

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

**Fourth Specific Contract - No 3436/R0-
COPERNICUS/EEA.57889**

HRL SMALL WOODY FEATURES 2015

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 | JR | Janik Deutscher, Petra Miletich | 17/07/2020 |
| Contributions | SIRS | Alexandre Pennec, Loïc Faucqueur | 22/07/2020 |
| Review | SIRS | Alexandre Pennec, Loïc Faucqueur | 30/07/2020 |
| Approval | EEA | Hans Dufourmont | |

Document Issue Record

| Issue | Date | Author(s) | Description of Change |
|-------|------------|----------------------------------|---|
| 1.0 | 17/07/2020 | Janik Deutscher, Petra Miletich | Document creation, Integration of SWF product specifications, user guidelines, validation strategy, examples and analysis, Integration of SWF/AWF/SWF+AWF results |
| 1.1 | 22/07/2020 | Alexandre Pennec, Loïc Faucqueur | Review |
| 1.2 | 28/07/2020 | Petra Miletich | Update chapter 2 and 4. |

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 | Copernicus Land Monitoring Service – High Resolution Layer Small Woody Features. Product Specifications (https://land.copernicus.eu/user-corner/technical-library/hrl_lot5_d5-1_product-specification-document_i3-4_public-1.pdf) |
| AD05 | Second Specific Contract N°3436/R0-COPERNICUS/EEA.56601 |
| AD06 | Third Specific Contract N°3436/R0-COPERNICUS/EEA.57056 |
| AD07 | Fourth Specific Contract - No 3436/R0-COPERNICUS/EEA.57889 |
| AD08 | Copernicus Land Monitoring Service – High Resolution Layer Small Woody Features. Validation Concept (https://land.copernicus.eu/user-corner/technical-library/clms_hrl_swf_validation_concept_sc03_1_1-2.pdf) |
| AD09 | HRL SMALL WOODY FEATURES VALIDATION CONCEPT. https://land.copernicus.eu/user-corner/technical-library/clms_hrl_swf_validation_concept_sc03_1_1-2.pdf |

Executive Summary

This report covers the validation of the 100m High Resolution Layer (HRL) on Small Woody Features (SWF) and Additional Woody Features (AWF). It also validates the combination of the two layers: AWF + SWF (SAWF).

We determine the thematic accuracy of the **HRL SWF 100m product (SWF, AWF and combined SWF+AWF Layer)** using a binary product that is based on a 1% density threshold - i.e. presence or absence of small or additional woody features within a 1ha reference polygon (Primary Sampling Unit). In addition, we provide density scatterplots at pan-European level and for biogeographic regions to visualize the density distribution and to estimate linear regression coefficients. Regression coefficients better show the overall agreement between interpretation and map values, while accuracy results for the chosen binary mask show the agreement on the presence or absence of small woody features for the PSUs.

Regarding the thematic accuracy assessment, the results show that the HRL SWF product (with layers SWF, AWF and SWF+AWF) meets most thematic classification target accuracy requirements (80% blind, 85% plausibility) based on a 1% threshold of SWF, AWF and SWF+AWF density values. Nevertheless, there is variability from country to country or based on biogeographical regions and some country zones and regions do not fulfil the target accuracies.

The main findings and recommendations for the HRL SWF / AWF / SWF+AWF products can be summarised as follows:

1. The SWF layer exceeds the overall target accuracy at pan-European level at the 1% density threshold for both blind and plausibility interpretation. The target Producer and User Accuracies are met after plausibility checks.
2. The AWF layer exceeds the overall target accuracy at pan-European level at the 1% density threshold for both blind and plausibility interpretation. The target Producer Accuracy is neither met for blind nor for the plausibility interpretation (59.73% and 72.43%) suggesting a high commission error. This might be related to the intermediate forest cover mask, which is based on the HRL TCD 2015 product. It is noticeable that accuracy issues for AWF (and partly SWF) occur in the same biogeographic regions and country zones than for the TCD 2015 product.
3. The combined SWF+AWF layer exceeds the target accuracy at pan-European level for overall accuracy and both omission and commission errors at the 1% density threshold for blind & plausibility validation.
4. Results are provided at lot, bio-geographical regions and country/group of country level and should provide a sound basis for further improving the product for future updates in areas that's show accuracy deficiencies.
5. Based on the lessons learnt from the previous Specific Contract, the chosen stratification procedure was effective for estimating both commission and omission errors with mostly low confidence intervals. However, for the analysis of the scatterplots the sampling does not seem to be ideal as the vast majority of plots has 0% density for both SWF and AWF. A more stratified sampling that includes a larger number of higher density plots is recommended for future validations.
6. There is some confusion between SWF and AWF and also at the 1% threshold (AWF/None, SWF/None).
7. The large number of geometric thresholds used during the production leads to a large number of ambiguous plots during interpretation, especially since the area specifications for patchy SWF and AWF partly overlap, meaning some woody features can indeed be both SWF or AWF.
8. In general, the applied geometric rules are very difficult to determine by visual interpretation (e.g. compactness of a feature) which results in a large number of ambiguous plots. This is reflected by a strong increase of the thematic accuracy during plausibility assessment.
9. The AWF product is closely linked to the SWF product (may not overlap with SWF, enhance connectivity of SWFs, only in areas that are neither SWF nor included in the forest cover mask). It is thus questionable if a blind product validation without knowledge of the presence and location of SWFs is very useful. Checking if the connectivity between SWFs is really enhanced by AWFs requires knowledge on the presence and location of both the SWFs and AWFs in the plot. This cannot be done properly in a blind interpretation.

10. The scatterplots for SWF and AWF mostly show low to medium quality regression coefficients (0.03-0.77). This can be explained by the very large number of plots that disagree at the 0/1% threshold compared to those that are > 0% in both map and reference. Density values are also not scattered evenly from 0-100% but tend to centre around 0% with 70-80 of plots showing density value of 0% in SWF and/or AWF. Thus R^2 is likely to be lower than for e.g. TCD where also many plots with 80-100% density occur. Due to the large number of plots at or around 0%, the scatterplots are also very difficult to read and to interpret. We recommend to increase the number of sample plots at higher density values for future analyses, to obtain more meaningful scatterplots.
11. A better compatibility of the SWF product with other HRL layers would be beneficial to the users. The product is closely linked to the HRL Forest. Currently the products partly overlap with TCD and FTY, which is not intuitive to the users. However, we do understand the problems that arise due to different definitions used and a full compatibility is currently not possible. Whether HRL layer definitions can be adapted to make products more compatible, should be part of a separate analysis.
12. The intermediate forest cover mask that was used during production seems to be an integral part of the product (exclusion areas) and we recommend to make this SWF forest cover mask available as a support layer to SWF. In some areas it may not be clear to the users why certain woody features are excluded from the different SWF Layers.

