimates of accuracy and area [52]. Having a fixed number of samples per AOI to allocate in two strata, proportional allocation would increase the variances of the performance metrics for the rarer class, as it would be poorly represented. Using equal allocation would allow a more robust validation of the poorer class at the expense of increasing the variances of the bigger class. The optimal sample allocation should minimize the variance of the desired target metric. Considering the overall accuracy (OA) as the metric whose variance needs to be minimized, the variance of an estimated OA can be computed using the following equation (Equation (3)) [26]:

$$\hat{V}(\hat{O}) = \sum_{i=1}^q W_i^2 \hat{U}_i (1 - \hat{U}_i) / (n_i - 1) \quad (3)$$

Being  $n_i$  the number of samples per stratum  $i$  in a specific AOI, the optimal sample allocation (i.e., optimal value of  $n_i$  per stratum) was computed through an iterative process estimating the variance of a desired OA for all possible values of  $n_i$ . The values of  $n_i$  which yielded the minimum variance of OA were selected as optimal sample allocation (Table 2 and Figure 3).

**Table 2.** Final distribution of points per AOI and stratum.

AOIs	Total Points	Forest Points	Non-Forest Points
CZE-1	97	58	30
ESP-1	115	73	42
ESP-2	135	71	64
ESP-3	142	82	60
ESP-4	226	115	109
ESP-5	176	78	98
FRA-1	129	72	57
FRA-2	105	52	53
HRV-1	96	64	32
HRV-2	159	103	56
LTU-1	217	94	123
LTU-2	96	46	50
PRT-1	88	41	47
PRT-2	127	69	58
PRT-3	94	42	52
PRT-4	98	54	44
<b>Total</b>	<b>2100</b>	<b>1114</b>	<b>984</b>



**Figure 3.** Changes in the variance of the overall accuracy ( $V(\hat{O})$ ) depending on the number of samples assigned to the forest class for the AOI ESP-1. The optimal number of forest samples to minimize the variance was 73.

### 3.2. Dataset Generation

Once sampling size and allocation were computed, the validation dataset was created. In each AOI, the random points were generated with a minimum distance of 100 m between them. Each point was subsequently assigned to forest or non-forest via photointerpretation by an independent trained consultant. Sentinel-2 and very high resolution images (see data, Section 2.4) were used as basis for the visual interpretation.

For the Sentinel-2 images, the RGB composition used was NIR/SWIR/red, as it allows clear vegetation discrimination due to the high reflectance of the vegetation in NIR wavelengths [53]. Forests were visualized in red tones for the coniferous species and as light reds or orange for deciduous species. Non-forest areas, such as crops, roads or buildings, were visualized in green, blue or yellow. High resolution images were visualized in true color (i.e., red/green/blue).

### 3.3. Performance Metrics

Confusion matrices (Table 3) were built for each area using the forest mask products and the reference dataset. A total confusion matrix was computed pooling the samples of all the areas.

**Table 3.** Sample confusion matrix for a forest/non-forest classification.

Predicted Condition	True Condition		Total
	Forest	Non-Forest	
Forest	True Positive (TP)	False Positive (FP)	Predicted Condition Positive (PCP)
Non-Forest	False Negative (FN)	True Negative (TN)	Predicted Condition Negative (PCN)
Total	Condition Positive (CP)	Condition Negative (CN)	sample size (n)

Different agreement and error metrics were computed from confusion matrices (Table 4) [35]. All metrics were multiplied by 100 to be expressed as percentage.

**Table 4.** Confusion matrix-derived agreement and errors metrics.

	Definition	Equation	
<b>Agreement metrics</b>	Overall Accuracy (OA): Proportion of pixels correctly classified.	$OA = \frac{TP+TN}{n}$	(1)
	Precision (P): Proportion of correctly predicted (i.e., classified) cases from all those predicted as positive.	$P = \frac{TP}{TP+FP}$	(2)
	Recall (R): Proportion of correctly predicted cases from all the real positives.	$R = \frac{TP}{TP+FN}$	(3)
	Dice similarity Coefficient (DC) or F1-score: Harmonic mean of precision and recall.	$DC = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$	(4)
<b>Error metrics</b>	Commission Error (CE): Proportion of misclassified pixels from all those predicted as positive.	$CE = \frac{FP}{TP+FP}$	(5)
	Omission Error (OE): Proportion of misclassified pixels from all the real positives.	$OE = \frac{FN}{TP+FN}$	(6)
	Relative Bias (relB): It quantifies the systematic error of the classification.	$relB = \frac{(FP-FN)}{TP+FN}$	(7)

#### 4. Results

Performance metrics for each forest/non-forest mask layer (i.e., MSF, HRL and FNF) can be found in Table 5. For the pooled set of AOIs, agreement metrics were high, the OA ranging from 88.7% to 99.5% and the DC from 91.1% to 99.5%. Highest scores were always found for MSF, followed by HRL and finally FNF. Errors and bias were low and balanced in the MSF product (OE = 2.4%; CE = 4.5; relB = 2%), while HRL (OE = 14.4%; CE = 6.3; relB = −8.6 %) and FNF (OE = 27.3%; CE = 19.7; relB = −9.5%) showed higher and imbalanced errors, which led to higher relative bias. HRL and FNF products had higher omission than commission errors (negative relB). As a summary, the relative bias of MSF product was lower and the commission error was slightly higher than the omission error, thus confirming the good calibration and balance of MSF errors.

