

GMES Initial Operations / Copernicus Land monitoring services – Validation of products

Validation Services for the geospatial products of the
Copernicus land Continental and local components
including in-situ data (lot 1)

Open Call for Tenders - EEA/MDI/14/010

ANALYSIS OF THE MIXED FOREST IN THE 20M HRL FTY FINAL VALIDATION REPORT



Consortium Partners

Prime Contractor	SIRS SAS (France)
Main Partners	GAF AG (Germany), Joanneum Research (Austria), e-GEOS (Italy)
Experts	Specto Natura (UK), ISPRA (Italy), IGN France International (France), Terranea (Germany), NLSI (Iceland)

Document Preparation and Release

	Affiliation	Name(s)	Date
Author	SIRS	Alexandre Pennec	28/11/2018
Contributions	SIRS	Renaud Vandeputte, Christophe Sannier	28/11/2018
	JR	Janik Deutscher	03/01/2019
	GAF	André Stumpf	08/01/2019
Review	SIRS	Christophe Sannier	07/01/2019
Approval	EEA	Hans Dufourmont	

Document Issue Record

Issue	Date	Author(s)	Description of Change
1.0	27/11/2018	Renaud Vandeputte	Document creation and Integration of results for FTY Mixed Study
1.1	28/11/2018	Alexandre Pennec	First final version
1.2	03/01/2019	Janik Deutscher	Review and comments
1.3	08/01/2019	André Stumpf	Review and comments
1.4	07/02/2019	Alexandre Pennec	Second final version with the integration of JR's comment

Applicable Documents

ID	Title
AD01	Tender Specifications « Open Call for Tenders n° EEA/MDI/14/010 - Framework service contract for the GMES Initial Operations / Copernicus Land monitoring services – Validation Services for the geospatial products of the Copernicus land Continental and local components including in-situ data (lot 1)”
AD02	Framework service contract No EEA/MDI/14/010
AD03	First Specific Contract N° 3436/B2015/R0-GIO/EEA.56159
AD04	HRL Forest Product Definitions
AD05	Second Specific contract N°3436/R0-COPERNICUS/EEA.56601
AD06	Third Specific Contract N°3436/R0-COPERNICUS/EEA.57056

Executive Summary

The High Resolution layer (HRL) 2012 on Tree cover Density (TCD) and Forest Type (FTY) was validated as part of the HRL Forest exercise during the Second Specific Contract¹.

If the results show that the 100m HRL TCD product meets the thematic classification accuracy requirements, the analysis for the 100m FTY product was more ambivalent since the product as it stands based on the reference data collected appeared not to fully meet the thematic classification accuracy requirements.

Indeed, based on the results available for the 2012 FTY product evaluation, substantial confusion was highlighted between forest types and particularly with the mixed forest type. When performing a plausibility analysis at Pan-European level, substantial improvement in the accuracy values was obtained for the non-mixed classes, but the accuracy of the mixed class remained low. In fact, technically, the presence of the mixed class is purely based on aggregation rule. This class only exists at 100m resolution and is not present in the 20m forest products (20m FTY and 20m FADSL) on which the 100m pixel is calculated and which can already include substantial mixed forest. Even with a plausibility analysis implemented to provide a better understanding of the causes of the classification errors, the relatively poor thematic accuracy of the mixed forest class was still confirmed. This was also confirmed for the HRL2015 FTY product for which the mixed class still exhibit lower accuracy compared to the other Forest Types.

Several reasons were at the time elaborated in the report and needed to be confirmed by a separate study with additional analysis to assess the accuracy of the 20m 2012 FTY layer over areas classified as mixed and non-mixed forest in the 100m layer. This is presumably due to the fact that the mixed class is typically associated with transition areas between broadleaved and coniferous forest whereas it is well known that transition areas exhibits lower accuracies than in homogeneous areas. Therefore, an additional analysis is requested to assess the accuracy of the 20m 2012 FTY layer over areas classified as mixed and non-mixed forest in the 100 m layer.

So, the current deliverable for this task comprises:

- the 20m FTY assessment (like the approach taken for the 100m assessment of the 2nd specific contract)
- A plausibility assessment based on the Forest Type nomenclature (broadleaf and coniferous; no mixed class) at 20m SSU level for the mixed plots of the 2012 100m FTY product identified during previous Service Contracts
- A comparison of accuracy results between the 100m and 20m assessment
- An interpretation of the results, conclusions and recommendations.

