

Copernicus Land monitoring services High Resolution land cover characteristics for the 2018 reference year

Small woody features

Validation report of the final dataset of Small
Woody Features 5m raster product
Deliverable D4.2

Service contract: EEA/DIS/R0/20/006

Issue 1 – 30/11/2021

With the support of:



Internal/Interne

CHRONOLOGY ISSUES

Issue	Date	Object	Written by	Checked by	Approved by
1	29/11/21	Final validation report	Kathrin Schweitzer, Thomas Renner	Nathalie Morin, Loïc Faucqueur	Carlos de Wasseige

DISTRIBUTION

Company	Means of distribution	Names
EEA	Email	Usue Donezar

LIST OF CONTENTS

1	Validation Framework.....	7
1.1	Product to be validated	7
1.2	Validation Criteria	9
2	Internal validation procedure.....	10
2.1	Stratification and Sample Design.....	10
2.2	Response design	14
3	Statistical Analysis.....	16
3.1	Procedure of statistical analysis.....	16
3.1.1	PSU-level	16
3.1.2	SSU-level	17
3.2	Results of statistical analysis.....	17
3.2.1	Confusion Matrix for accuracies at PSU-level.....	17
3.2.2	Confusion Matrix for accuracies at SSU-level.....	19
4	Conclusion.....	20

LIST OF TABLES AND FIGURES

Table 1: Thematic definition of SWF2018	8
Table 2: Geometric specifications of SWF2018.....	8
Table 3: Number of valid Primary Sample Units within EEA39 areas (including DROM) for commission and omission strata excluding invalid samples and samples with 50% and more forest cover, for the final delivery of 282 Production Units.	12
Table 4: Number of PSU containing still valid information (of 2015) and number of PSU re-assessed in the 2018 internal validation, as well as total number of SSU within valid PSU. Numbers are for the final delivery of 282 PU (EEA39: 277PU; DROM: 5PU).	13
Table 5: Unweighted accuracy figures for the SWF2018 5m product at PSU level for EEA39 areas including DROM corresponding to 282 PU of final delivery.....	18
Table 6: Weighted accuracy figures for the SWF2018 5m product at PSU level for EEA39 areas including DROM corresponding to 282 PU of final delivery.	18
Table 7: Unweighted accuracy figures for the SWF2018 5m product at SSU level for EEA39 areas including DROM corresponding to 282 PU of final delivery.....	19
Table 8: Weighted accuracy figures for the SWF2018 5m product at SSU level for EEA39 areas including DROM corresponding to 282 PU of final delivery.	19

LIST OF ACRONYMS

AWF	Additional Woody Features
CLC	Corine Land Cover
DROM	French Overseas Departments and Regions/Départements et Régions d'Outre Mer
EEA	European Environment Agency
EO	Earth Observation
EU-28	European Union members of 27 states including UK
FAO	Food and Agriculture Organization
HRL	High Resolution Layer
ITT	Invitation to Tender
LC/LU	Land Cover/Land Use
LR	Large Region
LUCAS	Land Use and Coverage Area frame statistical Survey
MMWmax	Maximum Mapping Width
MMWmin	Minimum Mapping Width
MMU	Minimum Mapping Unit
OA	Overall Accuracy
OSM	Open Street Map
PA	Producer Accuracy
PSU	Primary Sample Unit
PU	Production Unit
RGB	Real Colour Image, Red-Green-Blue
SDG	Sustainable Development Goals
SU	Sample Unit
SSU	Secondary Sample Unit
SWF	Small Woody Features
TCD	Tree Cover Density
UA	User Accuracy
UN	United Nations
VHR	Very High Resolution
WMS	Web Map Service

APPLICABLE DOCUMENTS

- AD01: ANNEX I – TENDER SPECIFICATIONS: Service contracts for the Copernicus Land monitoring services – High Resolution land cover characteristics for the 2018 reference year - Call for tenders EEA/DIS/R0/20/006
- AD02: Annex 7 to TS_ Detailed description of the services
- AD03: Deliverable 1.1 - Project Management Plan
- AD04: Validation Services for the geospatial products of the Copernicus land Continental and local components including in-situ data (lot 1): HRL SMALL WOODY FEATURES VALIDATION CONCEPT.
https://land.copernicus.eu/user-corner/technical-library/clms_hrl_swf_validation_concept_sc03_1_1-2.pdf

REFERENCE DOCUMENTS

- Büttner G., Kosztra B., Maucha G. and Pataki R. (2012) Implementation and achievements of CLC2006, ETC-LUSI, EEA, 65p.
- Olofsson, P., Foody, G. M., Stehman, S. V., & Woodcock, C. E. (2013). Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, 129, 122-131.
- Selkowitz, D. J., & Stehman, S. V. (2011). Thematic accuracy of the National Land Cover Database (NLCD) 2001 land cover for Alaska. *Remote Sensing of Environment*, 115(6), 1401-1407.
- Stehman, S. V., & Czaplewski, R. L. (1998). Design and analysis for thematic map accuracy assessment: fundamental principles. *Remote sensing of environment*, 64(3), 331-344.
- Van der Zanden, E. H., Verburg, P. H., & Mücher, C. A. (2013). Modelling the spatial distribution of linear landscape elements in Europe. *Ecological indicators*, 27, 125-136

1 Validation Framework

As described in the ITT documents and particularly in the Tender Specifications (AD01) and its Annex 7 (AD02) and in Deliverable 1.1 - Project management plan (AD03), this report refers to the internal validation activity linked to the final delivery of the Small Woody Feature 2018 (SWF2018) 5m raster product.