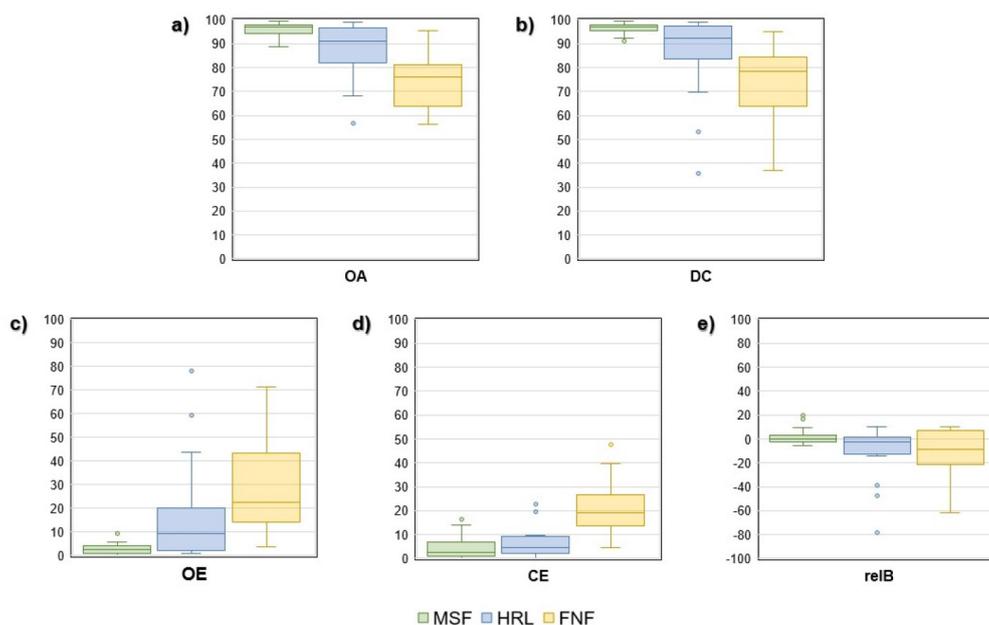
**Table 5.** Performance metrics of MSF, HRL and FNF products.

	MSF	HRL	FNF
<b>OA</b>	96.3	89.2	76.0
<b>DC</b>	96.5	89.4	76.8
<b>P</b>	95.5	93.6	80.3
<b>R</b>	97.6	85.6	72.6
<b>OE</b>	2.4	14.4	27.3
<b>CE</b>	4.5	6.3	19.7
<b>RelB</b>	2.0	−8.6	−9.5

Analyzing the agreement metrics by AOIs (Table 6, Figure 4), we observed that results follow the same trend in overall accuracy (OA) and Dice similarity Coefficient (DC). The Interquartile Range (IQR, i.e. the difference between the first and third quartiles) was narrower in MSF (IQR<sub>OA</sub> = 2.93 and IQR<sub>DC</sub> = 2.25), followed by HRL (IQR<sub>OA</sub> = 13.95 and IQR<sub>DC</sub> = 13.28) and FNF (IQR<sub>OA</sub> = 16.85 and IQR<sub>DC</sub> = 18.83). This confirmed a good performance of MSF forest mask across different European forest types. MSF and HRL present median values higher than 90% in the OA and DC, while, in FNF, these values are 76.05% and 78.55%. The high dispersion of agreement values in FNF also pointed out a high variability of the product across different forest types.