The analysis confirmed the low results obtained on the areas covered by mixed forest compared to broadleaved or coniferous forest areas. Therefore, it is suggested that the mixed class should be renamed to a "transition zones" class which would make it possible to keep the current product specifications.

¹ See <https://land.copernicus.eu/user-corner/technical-library/hrl-forest-2012-validation-report-1>

Table of Contents

.....	0
1. Validation Framework.....	8
1.1. Products to be validated.....	8
1.2. Validation Criteria.....	9
2. Validation approach	10
2.1. Thematic Accuracy.....	10
2.1.1. Level of reporting	10
2.1.2. Stratification and sample design.....	10
2.1.3. Response Design.....	13
2.1.4. Estimation and analyses procedures.....	13
2.1.5. HRL FTY 2012 100m thematic accuracy	15
3. Mixed Forest Analysis.....	16
4. Conclusions and recommendations	19

List of Figures

Figure 1: Example of SSUs organised in a 5x5 20m grid	11
Figure 2: Error matrices at pan-European level for the 2012 HRL FTY product for (a) blind interpretation and (b) plausibility analysis.	15
Figure 3: Extract of the HRL FTY 100m product illustrating the spatial distribution of the mixed forest class almost always occurring in transitional areas between coniferous and broadleaf forest types	16
Figure 4: FTY product thematic accuracy results per thematic classes for mixed class sample SSUs (20m).....	17
Figure 5: FTY product thematic accuracy results per thematic classes for (a) mixed class sample PSUs based on aggregated SSU interpretation (b) non – mixed class PSUs.	18

List of Tables

Table 1: Detailed specification of the 100m Forest Type product.....	9
Table 2: FTY Distribution of PSUs per main strata and substrata	12
Table 3: FTY Weight factor to be applied to each stratum and substratum for constructing confusion matrices	14

List of Abbreviations

BRME	Biogeographical Regions Map of Europe
CLC	CORINE Land Cover
DLT	Dominant Leaf Type
EEA	European Environment Agency
EU-DEM	Digital Elevation Model over Europe
ESA	European Spatial Agency
FAO	Food and Agriculture Organization of the United Nations
FTY	Forest Type
GIO	GMES Initial Operations
GMES	Global Monitoring for Environment and Security
HRL	High Resolution Layer
JRC	Joint Research Centre
LAEA	Lambert Azimuthal Equal-Area
LUCAS	Land Use/Cover Area frame Survey
MMU	Minimum Mapping Unit
MMW	Minimum Mapping Width
PSU	Primary Sample Unit
SP	Service Provider
SSU	Secondary Sample Unit
TCD	Tree Cover Density
UA	Urban Atlas

1. Validation Framework

The validation framework is defined by a comprehensive analysis of the product specifications to determine the criteria to be used for the validation exercise.

1.1. Products to be validated

Pan-European High Resolution Layers (HRL) provide information on specific land cover characteristics, and are complementary to land cover / land use mapping such as in the CORINE land cover (CLC) datasets. The HRLs are produced from 20 m resolution satellite imagery through a combination of automatic processing and interactive rule-based classification.

Five themes have been identified in **HRL 2012**, corresponding with the main themes from CLC, i.e. the level of sealed soil (imperviousness), Forest (tree cover density and **forest type**), natural grasslands, wetlands and water bodies. Pixels of 20 by 20 m are aggregated into 100 by 100 m grid cells for final products.

One HRL product is specifically addressed in this study: the **100m Forest Type (FTY)** layer part of the Forest products.

The **Forest Type** product allows to get as close as possible to the FAO forest definition. In its original (20m) resolution it consists of two products: 1) a dominant leaf type product that has a MMU of 0.5 ha, as well as a 10% tree cover density threshold applied, and 2) a support layer that maps, based on the dominant leaf type product, trees under agricultural use and in urban context (derived from CLC and imperviousness 2009 data). For the final 100m product trees under agricultural use and urban context from the support layer are removed.