The aim of this document is to inform about the procedures for internal validation and accuracy assessment within service contract (EEA/DIS/R0/20/006) and to provide results for the final delivery of the SWF2018 5m raster product. The main measure of the products' validity and quality is the thematic accuracy, which is assessed by comparing the mapped information within the product with reference data at selected representative sample locations. This approach has three components:

- (i) the stratification and sampling design,
- (ii) the suggested response design and
- (iii) the statistical Analysis procedures to evaluate the Overall Accuracy (OA), Producer's Accuracy (PA) and User's Accuracy (UA).

The internal validation approach follows scientifically accepted and operationally proven validation designs, aiming at reproducibility, consistency with EEA-endorsed approaches applied in previous productions (especially that of High Resolution Layer 2015SWF (AD04)) as well as for assessing various HRLs of the reference years 2012, 2015 and 2018, performed by the Copernicus Land external Validation team. The result of the sample interpretation is presented in the form of an error matrix, providing Overall Accuracies, Producer's Accuracies and User's Accuracies. These three accuracy measures shall be in the order of 80% with Confidence Interval of 95% (AD01).

1.1 Product to be validated

Small woody landscape features are important vectors of biodiversity and provide information on the fragmentation and connectivity of habitats, especially considering the importance of Green Infrastructure and hazard protection. In this light, the HRL Small Woody Features product contributes to monitor the United Nations' (UN) Sustainable Development Goals (SDGs) and in particular, supports the ecosystem assessment and land accounting in the context of SDG 15 "Life on Land".

From a thematic perspective, Small Woody Features ideally complement the HR Layers portfolio. They illustrate woody linear structures such as hedgerows, scrubs or trees along field boundaries, riparian or roadside vegetation as well as isolated patches of trees and scrubs. Non-woody vegetation elements are not part of the product, i.e., grassy elements, wet elements or artificial elements, as well as tree plantations, vineyards and orchards are excluded (Table 1).

Table 1: Thematic definition of SWF2018

Elements to be included in SWF Mapping 2018	Elements to be excluded in SWF mapping 2018
<ul style="list-style-type: none"> - linear hedgerows and scrubs - tree rows (along field boundaries), - isolated/scattered patches of trees 	<ul style="list-style-type: none"> - stone walls, - drainage ditches, - grass margins, - field boundaries without hedgerows or trees, - any kind of 'grey' infrastructure such as roads - artificial tree rows like olive tree plantations, vineyards, and orchards

For production and for validation, spatial patterns of the Small Woody Features within the landscape play an important role. Therefore, from a technical perspective, the SWF product comprises different geometries, linear woody elements, patchy elements as well as additional woody features, which ensure connectivity between various woody landscape features and improve the look and feel of the product (Table 2):

- Linear SWF: represent landscape features such as hedgerows or tree alignments that are defined by a compactness criterion less or equal to 0.785, up to 30m wide (=Maximum Mapping Width, MMWmax), at least 30m long (50m in 2015). A compactness criterion of 0.785 (0.75 in 2015) to the shorter and wider linear feature possible (e.g. 30m long and 30m wide).
- Patchy SWF: represent areas of isolated and scattered clumps of trees or scrubs defined by a compactness criterion greater than 0.785 with an area greater than 200m² and less than 5,000m² (no upper limit in 2015).
- Additional Woody features: Features that « enhance » connectivity (i.e. connected to a valid SWF) or isolated features with 1500m² < area < 50 000 m² (linear features wider than 30m, and out of specs patches)

Table 2: Geometric specifications of SWF2018

	linear SWF	patchy SWF	connecting AWF	isolated AWF
width	≤ 30m	n/a		
length	≥ 30m	n/a		
area	n/a	200m ² ≤ area ≤ 5000m ²	(no min/max criterion)	1500m ² ≤ area ≤ 50 000m ²
compactness	≤ 0.785	>0.785		

The Additional Woody Features (AWF) have been introduced during production in 2015 and are now fully integrated in the 2018 product in order to ensure connectivity between linear features that were disconnected due to application of the strict geometric rules for linear and patchy SWF. They help the SWF product to better match the variety of woody vegetation features.

Given the potential for confusion between these two elements (SWF and AWF) the 2018 reference year production will be limited to one harmonized SWF vector and raster product, and will NOT continue the separate mapping of AWF.