**Table 6.** Performance metrics of 16 AOIs grouped by product.

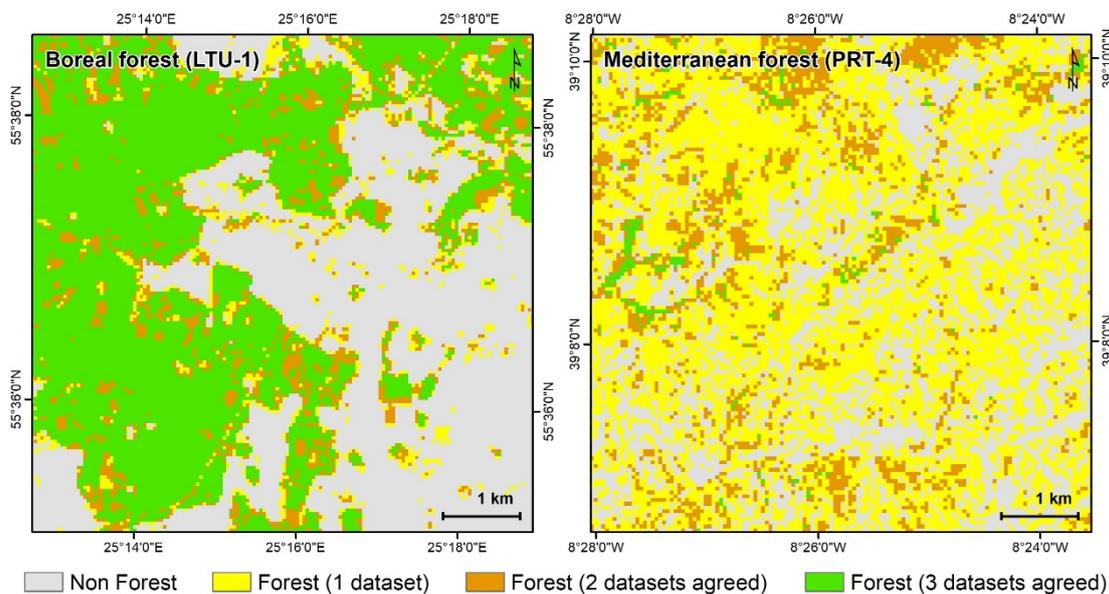
	CZE-1	ESP-1	ESP-2	ESP-3	ESP-4	ESP-5	FRA-1	FRA-2	HRV-1	HRV-2	LTU-1	LTU-2	PRT-1	PRT-2	PRT-3	PRT-4	
MSF	OA	96.9	98.2	97.0	88.7	97.3	97.1	97.7	97.2	93.7	98.7	99.5	97.9	96.5	95.2	92.5	90.9
	DC	97.4	98.6	97.2	91.1	97.4	96.8	97.9	97.0	95.5	99.0	99.5	97.8	96.3	95.6	91.6	92.4
	P	98.2	98.6	97.2	83.7	97.4	95.0	98.6	100	91.4	99.0	100	100	95.2	97.0	92.7	85.9
	R	96.5	98.6	97.2	100	97.4	98.7	97.2	94.2	100	99.0	98.9	95.6	97.6	94.2	90.5	100
	OE	3.4	1.3	2.8	0.0	2.6	1.2	2.8	5.8	0.0	0.9	1.0	4.3	2.4	5.8	9.5	0.0
	CE	1.7	1.3	2.8	16.3	2.6	4.9	1.4	0.9	8.6	0.9	0.0	0.0	4.7	3.0	7.3	14.0
	relB	-1.7	0.0	0.0	19.5	0.0	3.8	-1.4	-5.7	9.3	0.0	-1.0	-4.3	2.4	-3.0	-2.3	16.3
HRL	OA	96.9	96.5	92.6	88.7	89.7	98.3	82.9	99.0	88.5	96.2	97.2	94.8	77.3	81.9	68.1	56.6
	DC	97.4	97.3	92.7	89.7	89.9	98.1	83.6	99.0	91.5	97.1	96.7	94.4	69.7	84.1	53.1	35.8
	P	98.2	96.0	90.5	94.6	91.0	97.5	90.3	100	90.8	95.3	97.8	97.7	92.0	80.3	77.3	100
	R	96.5	98.6	90.1	85.4	88.7	98.7	77.8	98.0	92.2	99.0	95.7	91.3	56.0	88.4	40.5	21.8
	OE	3.4	1.3	9.8	14.6	11.3	1.2	22.2	1.9	7.81	0.9	4.2	8.7	43.9	11.6	59.5	78.2
	CE	1.7	4.0	4.5	5.4	8.9	2.6	9.7	0.0	9.23	4.7	2.1	2.3	8.0	19.7	22.7	0.0
	relB	-1.7	2.7	-5.6	-9.7	-2.6	1.3	-13.9	-1.9	1.6	3.9	-2.1	-6.5	-39.0	10.1	-47.6	-78.2
FNF	OA	81.4	63.5	77.8	79.6	73.2	81.2	60.5	95.3	76.0	76.1	91.7	91.6	64.7	72.4	56.3	57.6
	DC	85.0	65.0	77.6	81.0	72.7	79.5	59.8	95.2	82.4	82.2	90.0	90.9	63.5	75.8	36.9	44.7
	P	82.3	82.9	82.5	87.3	76.2	77.1	69.1	94.3	80.6	79.3	94.2	95.2	61.3	72.4	52.2	80.9
	R	87.9	53.4	73.2	75.6	69.6	82.0	52.8	96.1	84.4	85.4	86.2	87.0	65.8	79.7	28.6	30.9
	OE	12.1	46.6	26.7	24.9	30.4	17.9	47.2	3.8	15.6	14.5	13.8	13.0	34.1	20.3	71.4	69.1
	CE	17.7	17.0	17.4	12.7	23.8	22.9	30.9	5.6	19.4	20.7	5.8	4.7	39.6	27.6	47.8	19.0
	relB	6.9	-35.6	-11.2	-13.4	-8.7	6.4	-23.6	1.9	4.7	7.7	-8.5	-8.7	7.3	10.1	-45.2	-61.8



**Figure 4.** Box plots of performance metrics across the 16 AOIs for the three forest mask products: (a) overall accuracy; (b) Dice similarity coefficient; (c) omission errors; (d) commission errors, and (e) relative bias.

The  $IQR_{OE}$  (Table 6 and Figure 4) was higher in FNF (22.9%), followed by HRL (13.48%) and MSF (2.65%). The  $IQR_{CE}$  showed a similar behavior, being higher in FNF (8.83%) followed by HRL (6.73%) and finally MSF (4.3%). It should be noted that forest omissions were much higher than commissions in the case of HRL and FNF products. On the other hand, in MSF, the commission was greater than the omission, even though they achieved a great balance. The medians were lower than 10% for the MSF forest mask (OE = 2.5% and CE = 2.7%) and HRL (OE = 9.25% and CE = 4.6%), and it was around 20% in the case of FNF (OE = 22.6% and CE = 19.2%). The relative bias was higher for FNF and HRL than for MSF, highlighting more imbalanced errors. This fact was reinforced by the relatively high dispersion of relB in FNF and HRL. While for MSF the range of relB is 16.3, with areas with relB = 0%, the range of this metric reaches 88.3 for HRL and 71.9 for FNF. The low IQRs of MSF confirmed its consistent performance in forests with different characteristics compared to the other products analyzed.