20m pixel size status layers have undergone a verification and enhancement process by member states and/or service providers. HRL products should be validated as integrated pan-European mosaics at 100m pixel size. The following products should be validated:

- Forest type (FTY), 100m x 100m, European projection, LAEA

However, the 20m FTY product will be also evaluated as part of this study to provide more insight in the lower classification accuracy found in the validation of the HRL 2012 FTY 100m product²

² See <https://land.copernicus.eu/user-corner/technical-library/hrl-forest-2012-validation-report-1>

1.2. Validation Criteria

For the 100 m product, the overall target thematic accuracy is 90% with 85% considered as absolute minimum accuracy. The 90% accuracy value must be understood as follows: 10% for commission errors and 10% for omission errors for forest classes (either broadleaved, coniferous or mixed forests).

For the 20m, no requirements were planned.

The detailed specifications of the HRL 2012 Forest type 100m product is shown below:

Table 1: Detailed specification of the 100m Forest Type product

Product
<i>Forest type 2012, 100m x 100m, European projection</i>
Methodology
<i>Aggregation of the 20m x 20m forest type product: see Ch. 4 (forest type). 100m grid cells with a majority of underlying pixels indicating trees in agricultural use or urban context (from the additional support raster) are treated as Non-Forest.</i>
Geometric raster size
<i>Grid cell size 100m x 100m</i>
Coordinate Reference System
<i>European projection, LAEA</i>
Geometric accuracy (positioning scale)
<i>According to the ortho-rectified satellite imagery delivered by ESA (AD04)</i>
Thematic accuracy (in %)
<i>Verification: not planned</i>
<i>Validation: Overall target accuracy: 90% (85% is considered as absolute minimum accuracy). The 90% accuracy value has to be understood as follows: 10% for commission errors and 10% for omission errors. The validation will not be done by the SPs in the frame of the project.</i>
Data type
<i>Raster</i>
Raster coding
<i>Thematic pixel values</i>
<i>0: all non-forest areas (incl. trees in agricultural and urban context)</i>
<i>1: broadleaf forest</i>
<i>2: coniferous forest</i>
<i>3: mixed forest</i>
<i>254: unclassifiable (no satellite image available, or clouds, shadows, or snow)</i>
<i>255: outside area</i>
Metadata
<i>According to INSPIRE metadata standards</i>

2. Validation approach

The approach implemented as part of the validation of the HRL 2012 Forest products³ is summarized here and provides guidance on how the products were validated by defining suitable indicators or metrics. Detailed completeness and logical consistency checks were performed as part of the semantic checks undertaken by ETC ULS for most products.

The aim of this study is not to repeat the analysis conducted during the SC02, but to perform additional analysis to assess the accuracy of the 20m FTY layer over areas classified as mixed and non-mixed forest in the 100m layer.

2.1. Thematic Accuracy

2.1.1. Level of reporting

The level of reporting for the validation results of the study is at pan-European level. No further disaggregation of the results was performed as these are already presented in the HRL Forest 2012 validation report.

2.1.2. Stratification and sample design

2.1.2.1. Overview

A stratified systematic sampling approach based on LUCAS is used for all thematic layers adapting the number of replicates to each stratum. The LUCAS sampling is densified for small strata. Using LUCAS sampling ensures coherence between the different layers and traceability.

For the 100m FTY product, a stratification is applied at two levels:

1. Stratification according to countries or group of countries with an area greater than 90,000km²
2. Stratification based on a series of omission/commission strata

The number of primary sample units (PSUs) per stratum should be such to ensure a sufficient level of precision at reporting level. The minimum number of PSUs per stratum should be set at 50 if possible. Priority is given to strata which are known to be difficult to map: e.g. changes and difficult classes.

The validation exercise covers the whole study area to be valid (e.g. use of low and high probability omission strata for HRL with low sampling intensity in low probability stratum).

Each PSU corresponds to one 100m HRL pixel. Each PSU is then associated to secondary sampling units (SSUs) corresponding to a 5x5 grid with 20m between each SSU (Figure 1). The idea is that each SSU can then be associated with the corresponding HRL 20m layer pixel.