Therefore, the geometric specifications allow to capture meaningful small woody features into one single SWF class, that is broadly comparable with the combined SWF and AWF products from the 2015 reference year production, but with an upper limit in area for patchy structures (which has not been the case in 2015).

The final HRL2018 SWF 5m product is a binary one, comprising all types of small woody vegetation in one SWF class without further differentiation and a Non-SWF class.

1.2 Validation Criteria

As for thematic interpretation and geometric validity of the product, the interpretation of the samples has to take into account the specific mapping rules of the SWF production:

- Vegetation cover has to be woody and has to be compliant with what has to be included respectively excluded in Table 1 and Table 2**Erreur ! Source du renvoi introuvable..**
- Linear structures connected by features below the Minimum Mapping Width (MMWmin) are not artificially split.
- Patchy structures, parts of which are below the MMWmin, are not artificially cut.
- Trees/bushes are considered as green linear structures where the gaps between the trees/bushes are smaller than 5m.
- Linear and patchy features within open forest (to be understood as less dense, but still forest areas that might not be part of the SWF-forest mask) have to be excluded from the SWF product.

The validation procedure foresees to involve also the environment of the PSU for assessing the type of landscape with its regional characteristics. Interpretation and validation however is only be performed inside the 100 x 100m square of the Primary Sampling Unit.

2 Internal validation procedure

The quantitative assessment of the thematic accuracy is based on stratified random samples compared to external datasets. The OA, PA and UA should be at least 80 % with a 95% confidence interval. An internal validation procedure was done for each delivery batch, for the first delivery of 21% and for remaining Production Units (PU) to complement the 100% delivery.

The validity of the classification is evaluated and illustrated in confusion matrices.

2.1 Stratification and Sample Design

The stratification and sampling design consist primarily of selecting an appropriate sampling frame and sampling unit. It specifies the procedure for ensuring an adequate sample size and representative geographic spread as well as for selecting unbiased sample locations, the 5m product is compared with the ancillary data (mainly VHR2018 and other ancillary data).

The Sampling design for internal validation of the HRL SWF2018 combines systematic and stratified approaches. The design is based on the regular grid-based LUCAS (Land Use/Cover Area frame statistical Survey)¹ sampling approach, using LUCAS points to select Primary Sample Units (PSU) of 100 x 100 m squares. Since the SWF product is based on both linear and patchy elements, a point-based approach is not appropriate for assessing the spatial pattern. Therefore, a segment-based approach has been used). This approach ensures traceability and coherence between the different HRL layers' validation approaches as well as consistency with the validation approaches of previous SWF2015 product (AD04).

Stratification means a series of omission/commission strata based on the 100m SWF2015 raster product:

- Commission stratum: SWF 100m raster product 1-100%
- Omission stratum: SWF 100m raster product 0%

The commission stratum thus is defined by the presence of SWF as provided by the HRL layer, the omission stratum is defined by absence of SWF, both information taken from the SWF 100m raster product. As explained in Stehman & Czaplewski (1998), stratification ensures that each class is properly represented. The number of sample units to be allocated to each stratum has to be calculated as a function of a class' area proportion in the geographic reference area. In case of small strata (i.e. land cover/land use (LC/LU) classes with a very low areal share), where the intersecting area between the stratum and the LUCAS points may be nominally too small to reach the required number of samples, the LUCAS grid would be replicated in order to get more sample points. Although the SWF class is a small one compared to the Non-SWF class, sub stratification was not necessary to get a sufficient number of points.

The number of sample units per stratum should be defined to ensure a sufficient level of precision at reporting level, considering:

- that the number of sample units should allow comparison with HRL external validation results, although the latter will typically reach a higher precision through substantially more samples;

¹ LUCAS corresponds to a grid of approximately 1,100,000 points every 2km throughout the European Union where land cover / land use type is observed and reported.

- cost-effectiveness and operational feasibility within the framework of the applied high-intensity operational industrial production.
- that the number of sample units should allow for statistically rigorous analysis at pan-European level, while also allowing results with reasonable precisions at other levels, such as the level of biogeographic regions, country groups or production units. Although this is not the case for the internal validation of the SWF project 2018, where the level of reporting will be delivery related only, the validation strives for consistency and comparability with the standard for other HRL products in terms of number of samples.

A suitable sample size for each stratum considers an expected acceptable error rate and at the same time shall ensure that even small strata are properly represented in the sample. Since the error rate is expected to be less than 20%, a minimum of 30-50 PSU per PU should provide a standard error of 5-10%. This should be statistically sufficient for a reliable detection of issues at PU level and provide a high degree of precision at higher-aggregation reporting levels, particularly at pan-European level (Wack et al., 2012 and Büttner et al., 2012). In order to accommodate for the differently sized areas of the strata, both the adjustment of the number of samples per stratum and a subsequent area-dependent weighting in the analysis procedure are proven approaches (Selkowitz and Stehlman, 2011). The adjustment per stratum foresees a proportion of around 30% SWF samples to 70% non-SWF samples that reflects the average percentage of SWF in EEA39 vegetation cover. A subsequent area weighting of the samples according to their location in the respective stratum has been applied as standard, however, keeping a critical eye on the plausibility of the results, and considering to potentially increase the number of samples if needed.