There were some relevant differences among the AOIs (Figure 5). Independently of the forest mask used, the best metrics were obtained in Central and Northern Europe (CZE-1, FRA-2, LTU-1 and LTU-2), where continental forest have in general a continuous and dense canopy. All the forest mask products were less accurate for Portugal and Spain, especially for HRL and FNF, with OA and DC around 20% lower than MSF. Regarding the outliers of agreement metrics, most of them were located in the HRL product, mainly in Southern Portugal (PRT-4, OA; PRT-3 and PRT-4, DC). OE was up to 50% higher in areas with Mediterranean influence (PRT-3 and PRT-4). The greater omission problems were located in HRL and FNF datasets (PRT-3, OE = 71.4% and 59.5%, respectively) in areas dominated by tree–grass ecosystems, where also relative bias reached the highest values. Finally, there is an evidence that Eucalyptus plantations also presented some difficulties to be classified (CE and OE) by FNF.



**Figure 5.** Comparison between two forests in the three datasets: Boreal with high accuracy (left); and Mediterranean with low accuracy (right).

## 5. Discussion

The validation of the MSF high resolution forest masks on average yielded high scores for the agreement metrics and low for the error metrics. Nevertheless, we found some clear performance differences between the AOIs mainly related to the dominant forest types in each area. The best results were obtained for areas dominated by conifer, broadleaf or mixed forest in Central Europe. These forests are generally characterized by homogenous tree masses and crops which are relatively easy to discriminate using remote sensing data. The shrub stratum is common in these same areas under a dense tree canopy. On the other hand, the metrics were generally worse in areas closer to the Mediterranean climate, such as Portugal and Spain. The areas with the worst results (PRT-3 and

PRT-4) are characterized by the high presence of tree–grass ecosystems, which are similar to savannas. Tree–grass ecosystems are characterized by the co-existence of a sparse tree stratum and a grassland matrix with large seasonal contrasts [54,55]. Hence, reflectance contribution to the pixel highly depends on the behavior of the seasonal grass stratum, being this contribution variable depending on the density of the sparse tree stratum [56,57]. The complexity of these ecosystems in Southern Portugal (PRT-3 and PRT-4) led to relatively high omission errors compared to the other AOIs in all the products analyzed. Nevertheless, MSF forest masks had a maximum OE of 9% in these areas, while the HRL product obtained a maximum of 78.2% and FNF a maximum of 71.4%. This highlights the challenge of accurately detecting sparse tree–grass ecosystem [58]. Determining an exact threshold to discriminate pure cover types in transitional environments may lead to misinterpretation of natural landscapes, while providing fuzzy values might be more interesting to perform further analysis [59]. In the case of MSF forest mask, using a different threshold for assigning a pixel to forest (<50%) might change significantly the binary forest mask obtained for Southern Europe. The good performance of the MSF forest mask for this type of forest in comparison to the other products result from the great effort payed on this project to adapt the algorithm to different forest types across Europe, with a strong focus in Mediterranean tree–grass areas. Transitional woody shrublands are also present in some parts of AOIs in Southern Europe, resulting in higher commission errors in Portugal and Spain. For the AOIs located in Galicia (northwest of Spain; ESP-2, ESP-3, ESP-4 and ESP-5), the commission error was relatively high due to the large areas dominated by degraded shrublands, which have a similar spectral response as the dominant *Eucalyptus* plantations of these region. This effect seemed to be attenuated in HRL, while MSF and FNF showed higher CE in these AOIs.

HRL showed more outliers than the other products when comparing the metrics across AOIs, especially for the Dice coefficient (DC = 35.8%), commission error (CE = 78.2%) and relative bias (relB = −78.2%). These three outliers were yielded by the validation of PRT-4. In this area, HRL did not classify as non-forest any real forest pixel, but this happens at the cost of classifying misclassifying 78% of the predicted forest pixels. In this area, MSF also showed its highest relB but with opposite sign. Therefore, all the pixels classified as forest were correctly predicted (CE = 0%), but many real forest pixels are classified as non-forest (OE = 16.3%), which is common when working on sparse tree–grass ecosystem. This highlights the importance of considering relative bias for a complete performance vision even where the agreement metrics show high skill. Otherwise, it might lead to an optimistic interpretation of the results while the bias indicates the possible lack of robustness of the products in some areas.

The HRL official validation report [60] yielded a  $R^2 = 0.84$  for the Tree Cover Density (TCD) product, with significantly lower values for the Mediterranean region, especially for Portugal ( $R^2 = 0.66$ ), which coincides with the challenges found in this study. The HRL document also reports low  $R^2$  values for Iceland, the Arctic and Anatolian regions, which are not represented in MSF test areas or in this study. The best results in HRL report were found for Boreal, Pannonian and Continental bioregions ( $R^2 > 0.85$ ), which also coincide with the outcomes of this study. However, the methodology and results from both studies are not comparable since in the present work a pre-processing of the TCD was carried out to obtain a binary product.

The official FNF validation [33] reported a mean agreement of 86% for Germany and Eastern Europe. This is similar to the accuracy found for FNF in the present study in similar areas (CZE-1, LTU-1 and LTU-2). Nevertheless, Martone et al. [33] validated the FNF product using as ground truth the HRL product. Hence, this value of agreement should be taken with caution as the uncertainty of this validation and the errors of HRL products are not quantified in the FNF validation report.