³ <https://land.copernicus.eu/user-corner/technical-library/hrl-forest-2012-validation-report-1>



Figure 1: Example of SSUs organised in a 5x5 20m grid

Sampling units will be different for each layer due to different stratification approaches, but some sample locations will be shared thanks to using LUCAS as the basis for selecting sample units

The first level of stratification was defined according to countries or group of countries with an area greater than 90,000 km². The second level stratification was defined as follows:

- Commission Low Probability: Forest Type 1-2-3 & CLC forest classes (minimum of 50 PSUs per country / group of countries)
- Commission High Probability: Forest Type 1-2-3 & CLC non forest classes (minimum of 50 PSUs per country / group of countries)
- Omission High Probability: Forest Type 0 & CLC forest classes (minimum of 150 sample units per country / group of countries)
- Omission Low Probability: Rest of the area (minimum of 50 PSUs per country / group of countries)

Different sampling intensity were applied to focus on strata for which there is a higher probability that errors will be found. Weighting factors were calculated based on the final sample selected to ensure that the different sampling intensities are accounted for when constructing confusion matrices in order to avoid the introduction of a bias toward these strata.

CLC forest classes were defined as follows based on CLC2006:

- 2.4.4 = agroforestry
- 3.1.1 = broadleaved forest
- 3.1.2 = coniferous forest
- 3.1.3 = mixed forest
- 3.2.4 = transitional woodland, shrub

There was a total of 17,297 PSUs initially selected for the FTY product (Table 2). For this analysis, a total of 20,750 SSUs were selected to be re-interpreted corresponding to 830 PSUs (469 identified omissions and 361 as commission for the mixed class in the HRL FTY 2012 100m product).

Table 2: FTY Distribution of PSUs per main strata and substrata

LABEL	Commission		Omission		Total
	High	Low	High	Low	
AL+ME+MK+RS+XK	79	111	272	111	573
AT + CH + LI	75	99	227	99	500
BA + HR + SI	73	103	231	100	507
BE + LU+ NL + DK	69	95	225	95	484
BG	72	94	229	94	489
CZ + SK	75	100	229	100	504
DE	120	189	404	189	902
EE + LT + LV	84	121	272	118	595
EL + CY	78	105	260	105	548
ES	247	149	543	247	1,186
FI	145	182	413	178	918
FR	155	264	565	262	1,246
FR DOMs	76	80	214	88	458
HU	69	87	199	87	442
IE + UK	100	173	377	184	834
IS	65	86	207	90	448
IT + MT	109	167	384	167	827
NO	111	176	413	176	876
PL	111	172	341	172	796
PT	68	86	222	86	462
RO	97	142	332	143	714
SE	138	224	743	219	1,324
TR	202	353	756	353	1,664
TOTAL	2,418	3,358	8,058	3,463	17,297

2.1.3. Response Design

The sample design was based on the LUCAS sample frame and some sample units will coincide with some LUCAS points. However, the LUCAS points were re-interpreted based on available in situ data. LUCAS thematic information was not used directly.

The approach adopted in this study was to use the sample units already interpreted as part of the HRL FTY 2012 100m validation and compare them with the HRL FTY 2012 20m layer and isolate sample units in disagreement for both omission and commission errors. These were then re-interpreted in a plausibility analysis considering the information contained in the map layer.

The PSUs from the HRL FTY 2012 layer for the 20m product were evaluated according to the following definitions and aggregation rules used⁴:

- MMU 0.5 ha (i.e. 12.5 20m x 20m cells)
- 0-10% TCD = non-forest and TCD $\geq 10\%$ = forest
- MMW 20m

3 categories:

- Non-forest = 0
- Broadleaved Forest = 1
- Coniferous Forest = 2

It **includes** (according to FAO) forest nurseries and seed orchards that constitute an integral part of the forest, as well as forest roads, cleared tracts, firebreaks and other small open areas < 0.5 ha and/or < 20 m width. Forest in national parks, nature reserves and other protected areas such as those of specific scientific, historical, cultural or spiritual interest; windbreaks and shelterbelts of trees with an area of more than 0.5 ha and width of more or equal than 20 m; plantations primarily used for forestry purposes, including cork oak stands. Tree cover in traditional agroforestry system such as Dehesa / Montado is included.

It **excludes** (according to FAO) land predominantly used for agricultural practices. In this sense fruit trees and olive groves are also excluded. Gardens and urban parks are also not considered as forest.