Compliant to the previous internal and external HRL validation approaches, a probability sampling design is preferred due to its objectivity. The internal validation approach of HRL 2018 SWF uses a systematic unaligned sampling approach based on the LUCAS sampling grid. LUCAS offers a systematic grid of approximately 1,100,000 points with a 2 x 2 km spacing between the points, covering the area of the European Union + UK (EU-28). For areas outside the EU-28, the LUCAS grid has been spatially extended (i.e. extrapolated), following the original LUCAS grid scheme. The resulting grid serves as basis for the random sampling that fulfils the requirement for representative geographic distribution of the samples in the best way. The Primary Sampling Units of 100 x 100 m squares are snapped to this grid.

This approach has already been used for the SWF2015 internal validation and is also applied for the 2018 internal validation. Since change in woody vegetation coverage is moderate within a 3 years' period, the consortium decided to reuse the SWF2015 sample base to the largest possible extent. The sample base has been reduced by excluding PSU that are covered by more than 50% of forest (<50% coverage of forest mask will be tolerated, in case it is ≥50% coverage of forest mask within a PSU, the whole PSU will be skipped from validation). SWF2018 sees a more extended definition of forest coverage in the 2015 production in order to comply with the FAO-definition of forest. The production of the SWF forest mask (basing on the HRL Tree Cover Mask, TCD2018) comprises now a minimum of 10% tree cover in 2018 instead of minimum 30% tree cover in 2015. A PSU that has 50% or more of its area covered by the forest mask 2018 is excluded from the analysis.

Despite the consequential reduction of samples in 2018, the number of samples participating in the validation is still sufficient to guarantee reliable results for accuracies. The assessment in 2018 bases on 12742 samples for the EEA39 area (including DROM), out of which 8615 PSU are representing the omission stratum, and 4127 PSU

representing the commission stratum. The partial sample base for the EEA39 is 12368 PSU with 8415 PSU for the omission stratum and 3953 PSU for the commission stratum. Which is supplemented for DROM regions by 374 PSU with 200 PSU for the omission stratum and 174 PSU for the commission stratum.

Table 3: Number of valid Primary Sample Units within EEA38+UK areas (including DROM) for commission and omission strata excluding invalid samples and samples with 50% and more forest cover, for the final delivery of 282 Production Units.

	Omission stratum (number of PSU)	Commission stratum (number of PSU)	Total (number of PSU)
EEA39	8415	3953	12368
DROM	200	174	374
EEA39 (incl. DROMs)	8615	4127	12742

The accuracy assessment in the HRL 2018 SWF project foresees two levels: a first assessment at Primary Sampling Unit level (PSU) and a second one at Secondary Sampling Unit level (SSU). The assessment at PSU level aims to be consistent with the HRL 2015 SWF internal validation approach and provides information about the presence or absence of SWF within the 100 x 100m squares. Following the SWF2018 technical proposal recommendations, Secondary Sample Units (SSU) within the PSU complement this information at PSU level documenting the spatial accuracy of the product. SSU are 20 randomly distributed points within each PSU considering a 5 x 5m area (pixel size) for the validation analysis.

This approach of assessing Small Woody Features at PSU and at SSU level has also been used for the selection of PSU to be revisited during the 2018 validation. As already mentioned, it can be assumed that a large proportion of woody vegetation remained unchanged in the 3 years' period from 2015 to 2018, due to the highly stable nature of SWF vegetation cover. Hence the digitized SWF validation information (i.e. SWF that have been assessed, digitized and rasterized) for the year 2015 still remains valid to a large extent, meaning that SWF2015 validation result agrees with the SWF2018 product within a PSU. Where this is the case, no re-interpretation is needed for the reference year 2018. Unstable areas and their respective samples, however, have to be re-interpreted. The assessment of stability or instability of an area and the subsequent selection of the respective samples to be revisited are based on the rasterized 5m versions of both, the validation information of 2015 and the newly classified product for 2018. It comprises the following steps:

1. Using the initial sample base from 2015 and removing all PSU with 50% and more forest cover (using the SWF forest mask of 2018)
2. Comparison at PSU level: Comparing the rasterized validation information from 2015 (SWF15_ref) with the SWF 5m product information from 2018 (SWF18_product) within the 100 x 100m square, using a Minimum Mapping Unit (MMU) of 6 pixels, corresponding to a linear element of 30m length and 5m width:
 - if SWF15_ref is < 6 pixels within a PSU = absence of SWF
 - if SWF15_ref is ≥ 6 pixels within a PSU = presence of SWF
 - if SWF18_product is < 6 pixels within a PSU = absence of SWF

- if SWF18_product is \geq 6 pixels within a PSU = presence of SWF PSU

If this comparison shows an agreement of SWF15_ref vs. SWF18_product concerning SWF presence/absence characteristic results, this PSU is being flagged as not to be revisited in validation. In case of a disagreement between SWF15_ref vs. SWF18_product concerning SWF presence/absence, this PSU is being flagged as to be revisited in validation.