Table of Contents

| | |
|--|-----------|
| | 0 |
| 1. Validation Framework | 14 |
| 1.1. Products to be validated..... | 14 |
| 1.2. SWF product definition..... | 16 |
| 1.3. AWF product definition | 17 |
| 1.4. Validation Criteria..... | 17 |
| 1.4.1. SWF..... | 18 |
| 1.4.2. Complementarity and possible overlaps of the SWF product with the HRL Tree Cover Density (TCD)..... | 20 |
| 3. Validation approach | 22 |
| 3.1. Completeness | 22 |
| 3.2. Logical consistency | 22 |
| 3.2.1. Conceptual consistency..... | 22 |
| 3.2.2. Domain consistency | 23 |
| 3.2.3. Format consistency | 23 |
| 3.2.4. Topological consistency..... | 23 |
| 3.2.5. Additional logical consistency checks | 23 |
| 3.3. Positional Accuracy..... | 24 |
| 3.4. Thematic Accuracy..... | 24 |
| 3.4.1. Level of reporting | 24 |
| 3.4.2. Stratification and sample design..... | 26 |
| 3.4.3. Response Design..... | 34 |
| 3.4.5. Estimation and analyses procedures..... | 45 |
| 3.5. Temporal Quality | 46 |
| 3.6. Usability | 46 |
| 3.7. INSPIRE compliant metadata | 46 |
| 4. Validation check list..... | 47 |
| 4.1. Small Woody Features (SWF) | 47 |
| 5. Thematic accuracy..... | 52 |
| 5.1. Small Woody Features (SWF) | 52 |
| 6. Conclusions and recommendations | 61 |
| 7. References..... | 63 |
| Annex 1. Small Woody Features (SWF) Scatterplot for bio-geographical regions | 64 |

| | |
|--|----|
| Annex 2. Additional Woody Features (AWF) Scatterplot for bio-geographical regions..... | 68 |
| Annex 3. Small Woody Features (SWF) Scatterplot for countries and group of countries greater than 90,000km ² | 72 |
| Annex 4. Additional Woody Features (AWF) Scatterplot for countries and group of countries greater than 90,000km ² | 79 |
| Annex 5. SWF2015 - confusion matrices | 86 |

List of Figures

| | |
|---|----|
| Figure 1. SWF 100m density raster product over EEA39 | 15 |
| Figure 2. SWF 5m raster layer (left) and SWF 100m raster product (the darker the green, the higher the density of | 15 |
| Figure 3: AWF (dashed orange features) enhancing connectivity between SWF (light green), while partially overlapping TCD2015 (left), or being complementary with TCD 2015 (right). For more information on the mapping rules concerning the TCD (Tree Cover Density) product, please see below..... | 17 |
| Figure 4: Application of forest mask based on HRL TCD 2015. (a): VHR2015; (b): HRL TCD 2015; (c): small and additional woody feature (yellow outline) and newly generated forest cover mask (light green); Small and additional woody features (yellow outline) and HRL TCD2015: the 2 products are partially overlapping, but also complementary for features too small to be included in HRL TCD 2015. | 21 |
| Figure 5: Level of reporting at pan-European level..... | 25 |
| Figure 6: Level of reporting accordingly to the biogeographical regions (EEA, 2016)..... | 25 |
| Figure 7: Level of reporting according to the country or aggregated countries..... | 26 |
| Figure 8: Simple random (left) and random systematic (right) sampling designs..... | 27 |
| Figure 9: LUCAS points located on a regular grid..... | 27 |
| Figure 10. Number of sample points as a function of the expected error rate for two accepted standard error values (after Wack et al. 2012)..... | 28 |
| Figure 11: Replicates and sub-replicates used on LUCAS grid..... | 29 |
| Figure 12: SWF 2015 availability and production units. | 30 |
| Figure 13: Example of SSUs organised in a 5x5point 20m grid..... | 31 |
| Figure 14: Examples of different compactness factors based on a bounding rectangle..... | 35 |
| Figure 15: SWF Interpretation tool for reference data labelling (implemented by Joanneum Research). The tool provides relevant image chips, single grid cell value selection, density value summaries, and can handle a plausibility analysis. The red encircled area shows that there a discrepancy between map SWF value and interpreted SWF value. | 36 |
| Figure 16: SWF interpretation example: no SWF and AWF are present, all grid cells (SSUs) are None (0)..... | 38 |
| Figure 17: SWF interpretation example: olive trees or orchards are not SWF/AWF. | 39 |
| Figure 18: SWF interpretation example: areas within the forest cover mask are not considered. | 39 |
| Figure 19: SWF interpretation example: For some plots, the forest cover mask was not considered due to obvious thematic errors..... | 40 |
| Figure 20: SWF interpretation example: open forests are not SWF or AWF even though they are not excluded by a forest cover mask at this sample plot. | 40 |
| Figure 21: SWF interpretation example: Maximum mapping size for SWF exceeded. | 41 |
| Figure 22: SWF interpretation example: example of the 5m maximum distance between single trees to be considered as linear SWF..... | 41 |
| Figure 23: SWF interpretation example: ambiguous interpretations between SWF and AWF are possible due to the 30m width threshold applied to linear SWF. | 42 |
| Figure 24: SWF interpretation example: ambiguous interpretations are possible also for the 3m minimum mapping width of linear SWF. | 42 |
| Figure 25: SWF interpretation example: mountain dwarf pines are currently not considered in the product specifications..... | 43 |

| | |
|---|----|
| Figure 26: SWF interpretation example: Plot within a forest area > 5ha, but not removed by the forest cover mask..... | 43 |
| Figure 27: SWF interpretation example: A linear vegetation element of reed. It is not woody and thus not considered as SWF / AWF..... | 44 |
| Figure 28: SWF interpretation example: a linear SWF barely touches the PSU grid. | 44 |
| Figure 29: Representation of the behaviour of the 95% confidence interval for a 5x5point SSU grid over the whole range of density values (example from imperviousness analysis). | 45 |
| Figure 30: Distribution of sampling units by reference SWF density values. | 53 |
| Figure 31: Scatterplot of all SWF sample units darker green indicates a greater number of observations..... | 53 |
| Figure 32: Distribution of sampling units by reference AWF density values..... | 54 |
| Figure 33: Scatterplot of all AWF sample units darker green indicates a greater number of observations..... | 54 |
| Figure 34: SWF Layer - Alpine bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values..... | 64 |
| Figure 35: SWF Layer - Anatolian bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values..... | 64 |
| Figure 36: SWF Layer - Arctic bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values..... | 64 |
| Figure 37: SWF Layer - the Atlantic bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values..... | 65 |
| Figure 38: SWF Layer - Black Sea bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values..... | 65 |
| Figure 39: SWF Layer - Boreal bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values..... | 65 |
| Figure 40: SWF Layer - Continental bio-geographical: (left) Scatterplot of all sample units. (Right) Distribution of density values..... | 66 |
| Figure 41: SWF Layer - Macaronesian bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values..... | 66 |
| Figure 42: SWF Layer - Mediterranean bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values..... | 66 |
| Figure 43: SWF Layer - Pannonic bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values..... | 67 |
| Figure 44: SWF Layer - Steppic bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values..... | 67 |
| Figure 45: AWF Layer - Alpine bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values..... | 68 |
| Figure 46: AWF Layer - Anatolian bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values..... | 68 |
| Figure 47: AWF Layer - Arctic bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values..... | 68 |
| Figure 48: AWF Layer - the Atlantic bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values..... | 69 |
| Figure 49: AWF Layer - Black Sea bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values..... | 69 |
| Figure 50: AWF Layer - Boreal bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values..... | 69 |