For each SSU, the interpreter first identified whether the area could be considered as forest or not based on the definition described above by taking into account the surrounding areas and then when the SSU was identified as forest. As this was performed in a plausibility analysis, in its final decision, the interpreter also took into consideration the information from the HRL FTY 2012 20m map layer accepting the map classification when this was considered as plausible.

2.1.4. Estimation and analyses procedures

Thematic accuracy should be presented in the form of an error matrix. Unequal sampling intensity resulting from the stratified systematic sampling approach should be accounted for by applying a weight factor (p) to each sample unit based on the ratio between the number of samples and the size of the stratum considered:

$$\hat{p}_{ij} = \left(\frac{1}{N}\right) \sum_{x \in (i,j)} \frac{1}{\pi_{uh}^*}$$

Where i and j are the columns and rows in the matrix, N is the total number of possible units (population) and π is the sampling intensity for a given stratum. Based on this formula, weights were computed and applied to each stratum as shown in (Table 3):

⁴ <https://land.copernicus.eu/user-corner/technical-library/hrl-forest>

Table 3: FTY Weight factor to be applied to each stratum and substratum for constructing confusion matrices

LABEL	Commission		Omission	
	High	Low	High	Low
AL+ME+MK+RS+XK	1.95283933	0.28928673	0.15002733	2.15554249
AT + CH + LI	1.79489227	0.16374951	0.06548584	2.10452251
BA + HR + SI	2.35044509	0.32611892	0.10145303	1.53362087
BE + LU+ NL + DK	0.52715786	0.11841538	0.03857454	2.98337816
BG	1.49123689	0.14454644	0.08075001	2.04245338
CZ + SK	1.78536819	0.16266343	0.06110505	2.17608997
DE	2.40560103	0.24533189	0.06544269	3.71637624
EE + LT + LV	2.38322555	0.31395054	0.12408021	2.10773358
EL + CY	1.10732109	0.33136108	0.11110260	2.58590230
ES	1.49596098	0.93013556	0.21152256	3.58988059
FI	4.31379113	0.25107797	0.16137167	1.23129404
FR	2.46870694	0.29354368	0.14388578	4.18985237
FR DOMs	2.42656144	0.03016413	0.01850080	0.12669669
HU	0.68808972	0.05426091	0.06614315	2.44399048
IE + UK	0.57038033	0.19086100	0.08775483	4.42796084
IS	0.00311564	0.00186691	0.00728586	2.38608894
IT + MT	2.18486952	0.52649507	0.08610975	3.22702648
NO	1.99003065	0.27085622	0.21709962	2.91158698
PL	2.35975246	0.22433530	0.08026976	3.50556904
PT	0.86124260	0.11391927	0.29099300	1.63989818
RO	2.04593371	0.12908326	0.08065827	3.26689047
SE	4.98647649	0.24885208	0.21052241	1.67664679
TR	1.79088293	0.28359485	0.29281187	4.64654157

Overall accuracy and User and producer accuracy should be computed for all thematic classes and 95% confidence intervals should be calculated for each accuracy.

The standard error of the error rate can be calculated as follows: $\sigma_h = \sqrt{\frac{p_h(1-p_h)}{n_h}}$ where n_h is the sample size for stratum h and p_h is the expected error rate. The standard error is calculated for each stratum and an overall standard error is calculated based on the following formula:

$$\sigma = \sqrt{\sum w_h^2 \cdot \sigma_h^2}$$

In which w_h is the proportion of the total area covered by each stratum. The 95% confidence interval is +/- 1.96. σ .