The threshold of 6 pixels at SWF15_ref aims at taking account for the slightly changed specifications (cf. previous chapter, Table 2) as well as geometric inaccuracies/shifts within the EO data (VHR 2015 compared to VHR 2018).

3. The same information of SWF15_ref and SWF18_product will be compared at SSU level referring to all PU remaining after step 1. The analysis is pixel-based, any selected pixel/SSU is assessed and no threshold is used at SSU level:
 - any pixel of discrepancy between SWF15_ref and SWF18_product at SSU level will lead to whole PSU being flagged as to be revisited
 - an agreement between SWF15_ref and SWF18_product at SSU level will lead to whole PSU being flagged as no need to be revisited
4. Using the outcomes of the comparison at PSU and at SSU level:
 - An agreement between the result (=flagging) at PSU and at SSU level concerning “no need to be revisited” leads to whole PSU not to be revisited for validation in 2018 since it is a stable PSU and the validation information from 2015 is still valid in 2018.
 - Any disagreement at PSU or at SSU-level or a flagging at both levels as “to be revisited” will lead to a revision during the 2018 validation.

The assessment of stability/instability at PSU and SSU level for the first delivery results in the following sample numbers:

Table 4: Number of PSU containing still valid information (of 2015) and number of PSU re-assessed in the 2018 internal validation, as well as total number of SSU within valid PSU. Numbers are for the final delivery of 282 PU (EEA39: 277PU; DROM: 5PU).

Total number of PSU	12921
Number of PSU skipped due to insufficient quality of reference data	179
Number of PSU containing still valid information of 2015 internal validation	7993
Number of PSU reassessed in the 2018 internal validation	4749
Number of valid PSU for the 2018 internal validation	12742
Number of SSU within valid PSU	254840

The pre-selection workflow for the final 100% delivery resulted in a total number of 12921 PSU that have been part of the validation for the delivered 282 Production Units. After 179 PSU have been skipped from the validation due to insufficient quality in VHR reference data (shadows, shifts, incorrect geometric accuracy, low image resolution), a total of 12742 valid PSU and the corresponding total number of 254840 SSU samples remained in validation. 7993 PSU could be reused containing still valid information from 2015, and 4749 PSU had to be revisited due to certain degree of instability stemming either from:

- errors in the SWF2018 product whereas SWF2015 validation information is correct
- errors in SWF2015 validation information whereas SWF2018 product is correct
- existence of AWF within the PSU which was not part of the 2015 internal validation but is comprised and validated in the SWF2018 product
- technical change due to slightly changed specifications of the SWF from 2015 to 2018 (linear and patchy SWF, cf. Table 2)

For all of these four cases, reference interpretation for the year 2018 is needed to allow the update of validation information per PSU.

2.2 Response design

The response design is the protocol used for the retrieval of the land cover/land use reference/validation information for all sample units and for the definition of agreement between the mapped land cover/use class(es) with the validation information within the PSU. Given that in-situ data focussing on SWF with spatially, thematically and temporally appropriate ground truth are missing across the EEA39 area, the samples have been assessed by trained operators, through visual photointerpretation of each LUCAS point-based PSU in terms of SWF or Non-SWF presence. VHR EO data of higher spatial resolution and quality than the production data have been complementarily used for this step.

Two types of datasets against which the interpretation will be performed can be distinguished: guiding data and reference data. The guiding data are those used as basis for the production of the Small Woody Features layer. Hence, the available guiding data are:

VHR_IMAGE_2018 and VHR_IMAGE_2018_ENHANCED: very high resolution optical EO image coverage of the reference year 2018 (± 1 year) of the entire EEA39 area comprising Pléiades 1A & 1B, SuperView-1, Kompsat-3/3A and PlanetScope, as well as SPOT-6/7, Deimos-2 and TripleSat data.

The reference data are complementary independent data used to provide more spatial detail and strong landscape context to the assessment:

- Bing maps image / cartography layer
- Open Street Map (OSM) data
- GoogleEarth image / cartography data
- Further in-situ data, e.g. national and regional VHR (airborne) orthophoto, Web Map Services (RGB and/or CIR imagery, with varying spatial resolutions) or ancillary data

The interpretation workflow consists of thematic and geometric plausibility analyses within a PSU. Experienced photo interpreters assessed and manually digitized any valid SWFs included in a PSU, using the above listed guiding and reference imagery datasets. It is also indispensable to fully consider the applicable SWF product specifications (e.g., Minimum Mapping Unit (MMU), Minimum Mapping Width (MMWmin), minimum length, minimum and maximum area, class definitions) as given in Table 1 and **Erreur ! Source du renvoi introuvable..**. In order to facilitate the interpretation of SWF with their complex and difficult geometric rules, the interpreters use also the intermediary SWF2018 vector product including the three SWF classes (linear, patchy and AWF) prior to their merging.