| | |
|---|----|
| Figure 51: AWF Layer - Continental bio-geographical: (left) Scatterplot of all sample units. | 70 |
| (Right) Distribution of density values..... | |
| Figure 52: AWF Layer - Macaronesian bio-geographical region: (left) Scatterplot of all sample units. | 70 |
| (Right) Distribution of density values..... | |
| Figure 53: AWF Layer - Mediterranean bio-geographical region: (left) Scatterplot of all sample units. | 70 |
| (Right) Distribution of density values..... | |
| Figure 54: AWF Layer - Pannonian bio-geographical region: (left) Scatterplot of all sample units. | 71 |
| (Right) Distribution of density values..... | |
| Figure 55: AWF Layer - Steppic bio-geographical region: (left) Scatterplot of all sample units. | 71 |
| (Right) Distribution of density values..... | |
| Figure 56: SWF Layer – AL_MK_ME_RS_KS country zone: (left) Scatterplot of all sample units. | 72 |
| (Right) Distribution of density values..... | |
| Figure 57: SWF Layer – AT_CH_LI country zone: (left) Scatterplot of all sample units. (Right) | 72 |
| Distribution of density values. | |
| Figure 58: SWF Layer – BE_NL_LU_DK country zone: (left) Scatterplot of all sample units. | 72 |
| (Right) Distribution of density values..... | |
| Figure 59: SWF Layer – BG country zone: (left) Scatterplot of all sample units. (Right) | 73 |
| Distribution of density values. | |
| Figure 60: SWF Layer – CZ_SK country zone: (left) Scatterplot of all sample units. (Right) | 73 |
| Distribution of density values. | |
| Figure 61: SWF Layer – DE country zone: (left) Scatterplot of all sample units. (Right) | 73 |
| Distribution of density values. | |
| Figure 62: SWF Layer – EE_LT_LV country zone: (left) Scatterplot of all sample units. (Right) | 74 |
| Distribution of density values. | |
| Figure 63: SWF Layer – ES country zone: (left) Scatterplot of all sample units. (Right) | 74 |
| Distribution of density values | |
| Figure 64: SWF Layer – FI country zone: (left) Scatterplot of all sample units. (Right) | 74 |
| Distribution of density values. | |
| Figure 65: SWF Layer – FR_AD country zone: (left) Scatterplot of all sample units. (Right) | 74 |
| Distribution of density values. | |
| Figure 66: SWF Layer – GB_IE country zone: (left) Scatterplot of all sample units. (Right) | 75 |
| Distribution of density values | |
| Figure 67: SWF Layer – GR_CY country zone: (left) Scatterplot of all sample units. (Right) | 75 |
| Distribution of density values. | |
| Figure 68: SWF Layer – HR_BA_SI country zone: (left) Scatterplot of all sample units. (Right) | 75 |
| Distribution of density values. | |
| Figure 69: SWF Layer – HU country zone: (left) Scatterplot of all sample units. (Right) | 76 |
| Distribution of density values. | |
| Figure 70: SWF Layer – IS country zone: (left) Scatterplot of all sample units. (Right) | 76 |
| Distribution of density values. | |
| Figure 71: SWF Layer – IT_MT country zone: (left) Scatterplot of all sample units. (Right) | 76 |
| Distribution of density values. | |
| Figure 72: SWF Layer – NO country zone: (left) Scatterplot of all sample units. (Right) | 77 |
| Distribution of density values. | |
| Figure 73: SWF Layer – PL country zone: (left) Scatterplot of all sample units. (Right) | 77 |
| Distribution of density values. | |

| | |
|---|----|
| Figure 74: SWF Layer – PT country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values | 77 |
| Figure 75: SWF Layer – RO country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values | 77 |
| Figure 76: SWF Layer – SE country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values | 78 |
| Figure 77: SWF Layer – TR country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values | 78 |
| Figure 78: AWF Layer – AL_MK_ME_RS_KS country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values..... | 79 |
| Figure 79: AWF Layer – AT_CH_LI country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values | 79 |
| Figure 80: AWF Layer – BE_NL_LU_DK country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values..... | 79 |
| Figure 81: AWF Layer – BG country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values | 80 |
| Figure 82: AWF Layer – CZ_SK country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values | 80 |
| Figure 83: AWF Layer – DE country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values | 80 |
| Figure 84: AWF Layer – EE_LT_LV country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values | 81 |
| Figure 85: AWF Layer – ES country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values | 81 |
| Figure 86: AWF Layer – FI country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values | 81 |
| Figure 87: AWF Layer – FR_AD country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values | 81 |
| Figure 88: AWF Layer – GB_IE country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values | 82 |
| Figure 89: AWF Layer – GR_CY country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values | 82 |
| Figure 90: AWF Layer – HR_BA_SI country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values | 82 |
| Figure 91: AWF Layer – HU country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values | 83 |
| Figure 92: AWF Layer – IS country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values | 83 |
| Figure 93: AWF Layer – IT_MT country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values | 83 |
| Figure 94: AWF Layer – NO country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values | 84 |
| Figure 95: AWF Layer – PL country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values | 84 |
| Figure 96: AWF Layer – PT country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values | 84 |

| | |
|--|----|
| Figure 97: AWF Layer – RO country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values. | 84 |
| Figure 98: AWF Layer – SE country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values. | 85 |
| Figure 99: AWF Layer – TR country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values. | 85 |

List of Tables

| | |
|---|----|
| Table 1.1: Thematic definitions of SWF (included/excluded) | 16 |
| Table 1.2: Geometric specifications of SWF..... | 16 |
| Table 1.3: Detailed specification of the 100m Small Woody Feature products. | 19 |
| Table 2.1: Map projection details for LAEA-ERTS89. | 24 |
| Table 2.2: SWF sample plot distribution by level of stratification at pan-European level. | 31 |
| Table 2.3: SWF sample plot distribution per strata and area in km ² at pan-European level (excluding forest cover mask). | 32 |
| Table 2.4: SWF distribution of sample units per countries or groups of countries..... | 32 |
| Table 2.5: SWF weight factors to be applied to each stratum and level of stratification for constructing confusion matrices..... | 34 |
| Table 4.1: Details of SWF/AWF regression line parameter and coefficients and for the whole of Europe cover, per lot, bio-geographical regions and country and group of countries greater than 90,000km ² , n corresponds to the number of observations. Values in green are closest to optimal values (excellent when ≥ 0.8 and okay when ≥ 0.5 for R ² , close to 1 slope and close to 0 for intercept). | 55 |
| Table 4.2: SWF / AWF / SWF+AWF layers thematic accuracy results at the 1% threshold for the pan-European area + DOMs. Values in green indicate accuracies greater than the 80% threshold (blind/plaus), yellow greater than 80% threshold when the 95% CI is considered, red when less than 80% even when considering the 95% CI (Prod. = Producer). | 57 |
| Table 4.3: SWF / AWF / SWF+AWF product thematic accuracy results for the bio-geographical regions. Values in green indicate accuracies ≥ 80%. Values in yellow indicate accuracies ≥ 80% when adding the confidence values..... | 57 |
| Table 4.4: SWF / AWF / SWF+AWF product thematic accuracy results for countries and group of countries greater than 90,000km ² . Values in green indicate accuracies ≥ 80%. Values in yellow indicate accuracies ≥ 80% when adding the confidence values..... | 58 |
| Table 6.1: Estimated area proportion matrix for AWF mapping on pan-european level. Columns correspond to reference and rows to mapping..... | 86 |
| Table 6.2: Raw error matrix (counts) for AWF mapping on pan-european level. Columns correspond to reference and rows to mapping..... | 86 |
| Table 6.3: Estimated area proportion matrix for SWF mapping on pan-european level. Columns correspond to reference and rows to mapping..... | 86 |
| Table 6.4: Raw error matrix (counts) for SWF mapping on pan-european level. Columns correspond to reference and rows to mapping..... | 86 |

List of Abbreviations

| | |
|---------|---|
| BRME | Biogeographical Regions Map of Europe |
| AWF | Additional Woody Features |
| CLC | CORINE Land Cover |
| DWH | Data Warehouse |
| 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 |
| SWF+AWF | Small + Additional Woody Features (combined product) |
| SSU | Secondary Sampling Unit |
| SWF | Small Woody Features |
| 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 the CORINE land cover (CLC) datasets. The HRLs 2015 are produced from 20m resolution satellite imagery through a combination of automatic processing and interactive rule based classification.