2.1.5. HRL FTY 2012 100m thematic accuracy

The following section provides a summary of the final results obtained as part of the HRL FTY 2012 100m validation. A detailed plausibility analysis was conducted for the overall data set with particular attention given to the correct application of the aggregation procedure especially for the mixed class. The interpretation was conducted based on the SSUs (20m cells) the aggregation was performed using a simplified rule of the one used for the map production. The approach is based on a majority aggregation done on valid pixels without considering the Forest Additional Support Layer since the study especially focuses on areas already mapped as forest. The results are shown in Figure 2 and show a substantial improvement compared with the corresponding blind interpretation. The main reason for this improvement is, that many mixed plots were formerly considered as unreliable due to a lack of reliable reference data for 2012, but new reference data is now available to interpret these plots more accurately. Nevertheless, the accuracy of the Mixed forest class is still relatively low whilst the broadleaf class exhibits good producer and lower user accuracy and the coniferous class both satisfactory producer and user accuracies. A potential explanation is that the mixed class is typically associated with transition areas between broadleaved and coniferous forest as illustrated in Figure 3 and it is well known that transition areas are likely to exhibit lower accuracies than in homogeneous areas. The analysis conducted here clearly shows that there is an issue with the mixed forest class as defined in the HRL FTY 100m product.

(a)	Non-forest	Broadleaved	Coniferous	Mixed forest	Total	Overall User accuracy	User accuracy between Forest types
Non-forest	10804	352	114	124	11394		
Broadleaved	629	1600	127	462	2818	56.8%	73.1%
Coniferous	215	144	1336	591	2287	58.4%	64.5%
Mixed forest	158	100	202	338	798	42.4%	52.9%
Total	11806	2196	1779	1515	17297		
Overall Producer accuracy		72.9%	75.1%	22.3%		Overall Accuracy	81.4%
Producer accuracy between Forest types		86.8%	80.3%	24.3%			
(b)	Non-forest	Broadleaved	Coniferous	Mixed forest	Total	Overall User accuracy	User accuracy between Forest types
Non-forest	11063	215	84	32	11394		
Broadleaved	421	2125	87	185	2818	75.4%	88.6%
Coniferous	124	63	1962	137	2287	85.8%	90.8%
Mixed forest	98	39	135	526	798	66.0%	75.2%
Total	11706	2442	2269	881	17297		
Overall Producer accuracy		87.0%	86.5%	59.8%		Overall Accuracy	90.6%
Producer accuracy between Forest types		95.5%	89.8%	62.0%			

Figure 2: Error matrices at pan-European level for the 2012 HRL FTY product for (a) blind interpretation and (b) plausibility analysis.

It is also worth noting that after the plausibility analysis, the user and producer accuracy of the broadleaf and coniferous type are very close or exceeding the target accuracy when only considering the confusion between

forest types. This would suggest that the assessment of the HRL FTY layer could be done in two stages, first by assessing the accuracy of the forest mask and second by assessing the accuracy of the forest types within the forest mask. It is noticeable that most forest user and producer accuracy values are lower than the minimum accuracy threshold of 85%. However, most of the confusion is between the forest classes.