The result from this manual digitization is then converted to a 5 x 5m pixel size raster layer, using the same interpolation/rasterization rules that are used when deriving the SWF 5m raster product from the SWF vector product. This 5m raster layer resulting from the interpretation and digitization workflow then provides the reference information for the comparison with the original 5m product of 2018 and is the basis for evaluation:

The plausibility approach consists of checking the original SWF 5m raster product value to assess whether it can be considered correct or not, within the frame of accepted product specifications.

The validation workflow hence consists of the following steps:

1. Interpretation of SWF or Non-SWF within a PSU
2. Manual digitization of SWF within the PSU
3. Rasterization at 5m resolution as described above
4. Application of a threshold to each sample unit:
 - If number of pixels is <6, then assign this sample as Non-SWF
 - If number of pixels is ≥ 6 , then assign this sample as SWF
5. Assessment of thematic accuracy based on all sample units

The threshold of 6 pixels is applied as minimum mapping area within the area of the 100 x 100m samples, corresponding to a linear element of 30m length and 5m width, resulting in $150\text{m}^2/6$ pixels (same as in the pre-selection workflow, see previous chapter). This takes changed technical specifications in 2018 into account (in 2015, the MMU was 4 pixels corresponding to linear elements of 50m length and 2m width, resulting in $100\text{m}^2/4$ pixels).

3 Statistical Analysis

3.1 Procedure of statistical analysis

The thematic accuracy level of the HRL SWF product is requested to reach 80% in terms of Overall thematic Accuracy, User's accuracy, and Producer's Accuracy. The level of reporting is per delivery, where the 1st interim delivery comprised 58 PU covering about 21% of the EEA39 area (without the French Overseas Departments and Regions, DROM) and the final delivery comprises 282 PU covering the whole EEA39 including DROM regions. As last step of the internal validation, the statistical analysis consists of analysing the assessed samples in terms of agreement between the HRL SWF 5m product and the validation reference in order to provide final conclusions on whether the thematic accuracy of the product has been reached.

The final evaluations have been done

- at PSU-level, to provide information about the presence or absence of SWF
- at SSU-level, to provide information about the spatial accuracy of the product.

The evaluation procedures as well as the results will be part of the following chapters.

3.1.1 PSU-level

After manual digitization of SWF within the sample units and rasterization to 5m spatial resolution (in line with the interpolation/rasterization rules that are used when deriving the SWF 5m raster product from the SWF vector product) a threshold is applied to each 100 x 100 m sample unit, assigning it

- to Non-SWF, if the number of pixels is <6
- to SWF, if the number of pixels is ≥6

in compliance with the MMU of a linear element of 30m length and 5m width, corresponding to 6 pixels of the 5m raster product.

Once both SWF2018 map value (=SWF2018 product information) and SWF2018 reference value (=SWF validation result) are known, it is possible to create a confusion matrix presenting the sample interpretation versus the map class values for all PSU. Unequal sampling intensity resulting from the stratified systematic sampling approach (cf. chapter 2.1) is accounted for by applying a weighting factor (p) to each PSU, 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.

Since samples from the smaller strata show a higher sampling intensity than those from the larger strata, a correction for the sampling intensity has to be applied to the error matrices produced following the procedure described by Selkowitz & Stehman (2011) and applied by Olofsson et al. (2013). It results in a weighting factor inversely proportional to the inclusion probability of samples from a given stratum. Not applying this correction could result in

underestimating or overestimating map accuracies. This procedure has proven to be a state-of-theart approach in previous internal and external validations within Copernicus projects.

For the HRL SWF2018 internal validation, the error matrix at PSU level presents the unweighted and weighted Overall Accuracy, Producer's Accuracy, and User's Accuracy as well as the confidence interval at 95% confidence level.

3.1.2 SSU-level

The Secondary Sample Unit (SSU)-level builds on the PSU level, using the 100 x 100m PSU, in each of which 20 points = SSU per PSU are randomly distributed. The respective SWF2018 product value and the SWF validation reference value is then extracted for each SSU, which allows for creating a confusion matrix presenting the sample interpretation versus the map class values at SSU level. Whereas the PSU matrices illustrate that the presence/absence of SWF within the PSU has been correctly detected, the SSU matrices provide more information on the spatial accuracy of the SWF in order to prove that the woody features are correctly captured.

As for the PSU-level, also for the SSU-level weighting is crucial. The unequal sampling intensity resulting from the stratified systematic sampling approach has an impact on the SSU within the PSU, thus the SSU weighting follow their respective PSU weightings.