Five themes have been identified for the 2015 products, corresponding mainly with the main themes from CLC, i.e. imperviousness degree (IMD), tree cover density (TCD) and forest type (FTY), grasslands (GRA), water and wetness (WaW) and small woody features (SWF). For those the pixels of 20 by 20m are aggregated into 100 by 100m grid cells for final products. The SWF product is an exception as it is produced from VHR imagery as vector file and a 5 by 5m raster. The 100 by 100m product is derived from the 5 by 5m raster.

Pan-European wall to wall products cover all EEA39 countries. They were produced in a combined centralized and decentralized approach, involving service industry through market mechanisms and participating countries through grant agreements.

The **HRL Small Woody Features 2015** of the Copernicus Land Monitoring Services (CLMS) - High Resolution land cover characteristics for the 2015 reference year aims to deliver homogeneous information on small woody features across EEA39 countries. This includes linear structures such as hedgerows, but also patches of woody vegetation. The methodology for detecting these structures of small woody vegetation is based on the analysis of very-high resolution earth observation data for the reference year 2015 (+/- 1 year).

Small woody landscape features are important for a number of ecosystem services, among others related to biodiversity and habitat connectivity. They also improve air quality, water quality, water quantity, reduction of greenhouse gas emissions, carbon sequestration, climate change adaptation, regulation of soil erosion and soil quality, support biodiversity and pollination, and are important for cultural and recreational reasons. They are also highly relevant in terms of the green infrastructure strategy of the EU, and for monitoring of the effectiveness of the Common Agricultural Policy (CAP). The VHR_IMAGE_2015 dataset made available in the ESA Copernicus DWH is the main data source for the detection of Small Woody Features identifiable within the given image resolution (=<1m panchromatic, 2-4m multi-spectral).

Based on very-high resolution (VHR) satellite imagery from Copernicus Contributing Missions (CCMs) made available through the ESA Data Warehouse (DWH), the 3 SWF layers, a 100m AWF density layer and a 100m SWF+AWF density layer.

- SWF vector layer: separates the SWF class into 1: Linear & 2: Patchy; 3: Additional Woody Features
- SWF 005m spatial resolution raster layer: only separates between 1: SWF & 3: AWF
- SWF 100m spatial resolution raster aggregate layer: SWF density 0-100%
- AWF 100m spatial resolution raster aggregate layer: AWF density 0-100%
- SWF+AWF 100m spatial resolution raster aggregate layer: SWF+AWF density 0-100%

This validation report includes the 100m spatial raster products for SWF and AWF and a combination of the two products SWF+AWF. An example of the pan-European SWF2015 100m products is shown in Figure 1.

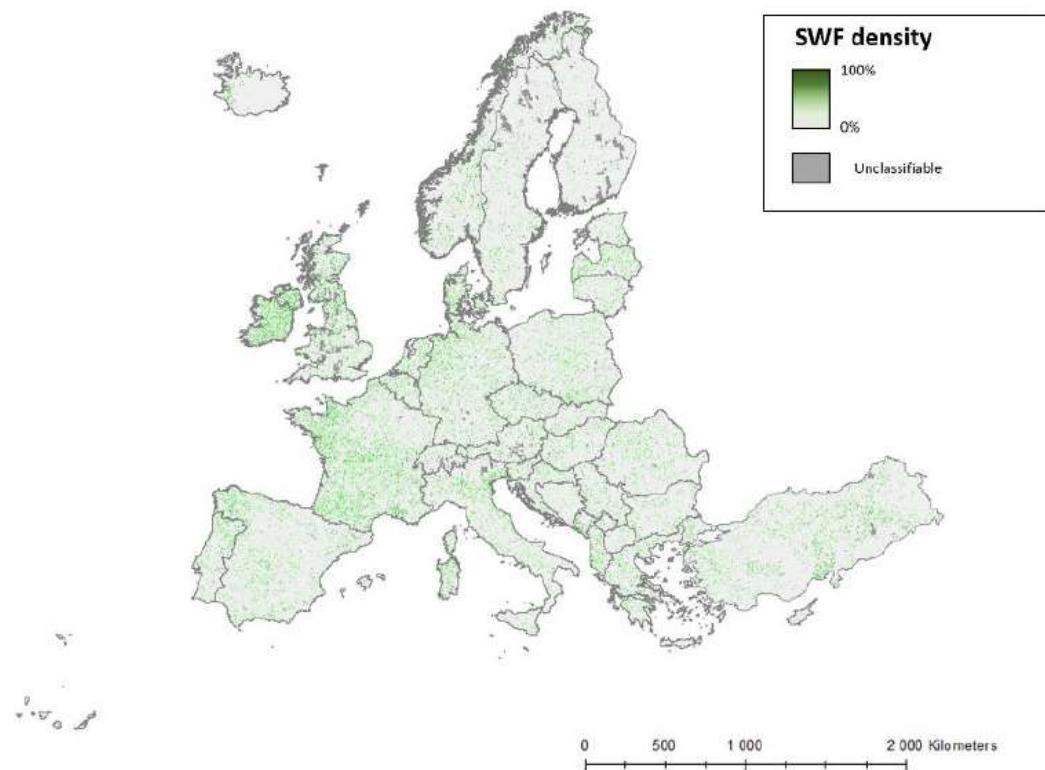


Figure 1. SWF 100m density raster product over EEA39.

The **HRL 2015 SWF 100m raster product** maps the level of small and/or additional woody feature density in a range from 0-100%. It is an aggregated version of the SWF 5m product raster layer. It can be used as a landscape descriptor of SWF density for large areas. It is fully aligned to the EEA 100m reference grid.

Following the integration of AWF in the HRL2015 layer, the SWF 100m raster now exists in three different versions, which are all validated in this validation exercise:

- SWF density, 100m x 100m, European projection, LAEA
- AWF density, 100m x 100m, European projection, LAEA
- SWF+AWF density, 100m x 100m, European projection, LAEA

Figure 2 shows a comparison of the SWF 5m raster layer and the SWF 100m raster layer. Darker green colours represent a higher SWF density values.