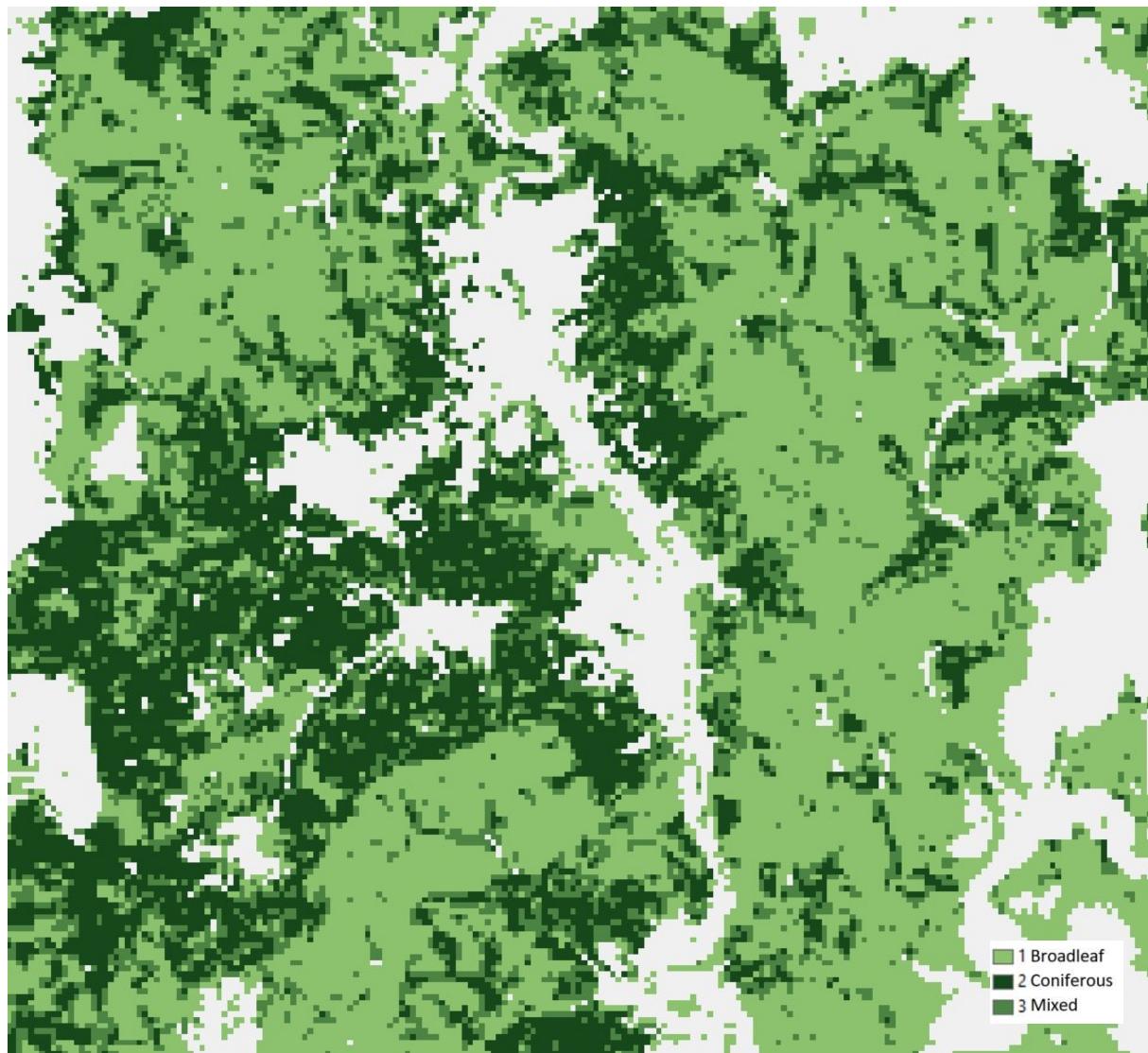


Figure 3: Extract of the HRL FTY 100m product illustrating the spatial distribution of the mixed forest class almost always occurring in transitional areas between coniferous and broadleaf forest types

3. Mixed Forest Analysis

To further explain the low precision rate of the mixed class, a specific analysis was performed on this thematic class as part of this study focusing on the input HRL FTY 2012 20m layer that was used. Therefore, the aim of the study is to perform an additional analysis to assess the accuracy of the 20m FTY layer over areas classified as mixed and non-mixed forest in the 100m layer. Considering that the mixed class is by definition mainly located over transition areas, the hypothesis is that omissions and commissions are more prevalent on the mixed class than for other classes.

The results of the re-interpretation of SSUs located in the mixed forest class and the analysis performed on the 20m HRL FTY 2012 are presented in the error matrix below in Figure 4. They seem to confirm the hypothesis that the accuracy of the underlying forest types from the 20m product is lower for the areas identified as mixed

forest in the 100m layer with user and producer accuracies substantially lower for both the broadleaf and coniferous classes than in the HRL FTY 2012 100m validation (see Figure 2 and Figure 4).

	Non-forest	Broadleaved	Coniferous	Total	Overall User accuracy	User accuracy between Forest types
Non-forest	1459	378	466	2304		
Broadleaved	1161	3408	2170	6738	50,6%	61,1%
Coniferous	704	924	4820	6448	74,8%	83,9%
Total	3323	4710	7456	15489		
Overall Producer accuracy		72,4%	64,6%		Overall Accuracy	62,5%
Producer accuracy between Forest types		78,7%	69,0%			

Figure 4: FTY product thematic accuracy results per thematic classes for mixed class sample SSUs (20m)