3.2 Results of statistical analysis

The confusion matrices and derived accuracies (Overall thematic Accuracy, Producer's and User's Accuracy) as well as the associated confidence intervals of at 95% confidence level are provided in the tables below. Using a stratified sampling approach usually ends up in an uneven distribution of validation points. SWF covers on average 30% of the EEA39-wide vegetation land cover, while Non-SWF is clearly dominating. The resulting imbalance of points and the potential bias in an evaluation has to be compensated for by area weighting to give a realistic representation of the accuracy of the SWF product. The weighting has been derived on the basis of the SWF2015 100m product and takes the 2015 division into Large Regions (LR) into account. The Large Regions were orientated towards biogeographic regions and the subsequent variations in vegetation land cover. The coverage of SWF versus Non-SWF at LR level allowed for deriving weighting figures for every PSU, still being valid for the 2018 product.

3.2.1 Confusion Matrix for accuracies at PSU-level

The confusion matrices below (Table 5, Table 6) illustrate the unweighted and weighted accuracies at PSU level for the EEA39 area including DROM, which corresponds to 282 PU of the final 100% delivery.

Table 5: Unweighted accuracy figures for the SWF2018 5m product at PSU level for EEA39 areas including DROM corresponding to 282 PU of final delivery.

	reference data		Row Total	User's Accuracy	Commission Error	95% Confidence Interval
SWF2018 product	SWF	Non SWF				
SWF	3779	348	4127	91.57%	8.43%	0.46
Non SWF	569	8046	8615	93.40%	6.60%	0.28
Column Total	4348	8394	12742			
Producer's Accuracy	86.91%	95.85%				
Omission Error	13.09%	4.15%				
95% Confidence Interval	0.58	0.22				

Overall Accuracy	92.80%
95% Confidence Interval	0.24
Kappa coefficient	0.84

Table 6: Weighted accuracy figures for the SWF2018 5m product at PSU level for EEA39 areas including DROM corresponding to 282 PU of final delivery.

	reference data		Row Total	User's Accuracy	Commission Error	95% Confidence Interval
SWF2018 product	SWF	Non SWF				
SWF	3223.91	322.25	3546.16	90.91%	9.90%	0.52
Non SWF	432.69	8803.18	9235.87	95.32%	4.68%	0.32
Column Total	3656.60	9125.43	12782.03			
Producer's Accuracy	88.17%	96.47%				
Omission Error	11.83%	3.53%				
95% Confidence Interval	0.59	0.20				

Overall Accuracy	94.09%
95% Confidence Interval	0.22
Kappa coefficient	0.85

Both tables above show the highly improved quality of the SWF product for the reference year 2018 as compared to 2015. Overall thematic Accuracy as well as User's and Producer's Accuracies exceed the required minimum of 80%

at 95% confidence level. The commission errors in the unweighted matrix are below 10% for both classes, SWF shows 8.43% commission error with a UA of 91.57% and Non-SWF shows 6.60% with a UA of 93.40%. PA for the SWF class is slightly lower with its 86.91%, giving an omission error of 13.09%. Compared to those figures, the PA for Non-SWF is very high, 95.85% with an omission error of 4.15% only.

As expected and intended to give a more realistic impression, the weighted figures confirm the very good results on PSU level. Commission errors for SWF is 9.09% at UA of 90.91%, for Non-SWF the commission error is 4.68% with a UA of 95.32%. The omission error for the Non-SWF is 3.53% with a PA of 96.47% and the omission error of the SWF class is slightly higher at 11.83% with a PA of 88.17%. This reveals that presence and absence of SWF and Non-SWF within a PSU could be detected with high precision, however, a certain degree of SWF land cover has not been detected. The OA of the weighted results is 94.09% with a confidence interval of +/-0.22 at 95% confidence level.

3.2.2 Confusion Matrix for accuracies at SSU-level

The confusion matrix below (Table 7, Table 8) illustrate the unweighted and weighted accuracies at SSU level for the final delivery of 282 PU corresponding to 100% of the EEA39 areas including DROM.

Table 7: Unweighted accuracy figures for the SWF2018 5m product at SSU level for EEA39 areas including DROM corresponding to 282 PU of final delivery.

SWF2018 product	reference data		Row Total	User's Accuracy	Commission Error	95% Confidence Interval
	SWF	Non SWF				
SWF	11315	3208	14523	77.91%	22.09%	0.43
Non SWF	5037	235280	240317	97.90%	2.10%	0.03
Column Total	16352	238488	254840			
Producer's Accuracy	69.20%	98.65%				
Omission Error	30.20%	1.35%				
95% Confidence Interval	0.51	0.02				

Overall Accuracy	96.76%
95% Confidence Interval	0.04
Kappa coefficient	0.72

Table 8: Weighted accuracy figures for the SWF2018 5m product at SSU level for EEA39 areas including DROM corresponding to 282 PU of final delivery.

reference data

SWF2018 product	SWF	Non SWF	Row Total	User's Accuracy	Commission Error	95% Confidence Interval
SWF	9079.27	2688.08	11767.35	77.16%	22.84%	0.49
Non SWF	3865.92	64171.31	68037.23	94.32%	5.68%	0.09
Column Total	12945.20	66859.39	79804.59			
Producer's Accuracy	70.14%	95.98%				
Omission Error	29.86%	4.02%				
95% Confidence Interval	0.56	0.08				

Overall Accuracy	91.79%
95% Confidence Interval	0.10
Kappa coefficient	0.69

The unweighted and weighted OA figures at SSU-level are very high, confirming that the spatial accuracy, respectively the correct location of SWF and Non-SWF could well be captured.