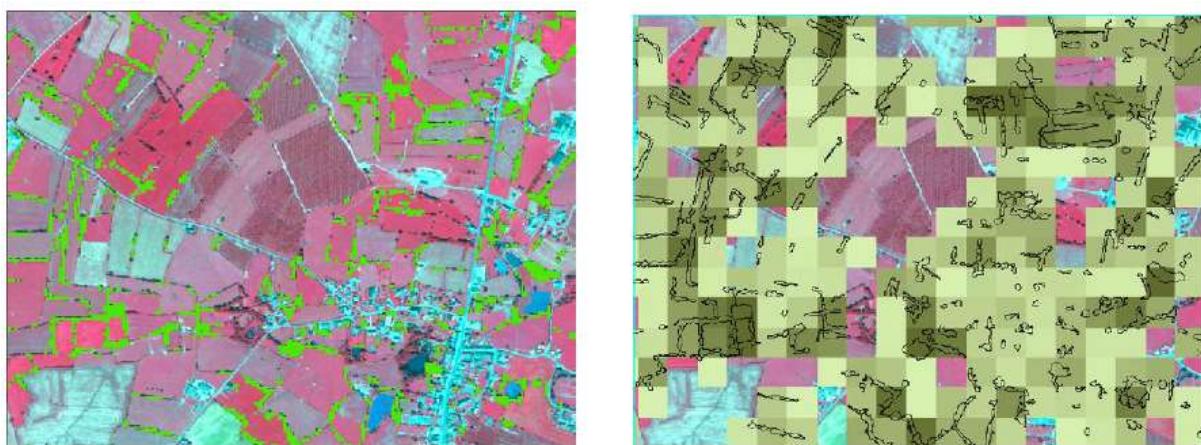


Figure 2. SWF 5m raster layer (left) and SWF 100m raster product (the darker the green, the higher the density of SWF) with the SWF vector product overlaid with black outline (right).

1.2. SWF product definition

The SWF product includes woody linear structures such as hedgerows, shrubs or tree rows along field boundaries, riparian and roadside vegetation. Isolated patches of trees and scrubs are included.

The product excludes grassy elements (e.g. margins along field boundaries), wet elements (drainage ditches, watercourses) or artificial elements (any kind of 'grey' infrastructure such as roads or stonewalls). Tree plantations, vineyards and orchards are also not included in the SWF product.

An overview of the thematic definitions of the SWF product is provided in Table 1.1, which is derived from the SWF 2015 Product Specification Document (https://land.copernicus.eu/user-corner/technical-library/hrl_lot5_d5-1_product-specification-document_i3-4_public-1.pdf).

Table 1.1: Thematic definitions of SWF (included/excluded).

| Elements to be included in SWF mapping 2015 | Elements to be excluded in SWF mapping 2015 ¹ |
|---|---|
| linear hedgerows and scrubs | stone walls, |
| tree rows (e.g. along field boundaries), | drainage ditches, |
| isolated/scattered patches of trees | 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 ² |

In addition to the thematic definitions, the SWF product is also constrained by geometric specifications. The geometric specifications distinguish between linear and patchy structures and consider width, length, compactness and area of woody features as indicated in Table 1.2, which is derived from the SWF 2015 Product Specification Document (https://land.copernicus.eu/user-corner/technical-library/hrl_lot5_d5-1_product-specification-document_i3-4_public-1.pdf).

Examples of geometric mapping rules are:

- Linear structures may contain feature parts wider than 30m if connecting 2 features of less than 30m width, each longer than 50m, over a distance of less than 50m – otherwise this would result in rather artificial cuts in linear structures;
- Patchy structures may have a width of less than 30m over a distance below 50m. A cut is not applied in order to keep the overall, natural characteristic of the identified patch.
- Trees are considered as green linear structure when the gaps between the trees are smaller than 5m.
- Linear and patchy features within open forest are excluded from the SWF product

Table 1.2: Geometric specifications of SWF.

| | Linear Structures | Patchy Structures |
|-------------|-------------------|--|
| Width | $\leq 30m$ | n/a |
| Length | $\geq 50m$ | n/a |
| Compactness | ≤ 0.75 | > 0.75 |
| Area | n/a | $200m^2 \leq \text{area} \leq 5000m^2$ |

1.3. AWF product definition

Given the limitations with any geometric rules and the need for good “green” connectivity, an additional class was added: the “Additional Woody Features” (AWF). This AWF class (thematic code = 3 in the vector and 5m raster file) includes the woody elements identified by the pre-classification but rejected by the geometric rules. This will provide to the user more relevant woody features and better ensure the link between Tree Cover Density (TCD) and SWF. The purpose of this AWF class is to re-inject meaningful features detected by pre-classification and removed by post-processing due to the applications of SWF geometric rules (see Figure 3).

To avoid re-introducing every feature in the product, potential AWF are selected with the following rules:

- Features that « enhance » connectivity (i.e. features connected to a valid SWF)
- Isolated features with area > 1500 m² (linear wider than 30m, and out of specs for patchy SWF structures)

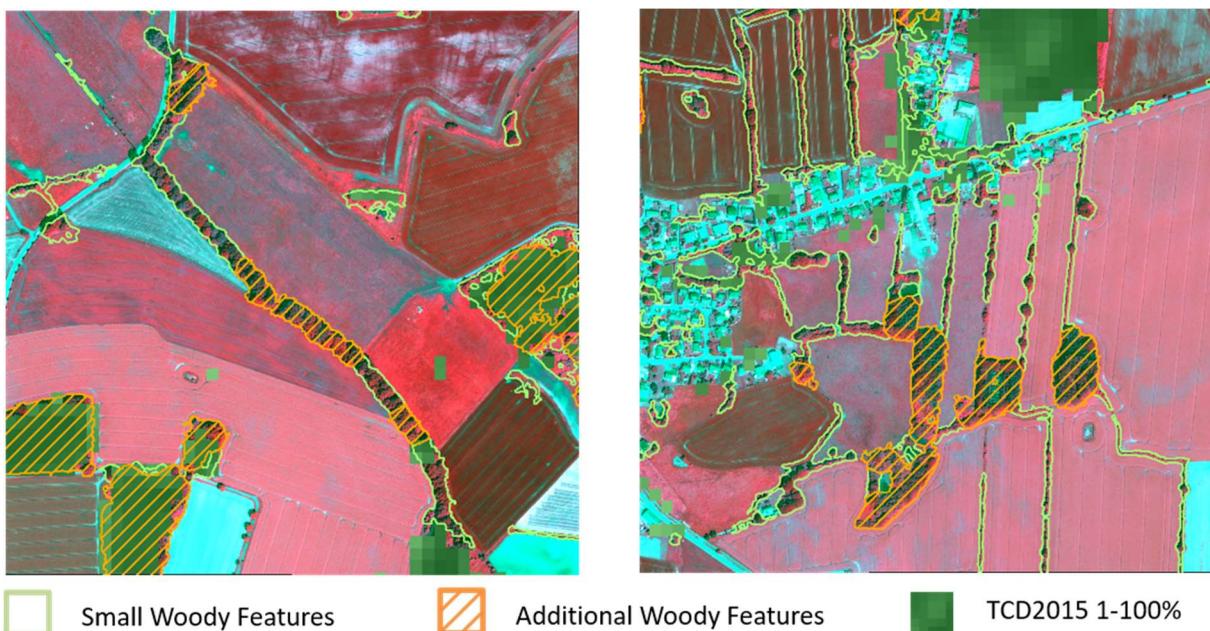


Figure 3: AWF (dashed orange features) enhancing connectivity between SWF (light green), while partially overlapping TCD2015 (left), or being complementary with TCD 2015 (right). For more information on the mapping rules concerning the TCD (Tree Cover Density) product, please see below.