External validations were carried out by local stakeholders within MSF project. ESP-2, ESP-3, ESP-4 and ESP-5, located in northwestern Spain, reported accuracies ranging between 93% and 100%. However, these AOIs were validated with sets of 8, 14, 31 and 20 points, making the final accuracy not significant enough. LTU-1 reported an overall accuracy of 90.4% and a kappa coefficient (k) [61] of 0.80 comparing MSF forest mask with information from the Lithuanian National Forest Inventory. The OA

given by external stakeholders in this case is 10% lower than the OA achieved in this study. For LTU-2, the difference was lower (external OA = 91.1%; external  $k = 0.82$ ). FRA-1 and FRA-2 were validated by stakeholders using non-systematic visual field control, reporting accuracies of 90% for both areas, around 8% below the OA found in this work. Stakeholders from CZE-1 reported an accuracy of more than 95% through visual inspection using high resolution ortophotos and cadaster data. In this last case, the difference with our OA was under 2%. At the time of writing this article, local validation from Portuguese and Croatian stakeholders is still pending, being the former particularly interesting to compare as it was clearly the most problematic area. However, the results of these external validations are hard to interpret and compare given the variability of methods and data sources used to carry out the accuracy assessments. The differences between the independent validation here presented and the external validation carried out by local stakeholders may be explained by two main points. Firstly, local forestry institutions are not always used to deal with continuous remote sensing data (i.e., based on pixels). Hence, these institutions may not have clearly defined validation protocols for this type of products, leading to several problems such as under-sampling or not taking into account the confidence level or standard error of the metrics computed for a certain area of interest. Secondly, the main purpose of local institutions may differ. Within the MSF project consortium, some institutions were centered in public forest management at national scale (e.g., CFRI), while others were associations of private owners (e.g., FORESNA) or privately owned companies with specific objectives (e.g., RAIZ is part of the pulp industry, while MADERA+ is focused on wood technological properties). This fact, together with the different nationalities involved, led to different definition of forests depending on national forest inventories or other sources. These considerations led to significantly different external validation approaches, which highlighted the need of standard definitions and methods to validate remote sensing-based forestry products, as the protocol proposed in this study.

There are several limitations of the validation exercise presented in this study and are explained below. The characteristics of the AOIs, namely the size, locations and forest types, were defined on the basis of local stakeholders within the MSF project consortium. Hence, while some forest types and regions are well represented (e.g., *Eucalyptus* plantations in northern Spain or mixed and broadleaf forests), other forest types, such as Northern European boreal forests, are poorly represented within the AOIs. A more detailed analysis and validation could take into account a more proportional representation of all forest types present in Europe. It is also important to bear in mind that, while the MSF forest masks were specifically calibrated for the European continent, the other two datasets analyzed (i.e., Copernicus HRL and Tandem-X FNF) are global datasets. This might explain the apparent better quality of the MSF product in the selected AOIs, which is likely to decrease if the algorithm is calibrated for other environments such as tropical forests or arid regions.

We are also aware that some of the products characteristics such as the date of the imagery used and the spatial resolution may have an impact on the performance and comparison of the products (MSF, HRL and FNF). As previously explained, the three products have different spatial resolutions and were developed using satellite images from different years. The products were validated using a reference dataset built using images from 2017–2018. As there is more consistency between the satellite images used to build the MSF layers and those for the validation dataset, this might have had a penalty of the results of the other two products. However, it is unlikely that difference between the dates used on the reference dataset and product (maximum three years) would have had such a high impact. We assume that the studied forests did not experience significant changes in the last five years, as previously stated directly by local stakeholders. A multi-temporal dataset (i.e., with reference samples for different years) would be needed to identify how much is the impact of difference between products and reference datasets affecting the performance metrics. In the case of the different spatial resolutions used, even when this may have had an effect on the final results, it is not likely to be the main factor explaining such remarkable differences in performance metrics, as we confirmed by doing a visual analysis of the most problematic areas (PRT-3 and PRT-4).

## 6. Conclusions

In this study, a protocol was established to validate remote sensing derived forest/non-forest classification maps across Europe. Based on a stratified random sampling, a reference dataset was created using visual interpretation of satellite images in order to validate MSF forest mask. Performance metrics were compared with two similar products: HRL and FNF.

The validation protocol allowed comparing the results of the accuracy assessment for different areas across Europe and different forest mask products. Problems were found when trying to compare the results of this assessment with the validation results provided by local experts. This was caused by the high variability of techniques used by stakeholders and justified the need of clearly established operational protocols for carrying out accuracy assessments of remote sensing-based products.

Accuracies were generally high for all forest mask products (OA = 76–96.3%), with MSF showing a more precise calibration for different European forest types (MSF accuracy ranges from 89% to 99.5%, while FNF accuracy ranges from 56% to 95%), which confirms its suitability as an adequate data source for large scale forest mapping in this continent. The greatest problems to discriminate forest from other land covers were found in areas with Mediterranean influence characterized by the presence of tree–grass ecosystems. Large shrublands in degraded areas also hindered forest discrimination. Performance metrics were significantly better for MSF in these areas compared to the other datasets analyzed, while FNF performance seemed to yield worse results. Note that the reference dataset was built with satellite imagery dates similar to the MSF masks, which also have higher spatial resolution (10 m). The other products used imagery up to three years older and with slightly lower resolution (20–50 m). This difference of dates and spatial resolutions might have slightly impacted the comparison of products. The performance of MSF satellite forest mask in other continents was out of the scope in this study but should be carried out in the near future.

The design of the accuracy assessment and selection of performance metrics highlight the need of establishing clear validation protocols for remote sensing-based forestry products. Efforts should be made in creating unified reference datasets at continental or global scales in order to clearly define the advantages and limitations of incorporating remote sensing-derived forest data in operational forestry processes.