Unweighted UA and PA (97.90% and 98.65%) for the Non-SWF class are very high, with very low commission error (2.10%) respectively omission error (1.35%). Contrary to that, UA for the SWF class is only at 77.91%, and PA is even lower at 69.20%. The quite high commission error of 22.09% and omission error at 30.80% indicate that the spatial accuracy of the SWF class is less than expected. OA for the unweighted SSU assessment is still high at 96.76% (confidence interval of +/-0.04).

The weighting balances the impression of spatial accuracies' limitations. Whereas the results for the Non-SWF class are far above 90%, with a UA of 94.32% (commission error of 5.68%) and PA of 95.98% (omission error of 4.02%), the weighted figures for the SWF class do not reach the desired 80%. Weighted UA for SWF is 77.16% with a commission error of 22.84% and weighted PA is 70.14% with an omission error of 29.86%. The results do not generally indicate a low performing spatial accuracy. There might be two reasons for the lower UA and PA values:

- Even if it could be confirmed that woody features have been captured by the semi-automatic production, the exact outline or shape is not matching the reference data in all cases, especially if we consider that the reference dataset is firstly manually digitized (vector format) then converted to raster format to be compared with the SWF2018 5m product. This may amplify a "border discrepancy" between reference and SWF datasets.
- Since the SSU have been extracted randomly, the allocation of SSU on SWF and Non-SWF pixel might be slightly unbalanced, matching more Non-SWF pixels than SWF pixels.

The weighted OA at SSU level is 91.79% with a confidence interval of +/-0.10 at 95% confidence level indicating generally high reliability of the SWF product. A potential adaptation of the validation workflow referring to a higher number of SSU sampling points was tested but showed no significant influence on the resulting accuracies at SSU level.

4 Conclusion

The look and feel of the SWF2018 product has significantly improved compared to the SWF2015. The results of the internal validation on the SWF for this final delivery of EEA39 areas, including the French Overseas Departments and Regions, show high Overall Accuracies at both levels: weighted OA of 94.09% at PSU level with a confidence interval of +/- 0.22 at 95% confidence level; and weighted OA of 91.79% at SSU level with a confidence interval of +/- 0.10 at 95% confidence level. These results fully meet the required product specifications of minimum of 80% thematic accuracy, illustrating that SWF and Non-SWF could well be captured. PA as well as UA are also high at the PSU level and comply with the ITT requirements being above 80%. At SSU level, PA and UA for the SWF class are lower than expected (weighted figures for UA of 77.16% and PA of 70.14%). All accuracy show already at unweighted level high values , indicating that the detection of SWF being an underrepresented class compared to the Non-SWF was successful.

The figures indicate that presence and absence of Small Woody Features have been detected with very high precision at PSU level. The slightly lower figures for PA of the SWF class at PSU level might derive from the complex differentiation between woody vegetation and the different types of forest (types of trees, density, height, coverage), which in some cases is not that obvious and straightforward. Since spectral characteristics are similar, a semi-automated production workflow might have its limitations in correctly capturing the differences.

A similar aspect counts for the SSU level. The OA at SSU level is very good, highlighting a high level of spatial accuracy in the product. However, PA and UA of the SWF class is lower than expected but also a direct consequence of the omission error of the SWF class at PSU level. Both, omission error and commission error at SSU level indicate that detecting woody vegetation according to the thematic and geometric requirements can be done with high correctness, capturing exact shapes and correct locations of the same features. However, the reliability of the EO data is an issue and has impact on both, production as well as on validation. High geolocation-accuracy and a suitable geometric resolution, have direct impact on both, production as well as on validation. Lacking precise co-registration, varying geometric resolution of the input data from different sensors, presence of shadow areas in mountainous regions as well as distorted spectral values at steep slopes, or any further limitations in preprocessing affect production but also aggravate validation efforts.

Since SWF are by nature often small-scale and located in areas with highly heterogeneous land cover, suitable reference data are necessary for correct assessment as well as clear working definitions. The working definition of SWF has turned out to be sometimes vague and less suitable for reflect the broad range of phenotypes and phenology of woody vegetation, and especially the thematic differentiation between the different types of forest (types of trees, density, height, coverage) and SWF for example, can be challenging.

To enhance the quality of the validation and strengthen the reliability of the results, those cases where validation could not be performed due to insufficient reference data, have been excluded from the final evaluation. Given this background, the overall accuracy results at both levels are more than satisfying and confirm that the SWF2018 is a highly meaningful and reliable product.