AWF and SWF classes are mutually exclusive (non-overlapping). Every detected feature is first compared with the SWF geometric rules. If this feature meets SWF linear or SWF patch specifications, it is labelled as SWF. If not, it is eventually labelled as AWF if it is meeting the AWF conditions listed above.

1.4. Validation Criteria

Detailed completeness and logical consistency checks are already performed as part of the semantic checks undertaken by the QC Tool during upload of the products. Therefore, the aim of this validation exercise is not to repeat these, but to review the existing documents and perform additional checks if deemed necessary. We recommend however, to make the QC Tool reports available openly as additional quality information to the users.

For the SWF layers, the main criteria selected for the validation are:

- Thematic accuracy: the correspondence with reference data.
- Completeness: the amount of omission and commission.
- Logical consistency, the adherence to formats, conventions and conceptual aspects.
- Temporal quality, the alignment of the results with the reference year.

- Usability.
- Metadata, the presence of sufficient metadata to describe the product.

Other validations are either not applicable (topological consistency is not relevant for a raster dataset) or being dealt with by other aspects of the project (positional accuracy is an assessment of the image data).

1.4.1. SWF

In the validation framework, the assessment of the thematic accuracy of the HRL 2015 SWF, AWF and SWF+AWF raster layers at 100m spatial resolution is performed at two levels:

- A scatterplot of the density values extracted from the sample units for both the reference and layer data is made with a view to assess the correlation between reference and map values and identify any systematic bias (slope and intercept of the regression line significantly different for 1 and 0 respectively)
- A threshold applied to the density values for reference and map data to produce binary attributes for both the reference and map data layers. For SWF, AWF and SWF+AWF, a threshold of 1% density was defined for the validation process (cf. HRL Small Woody Features Validation Concept, AD09). That means the analysis is based on the presence or absence of SWF and AWF in the respective PSU (Primary Sampling Unit) of 1ha.
 - Density values 1 to 100 (class 1) against density value 0 (class 0)

In the validation framework, the general accuracy level of the HRL SWF products shall be in the order of 80% (Overall thematic Accuracy, User's Accuracy and Producer's Accuracy). The product specification are only defined for the SWF layer since the AWF class was added during production as a “complementary class”. For comparability, also the AWF and the SWF+AWF layer are also analysed with the 80% threshold. The 80% accuracy value shall be understood as follows: 20% for commission errors and 20% for omission errors for the SWF/AWF/SWF+AWF classes.

The detailed specifications of the Small Woody Feature 100m product are shown below:

Table 1.3: Detailed specification of the 100m Small Woody Feature products.

| Small Woody Features 100m | Acronym | Product category |
|--|----------------|-------------------------|
| | SWF | Aggregated status layer |
| Reference year | | |
| 2015 (+/- 1 year) | | |
| Geometric resolution | | |
| <i>Pixel resolution 100m x 100m, fully conform with the EEA reference grid</i> | | |
| Coordinate Reference System | | |
| <i>European ETRS89 LAEA projection</i> | | |
| Geometric accuracy (positioning scale) | | |
| <i>According to ortho-rectified satellite image base delivered by ESA.</i> | | |
| Thematic accuracy | | |
| <i>Determined by the accuracy of the source SWF 5m raster</i> | | |
| Data type | | |
| <i>8bit unsigned raster with LZW compression</i> | | |
| Minimum Mapping Unit (MMU) | | |
| <i>Pixel-based (no MMU)</i> | | |
| Raster coding (thematic pixel values) | | |
| <i>0: no small or additional woody features present in grid cell</i> | | |
| <i>1-100: small or additional woody feature density values</i> | | |
| <i>254: unclassifiable (no satellite image available, or clouds, shadows, or snow)</i> | | |
| <i>255: outside area</i> | | |
| Metadata | | |
| <i>XML metadata files according to INSPIRE metadata standards</i> | | |
| Delivery format | | |
| <i>GeoTIFF or TIFF</i> | | |

1.4.2. Complementarity and possible overlaps of the SWF product with the HRL Tree Cover Density (TCD)

In order to exclude SWF from larger areas with dense tree cover, which are well covered by the TCD HRL, the production workflow considers a mask based on HRL TCD 2015. To obtain the forest cover mask used in production,

- a 30% threshold is applied to HRL TCD 2015.
- In a second step, a morphological filter (erosion + dilatation) is applied to the intermediate tree cover mask. This allows to exclude from the tree cover mask linear elements connected to larger forest area (see Figure 4).
- In a third step, a minimal mapping unit (MMU) of 5 ha is applied. Due to the coarser spatial resolution of HRL TCD2015 compared to SWF product, this allows to identify relevant small woody features over areas which are covered by HRL TCD 2015, especially in areas of lower tree cover density and urban areas (see Figure 4).
- To avoid boundary effects at forest edges due to difference in spatial resolution between HRL TCD2015 and SWF products, a 10m buffer zone is then applied to the resulting mask (see Figure 6).
- SWFs located inside this TCD based new forest cover mask are systematically removed from the SWF layer. Figure 4 shows the difference between the HRL TCD 2015 and forest cover mask used during production, as well as why SWF and TCD layers can overlap while being complementary. (For complementary between SWF and TCD, see also Figure 3). SWFs located outside this mask (including low density forest areas with TCD values lower than 30%) are treated in a separate step in order to ensure the inclusion of SWF in fragmented areas representing isolated groups of trees and avoid SWF located in continuous forest (even of low density). These detected features are removed during the manual enhancement step. Further details on how which features are excluded or included by the new forest cover mask can be found in the SWF Product Specification Document (AD04).