**Author Contributions:** Conceptualization, A.F.-C.; Methodology, A.F.-C., A.F.-N. and E.P.-B.; Software, M.B.-M.; Validation, E.P.-B., A.F.-C. and A.F.-N.; Formal Analysis, A.F.-C.; Investigation, A.F.-C. and B.R.-R.; Resources, GMV; Data Curation, M.B.-M. and A.F.-C.; Writing—Original Draft Preparation, A.F.-C., B.R.-R. and A.F.-N.; Writing—Review and Editing, B.R.-R.; Visualization, A.F.-N.; Supervision, B.R.-R.; Project Administration, MSF; and Funding Acquisition, MSF. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by European Union’s Horizon 2020 research and innovation programme, under grant agreement 776045.

**Acknowledgments:** The authors are grateful to MySustainableForest stakeholders from Centre Nationale de la Propriété Forestière (CNPF), Croatian Forest Research Institute (CFRI), Forest Owners Association of Lithuania (FOAL), Forest Owners Association of Navarra (FORESNA-ZURGAIA), Instituto de Investigação da Floresta e Papel (RAIZ), Madera Plus Calidad Forestal (MADERA+) and University Forest Enterprise Masaryk Forest Křtiny of Mendel University in Brno (UFE) (<https://mysustainableforest.com>).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Pause, M.; Schweitzer, C.; Rosenthal, M.; Keuck, V.; Bumberger, J.; Dietrich, P.; Heurich, M.; Jung, A.; Lausch, A. In situ/remote sensing integration to assess forest health—a review. *Remote Sens.* **2016**, *8*, 471. [[CrossRef](#)]
2. Masek, J.G.; Hayes, D.J.; Joseph Hughes, M.; Healey, S.P.; Turner, D.P. The role of remote sensing in process-scaling studies of managed forest ecosystems. *For. Ecol. Manag.* **2015**, *355*, 109–123. [[CrossRef](#)]
3. Trumbore, S.; Brando, P.; Hartmann, H. Forest health and global change. *Science* **2016**, *349*. [[CrossRef](#)] [[PubMed](#)]

4. White, J.C.; Coops, N.C.; Wulder, M.A.; Vastaranta, M.; Hilker, T.; Tompalski, P. Remote Sensing Technologies for Enhancing Forest Inventories: A Review. *Can. J. Remote Sens.* **2016**, *42*, 619–641. [[CrossRef](#)]
5. Forest Europe; UNECE; F.A.O. *FOREST EUROPE: State of Europe's Forests 2015*; Liaison Unit: Madrid, Spain, 2015.
6. Bonan, G.B. Forests and climate change: Forcings, feedbacks, and the climate benefits of forests. *Science* **2008**, *320*, 1444–1449. [[CrossRef](#)]
7. Seidl, R.; Thom, D.; Kautz, M.; Martin-Benito, D.; Peltoniemi, M.; Vacchiano, G.; Wild, J.; Ascoli, D.; Petr, M.; Honkaniemi, J.; et al. Forest disturbances under climate change. *Nat. Clim. Chang.* **2017**, *7*, 395–402. [[CrossRef](#)]
8. Lindner, M.; Fitzgerald, J.B.; Zimmermann, N.E.; Reyer, C.; Delzon, S.; van der Maaten, E.; Schelhaas, M.-J.; Lasch, P.; Eggers, J.; van der Maaten-Theunissen, M.; et al. Climate change and European forests: What do we know, what are the uncertainties, and what are the implications for forest management? *J. Environ. Manag.* **2014**, *146*, 69–83. [[CrossRef](#)]
9. Milad, M.; Schaich, H.; Bürgi, M.; Konold, W. Climate change and nature conservation in Central European forests: A review of consequences, concepts and challenges. *For. Ecol. Manag.* **2011**, *261*, 829–843. [[CrossRef](#)]
10. Holmgren, P.; Thuresson, T. Satellite remote sensing for forestry planning—A review. *Scand. J. For. Res.* **1998**, *13*, 90–110. [[CrossRef](#)]
11. Boyd, D.S.; Danson, F.M. Satellite remote sensing of forest resources: Three decades of research development. *Prog. Phys. Geogr. Earth Environ.* **2005**, *29*, 1–26. [[CrossRef](#)]
12. Wood, J.E.; Gillis, M.D.; Goodenough, D.G.; Hall, R.J.; Leckie, D.G.; Luther, J.E.; Wulder, M.A. Earth Observation for Sustainable Development of Forests (EOSD): Project overview. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Toronto, ON, Canada, 24–28 June 2002; Volume 3, pp. 1299–1302.
13. Wulder, M.A.; White, J.C.; Nelson, R.F.; Næsset, E.; Ørka, H.O.; Coops, N.C.; Hilker, T.; Bater, C.W.; Gobakken, T. Lidar sampling for large-area forest characterization: A review. *Remote Sens. Environ.* **2012**, *121*, 196–209. [[CrossRef](#)]
14. Lim, K.; Treitz, P.; Wulder, M.; St-Onge, B.; Flood, M. LiDAR remote sensing of forest structure. *Prog. Phys. Geogr. Earth Environ.* **2003**, *27*, 88–106. [[CrossRef](#)]
15. Dubayah, R.O.; Drake, J.B. Lidar Remote Sensing for Forestry. *J. For.* **2000**, *98*, 44–46. [[CrossRef](#)]
16. My Sustainable Forest. Earth Observation Services for Silviculture. Available online: <https://mysustainableforest.com/> (accessed on 30 August 2020).
17. Pekkarinen, A.; Reithmaier, L.; Strobl, P. Pan-European forest/non-forest mapping with Landsat ETM+ and CORINE Land Cover 2000 data. *ISPRS J. Photogramm. Remote Sens.* **2009**, *64*, 171–183. [[CrossRef](#)]
18. Fernandez-Carrillo, A.; de la Fuente, D.; Rivas-Gonzalez, F.W.; Franco-Nieto, A. A Sentinel-2 unsupervised forest mask for European sites. In Proceedings of the SPIE, Strasbourg, France, 9–12 September 2019; Volume 11156.
19. Fassnacht, F.E.; Latifi, H.; Stereńczak, K.; Modzelewska, A.; Lefsky, M.; Waser, L.T.; Straub, C.; Ghosh, A. Review of studies on tree species classification from remotely sensed data. *Remote Sens. Environ.* **2016**, *186*, 64–87. [[CrossRef](#)]
20. Kempeneers, P.; Sedano, F.; Seebach, L.; Strobl, P.; San-Miguel-Ayanz, J. Data Fusion of Different Spatial Resolution Remote Sensing Images Applied to Forest-Type Mapping. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4977–4986. [[CrossRef](#)]
21. Saksa, T.; Uutera, J.; Kolström, T.; Lehtikoinen, M.; Pekkarinen, A.; Sarvi, V. Clear-cut Detection in Boreal Forest Aided by Remote Sensing. *Scand. J. For. Res.* **2003**, *18*, 537–546. [[CrossRef](#)]
22. Lu, D. The potential and challenge of remote sensing-based biomass estimation. *Int. J. Remote Sens.* **2006**, *27*, 1297–1328. [[CrossRef](#)]
23. Gleason, C.J.; Im, J. A Review of Remote Sensing of Forest Biomass and Biofuel: Options for Small-Area Applications. *GIScience Remote Sens.* **2011**, *48*, 141–170. [[CrossRef](#)]
24. Hall, R.J.; Castilla, G.; White, J.C.; Cooke, B.J.; Skakun, R.S. Remote sensing of forest pest damage: A review and lessons learned from a Canadian perspective \*. *Can. Entomol.* **2016**, *148*, S296–S356. [[CrossRef](#)]
25. Foody, G.M. Assessing the accuracy of land cover change with imperfect ground reference data. *Remote Sens. Environ.* **2010**, *114*, 2271–2285. [[CrossRef](#)]

26. Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* **2014**, *148*, 42–57. [[CrossRef](#)]
27. Foody, G.M. Status of land cover classification accuracy assessment. *Remote Sens. Environ.* **2002**, *80*, 185–201. [[CrossRef](#)]
28. Stehman, S.V. Practical Implications of Design-Based Sampling Inference for Thematic Map Accuracy Assessment. *Remote Sens. Environ.* **2000**, *72*, 35–45. [[CrossRef](#)]
29. Stehman, S.V. Statistical rigor and practical utility in thematic map accuracy assessment. *Photogramm. Eng. Remote Sens.* **2001**, *67*, 727–734.
30. Padilla, M.; Olofsson, P.; Stehman, S.V.; Tansey, K.; Chuvieco, E. Stratification and sample allocation for reference burned area data. *Remote Sens. Environ.* **2017**, *203*, 240–255. [[CrossRef](#)]
31. Fernandez-Carrillo, A.; Belenguer-Plomer, M.A.; Chuvieco, E.; Tanase, M.A. Effects of sample size on burned areas accuracy estimates in the Amazon Basin. In Proceedings of the SPIE—The International Society for Optical Engineering, Berlin, Germany, 10–13 September 2018; Volume 10790.
32. Langanke, T.; Herrmann, D.; Ramminger, G.; Buzzo, G.; Berndt, F. Copernicus Land Monitoring Service – High Resolution Layer Forest. Available online: <https://land.copernicus.eu/user-corner/technical-library/hrl-forest> (accessed on 20 July 2020).
33. Martone, M.; Rizzoli, P.; Wecklich, C.; González, C.; Bueso-Bello, J.L.; Valdo, P.; Schulze, D.; Zink, M.; Krieger, G.; Moreira, A. The global forest/non-forest map from TanDEM-X interferometric SAR data. *Remote Sens. Environ.* **2018**, *205*, 352–373. [[CrossRef](#)]
34. Esch, T.; Heldens, W.; Hirner, A.; Keil, M.; Marconcini, M.; Roth, A.; Zeidler, J.; Dech, S.; Strano, E. Breaking new ground in mapping human settlements from space—The Global Urban Footprint. *ISPRS J. Photogramm. Remote Sens.* **2017**, *134*, 30–42. [[CrossRef](#)]
35. Padilla, M.; Stehman, S.V.; Ramo, R.; Corti, D.; Hantson, S.; Oliva, P.; Alonso-Canas, I.; Bradley, A.V.; Tansey, K.; Mota, B.; et al. Comparing the accuracies of remote sensing global burned area products using stratified random sampling and estimation. *Remote Sens. Environ.* **2015**, *160*, 114–121. [[CrossRef](#)]
36. Viana-Soto, A.; Aguado, I.; Salas, J.; García, M. Identifying post-fire recovery trajectories and driving factors using landsat time series in fire-prone mediterranean pine forests. *Remote Sens.* **2020**, *12*, 1499. [[CrossRef](#)]
37. European Commission Joint Research Centre. Forest Type Map 2006. Available online: <http://data.europa.eu/89h/62ec23aa-2d47-4d85-bc81-138175cdf123> (accessed on 20 July 2020).
38. FAO Global Forest Resources Assessment 2020: Terms and Definition. Available online: <http://www.fao.org/forest-resources-assessment/2020> (accessed on 20 July 2020).
39. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
40. Tucker, C.J.; Elgin, J.H.; McMurtrey, J.E.; Fan, C.J. Monitoring corn and soybean crop development with hand-held radiometer spectral data. *Remote Sens. Environ.* **1979**, *8*, 237–248. [[CrossRef](#)]
41. Jiang, Z.; Huete, A.R.; Didan, K.; Miura, T. Development of a two-band enhanced vegetation index without a blue band. *Remote Sens. Environ.* **2008**, *112*, 3833–3845. [[CrossRef](#)]
42. Haboudane, D.; Miller, J.R.; Tremblay, N.; Zarco-Tejada, P.J.; Dextraze, L. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sens. Environ.* **2002**, *81*, 416–426. [[CrossRef](#)]
43. Rikimaru, A.; Roy, P.S.; Miyatake, S. Tropical forest cover density mapping. *Trop. Ecol.* **2002**, *43*, 39–47.
44. Crist, E.P.; Cicone, R.C. A Physically-Based Transformation of Thematic Mapper Data—The TM Tasseled Cap. *IEEE Trans. Geosci. Remote Sens.* **1984**, *GE-22*, 256–263. [[CrossRef](#)]
45. Haralick, R.M. Statistical and structural approaches to texture. *Proc. IEEE* **1979**, *67*, 786–804. [[CrossRef](#)]
46. Browne, M.W. Cross-Validation Methods. *J. Math. Psychol.* **2000**, *44*, 108–132. [[CrossRef](#)]
47. Vidal, C.; Alberdi, I.; Redmond, J.; Vestman, M.; Lanz, A.; Schadauer, K. The role of European National Forest Inventories for international forestry reporting. *Ann. For. Sci.* **2016**, *73*, 793–806. [[CrossRef](#)]
48. McRoberts, R.E.; Tomppo, E.O. Remote sensing support for national forest inventories. *Remote Sens. Environ.* **2007**, *110*, 412–419. [[CrossRef](#)]
49. Tomppo, E.; Gschwantner, T.; Lawrence, M.; McRoberts, R.E.; Gabler, K.; Schadauer, K.; Vidal, C.; Lanz, A.; Ståhl, G.; Cienciala, E. National forest inventories. *Pathw. Common Rep. Eur. Sci. Found.* **2010**, *1*, 541–553.
50. Biging, G.S.; Congalton, R.G.; Murphy, E.C. *A Comparison of Photointerpretation and Ground Measurements of Forest Structure*; American Congress on Surveying and Mapping and American Soc for Photogrammetry and Remote Sensing: Baltimore, MD, USA, 25–29 March 1991.