To further confirm this, the results at SSU level were re-aggregated at PSU level simply using the majority rule. This was then compared to the results obtained with PSUs that were not identified as mixed forest in the product or the validation. These results are shown below in Figure 5. The results obtained for the mixed class PSUs (Figure 5a) are very similar to that obtained for the 20m analysis with only a slight positive effect on the resulting producer, user and overall accuracies suggesting that the aggregation based on the majority rule did not have any substantial effect on the analysis. The corresponding results for the non-mixed class PSUs (Figure 5b) are substantially better. Several factors could explain this phenomenon such as the fact that the mixed class as defined in the HRL FTY 2012 100m product does indeed correspond to transition areas exhibiting a lower classification accuracy from the input HRL FTY 2012 20m layer. Other factors could also explain the results such as: the overestimation of the broadleaf percentage in mixed forests; the confusion with the non-forest class is more problematic for mixed forests than for pure forests (e.g. for low TCD forests); and in general, one would always expect more confusion between forest types within a mixed forests than within pure forests since both forest types/tree types are part of the mixed forest.

(a)	Non-forest	Broadleaved	Coniferous	Total	Overall User accuracy	User accuracy between Forest types
Non-forest	36	11	42	89		
Broadleaved	40	141	85	265	53,1%	62,5%
Coniferous	24	28	215	266	80,6%	88,6%
Total	100	180	341	621		
Overall Producer accuracy		78,4%	63,0%		Overall Accuracy	63,1%
Producer accuracy between Forest types		83,6%	71,8%			
(b)	Non-forest	Broadleaved	Coniferous	Total	Overall User accuracy	User accuracy between Forest types
Non-forest	11034	214	75	11324		
Broadleaved	415	2062	77	2555	80,7%	96,4%
Coniferous	124	62	1842	2028	90,8%	96,7%
Total	11573	2339	1994	15907		
Overall Producer accuracy		88,2%	92,4%		Overall Accuracy	93,9%

Producer accuracy
between Forest types 97,1% 96,0%

Figure 5: FTY product thematic accuracy results per thematic classes for (a) mixed class sample PSUs based on aggregated SSU interpretation (b) non – mixed class PSUs.

4. Conclusions and recommendations

Based on the results available for the FTY product evaluation, there seemed to be substantial confusion between forest types and particularly with the mixed forest type. This may be due to a number of reasons:

- The aggregation rules for constructing the 100m FTY layer are complex and the integration process from the 20m layer was further investigated in constructing the validation data as part of the plausibility analysis
- The notion of mixed class is problematic and sometimes may introduce more rather than less confusion with the other two classes. Perhaps the dominant forest type would be more appropriate
- Some countries such as in Scandinavia often have very low TCDs and forest type is therefore more difficult to estimate.

A plausibility analysis was conducted at pan-European level. Substantial improvement in the accuracy values is shown, but the accuracy of the mixed class remains low. It was suggested that the mixed class is typically associated with transition areas between broadleaved and coniferous forest and it is well known that transition areas exhibits lower accuracies than in homogeneous areas. Therefore, an additional analysis was performed to assess the accuracy of the 20m FTY layer over areas classified as mixed and non-mixed forest in the 100m layer.

The analysis confirmed that mixed forest areas identified in the product or the reference data for the validation of the HRL FTY 2012 100m exhibited lower accuracies in the 20m HRL FTY product than for areas outside of the mixed class for the broadleaved or coniferous classes.

The mixed forest class as it is defined does seem to pose a problem from a user perspective because it does not necessarily correspond to actual mixed forest but is a result of the aggregation process. In other word, what could be considered as a homogeneous forest stand containing a mix of broadleaf and coniferous trees will either be classified as broadleaf or coniferous forest in the 20m product depending on which forest type is dominant in the 20m cell, resulting in either coniferous or broadleaf pixels in the 100m product. However, two separate stands – one pure broadleaf and one pure coniferous forest - contained within a 100m pixel will result in a mixed forest, which seems to be illogical. Therefore, we would suggest that the mixed class should be dropped or adapted to be more in line with the reality, but this would probably mean introducing a mixed class for the 20m product. Therefore, **it is suggested that the current “mixed forest” class be renamed as “complex forest area” class** (including transition areas or mixed forests or a mix of different pure stands) thus making it possible to keep the existing products as they are and providing more clarity from a user perspective. Alternatively, the creation of a more realistic mixed forest class would probably require the inclusion of this class at the thematic classification stage in the 20m product thus requiring substantial additional processing.