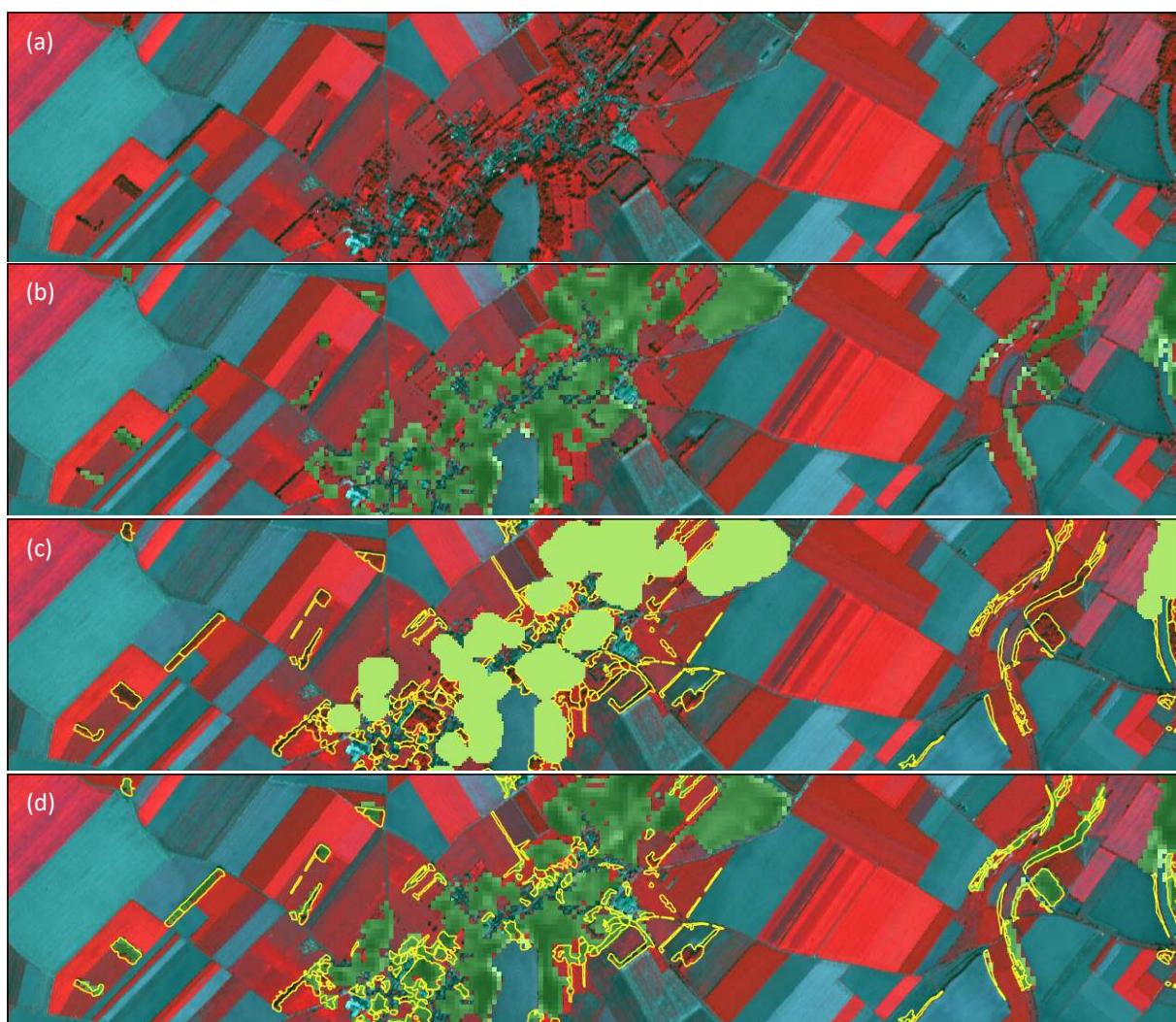


Figure 4: Application of forest mask based on HRL TCD 2015. (a): VHR2015; (b): HRL TCD 2015; (c): small and additional woody feature (yellow outline) and newly generated forest cover mask (light green); Small and additional woody features (yellow outline) and HRL TCD2015: the 2 products are partially overlapping, but also complementary for features too small to be included in HRL TCD 2015.

3. Validation approach

The validation approach will provide guidance on how the products will be validated by defining suitable indicators or metrics.

Detailed completeness and logical consistency checks are already performed as part of the semantic checks undertaken by the QC Tool during upload of the products. Therefore, the aim of this validation exercise is not to repeat these, but to review the existing documentation and perform additional checks if deemed necessary. We recommend however, to make the QC Tool reports available openly as additional quality information to the users.

The quality assessment is performed according to INSPIRE Data Specifications. The data quality elements considered are: (i) Completeness, (ii) Logical Consistency, (iii) Positional Accuracy, (iv) Thematic Accuracy, (v) Temporal quality and (vi) Usability. Each of them forms a section in the Validation Check list.

Logical consistency checks do not consist in a duplication of semantic checks but are performed to identify missing information if relevant.

The thematic accuracy will represent the bulk of the work undertaken as part of this validation exercise.

3.1. Completeness

Description: For land cover and land use products (both raster & vector), the notion of completeness in INSPIRE provides an indication of omission and commission errors. Commission - excess data present in the dataset, as described relative to the scope. Omission - data absent from the dataset, as described relative to the scope. This operation will be applied at the dataset level, rather than spatial object, and related to area extent. It can also include attributes and whether they are set etc., but this is actually covered by further checks below.

Indicators: the rate of excess items is used for areas mapped beyond the intended area and the rate of missing items is used to verify gaps in the intended area to be mapped.

3.2. Logical consistency

Logical consistency evaluates the degree of adherence to logical rules of data structure, attribution and relationships. In INSPIRE Data Specifications, Logical Consistency comprises four sub-elements described hereafter: conceptual consistency, domain consistency, format consistency and topological consistency. As the SWF, AWF and SWF+AWF products are in raster format the topological consistency is not relevant.

3.2.1. Conceptual consistency

Description: Conceptual consistency relates to the data structure and follows the data specifications in terms of data model and relationships. It is also related to the adherence to the rules of the conceptual schema.

Indicators: For the SWF, AWF and SWF+AWF layers the properties of conceptual consistency are mainly defined for the 5m SWF product, with the inference that the same are applied to the versions aggregated to 100m spatial resolution. The properties appropriate to raster data can be defined as:

- Minimum Mapping Unit (MMU). This is the minimum size a feature may have within the dataset. In the case of raster datasets this is the spatial resolution or grid cell size. For the SWF layers the spatial resolution should be 100 m.
- Coordinate Reference System (CRS): A coordinate-based local, regional or global system used to locate geographical entities and defines a specific map projection. CRS will be dealt with in detail later, but for the Forest layers it is the European LAEA.
- Pixel size and origin
- Additional attributes, symbology

3.2.2. Domain consistency

Description: involves the detection of attribute values that are outside the pre-defined range of values. For vector data each attribute has a pre-defined set of range of values. Domain consistency in raster datasets relates to the various range structures for bands and attributes, e.g. number of available bands with their names, the units of measure, the data type and the null value used. Checking domain consistency involves assessing the numbers of bands and the detection of attribute / pixel values that are outside the pre-defined ranges or sets of values. For raster data such as SWF, AWF, SWF+AWF, the correct encoding of data is checked.

Indicator: Value domain non-conformance: number of items not in conformance with their expected value domain.

3.2.3. Format consistency

Description: Format consistency includes detection of file format, file or attribute names or attribute types which do not correspond to the specifications. In addition, for raster data the pixel depth is also considered here. File format, schema, naming conventions etc. Degree to which data is stored in accordance with the physical structure of the dataset, as described by the scope.

Indicators: For the SWF layers which are raster-based products plus documentation, only the following format consistency properties need to be checked:

- File format conformance
 - o Raster data is required in the GeoTIFF or TIFF with World file formats.
 - o Compression: It was proposed that the image data should be optimised for storage by use of the LZW compression, which is supported in all image processing and GIS software, including GDAL.
 - o A PDF file providing CRS information, including details of parameters used to transform to ETRS89.
 - o The metadata should be supplied in XML format.
- File name conformance: A set of filename conventions were established in the product specifications.
- Attribute names conformance
- Attribute types conformance

3.2.4. Topological consistency

Description: topological consistency is applicable to vector data and describes the degree of correctness of the topological characteristics described in the product specification of the dataset.

Indicators: Not applicable to raster data

3.2.5. Additional logical consistency checks

There are further logical consistency checks that should be made to make sure the data confirms with the specifications to allow ease of use.

- **Labelling or symbology:** The conformity of a layer with the symbology or style given in the product specifications should be checked. The EEA provided colour tables to be built into raster datasets for all the HRLs.
- **Map projection:** The conformity of the map projection parameters is also checked. The selected projection for the HRL data is the LAEA-ERTS89 (Table 3.1).