51. Cochran, W.G. *Stratified Random Sampling*; John Wiley & Sons Inc.: Toronto, Canada, 1977; ISBN 047116240X.
52. Stehman, S.V. Impact of sample size allocation when using stratified random sampling to estimate accuracy and area of land-cover change. *Remote Sens. Lett.* **2012**, *3*, 111–120. [CrossRef]
53. Guyot, G. Optical properties of vegetation canopies. In *Applications of remote sensing in agriculture.*; Steven, M.D., Clark, J.A., Eds.; Butterworth-Heinemann: London, UK, 1990; pp. 19–44. ISBN 0408047674.
54. Pacheco-Labrador, J.; El-Madany, T.S.; Martín, M.P.; Migliavacca, M.; Rossini, M.; Carrara, A.; Zarco-Tejada, P.J. Spatio-temporal relationships between optical information and carbon fluxes in a mediterranean tree-grass ecosystem. *Remote Sens.* **2017**, *9*, 608. [CrossRef]
55. Vlassova, L.; Perez-Cabello, F.; Nieto, H.; Martín, P.; Riaño, D.; Riva, J.D. La Assessment of methods for land surface temperature retrieval from landsat-5 TM images applicable to multiscale tree-grass ecosystem modeling. *Remote Sens.* **2014**, *6*, 4345–4368. [CrossRef]
56. Moore, C.E.; Beringer, J.; Evans, B.; Hutley, L.B.; Tapper, N.J. Tree–grass phenology information improves light use efficiency modelling of gross primary productivity for an Australian tropical savanna. *Biogeosciences* **2017**, *14*, 111–129. [CrossRef]
57. Liu, Y.; Hill, M.J.; Zhang, X.; Wang, Z.; Richardson, A.D.; Hufkens, K.; Filippa, G.; Baldocchi, D.D.; Ma, S.; Verfaillie, J.; et al. Using data from Landsat, MODIS, VIIRS and PhenoCams to monitor the phenology of California oak/grass savanna and open grassland across spatial scales. *Agric. For. Meteorol.* **2017**, 237–238, 311–325. [CrossRef]
58. Whiteside, T.G.; Boggs, G.S.; Maier, S.W. Comparing object-based and pixel-based classifications for mapping savannas. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 884–893. [CrossRef]
59. Arnot, C.; Fisher, P.F.; Wadsworth, R.; Wellens, J. Landscape metrics with ecotones: Pattern under uncertainty. *Landsc. Ecol.* **2004**, *19*, 181–195. [CrossRef]
60. SIRS. GMES Initial Operations/Copernicus Land monitoring services—Validation of products: HRL Forest 2015 Validation Report. Available online: [https://www.google.com.hk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjLuq\\_F1oXsAhVGXsSKHWOeBvEQFjAAegQIBxAB&url=https%3A%2F%2Fland.copernicus](https://www.google.com.hk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjLuq_F1oXsAhVGXsSKHWOeBvEQFjAAegQIBxAB&url=https%3A%2F%2Fland.copernicus) (accessed on 20 July 2020).
61. McHugh, M.L. Interrater reliability: The kappa statistic. *Biochem. Medica* **2012**, *22*, 276–282. [CrossRef]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).