Table 3.1: Map projection details for LAEA-ERTS89.

| European | | |
|----------------|--------------------|---|
| Datum | Name | ETRS89 (European Terrestrial Reference System 1989) |
| | Type | geodetic |
| | Valid area | Europe / EUREF |
| Prime meridian | Name | Greenwich |
| | Longitude | 0° |
| Ellipsoid | Name | GRS 80 (New International) |
| | Semi major axis | 6 378 137 m |
| | Inverse flattening | 298.257222101 |
| Projection | | Geographic (Ellipsoidal Coordinate System) |

3.3. Positional Accuracy

Detailed positional accuracy as described below is only required for the validation of image mosaics. Positional accuracy of the HRLs is directly related to the underlying HR and VHR imagery. Visual checks were undertaken in relation to imagery used for validation and during the collection of sample units.

3.4. Thematic Accuracy

3.4.1. Level of reporting

The level of reporting for the validation results is at pan-European level. However, results are also provided at different levels of aggregation (as indicated in the Request for Services for this first specific contracts). The analysis at disaggregated levels will contribute to assess regional differences, if any, and the nature of these differences.

The internal validation needs to find a compromise between the number of sample units and representativeness of the results at sub-European level. Therefore, the envisaged levels of reporting are:

1. Pan-European

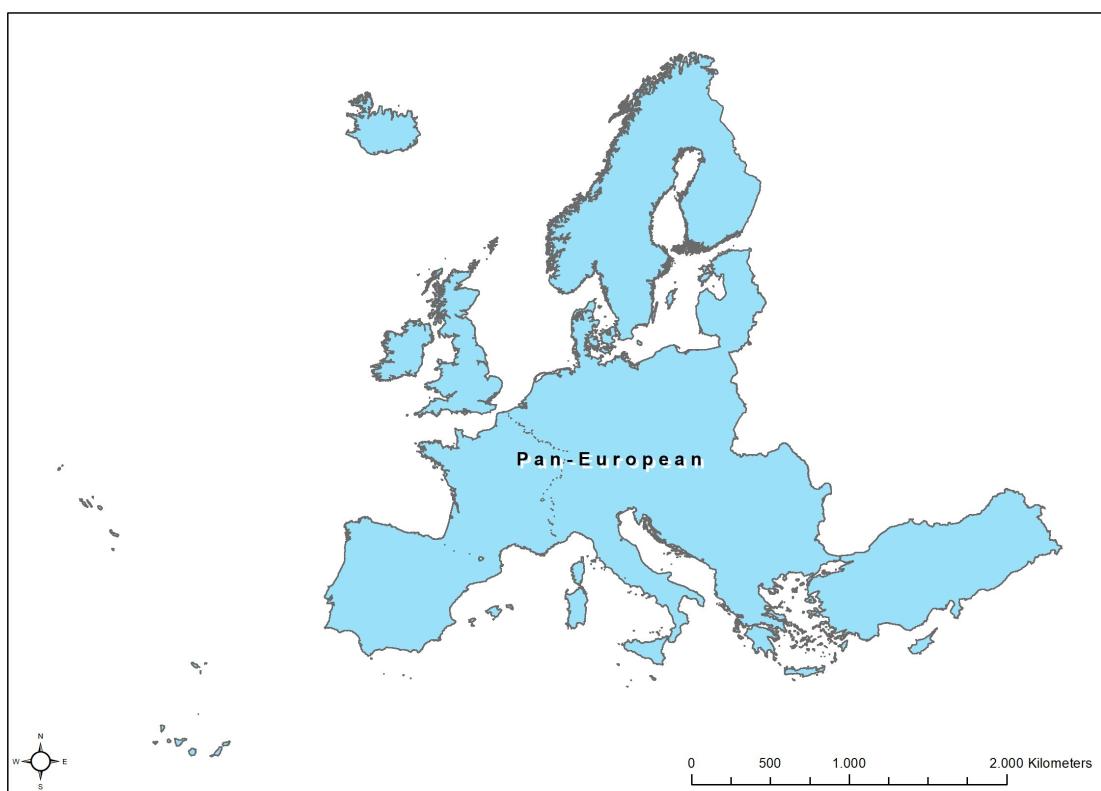


Figure 5: Level of reporting at pan-European level.

2. Biogeographical regions 2016

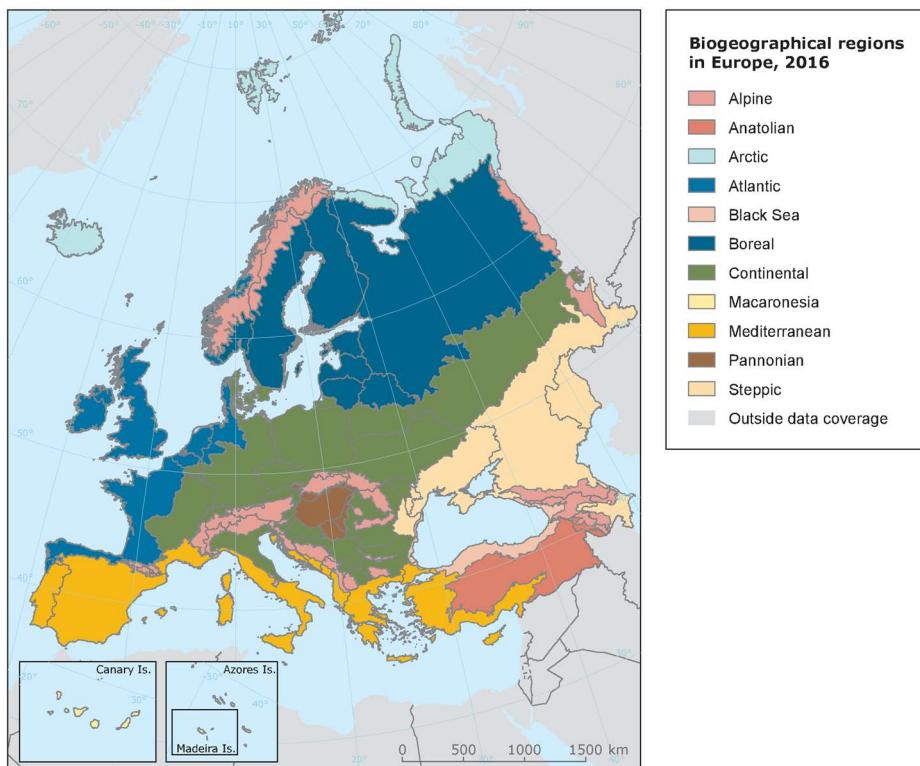


Figure 6: Level of reporting accordingly to the biogeographical regions (EEA, 2016).

3. Country or aggregated groups of countries as described in the previous Specific Contracts. 23 main countries or groups of countries (including French DOMs) are identified. Countries < 90,000 km² shall be grouped into contiguous groups of countries > 90,000 km².



Figure 7: Level of reporting according to the country or aggregated countries

3.4.2. Stratification and sample design

The following sub-sections provide a description of the procedure of a scientifically and statistically sound sampling scheme for assessing the thematic quality of the HRL SWF products. This comprises descriptions of the stratification approach, the sample size calculation procedure and of the strategy for ensuring representative sample distribution and sufficient regional spread. The validation concept for SWF layers and the corresponding sampling design are described in detail in the SWF Validation Concept (https://land.copernicus.eu/user-corner/technical-library/clms_hrl_swf_validation_concept_sc03_1_1-2.pdf).

3.4.2.1. Overview

The stratification and the sampling design primarily consists in selecting an appropriate sampling frame and sampling unit. The sampling units can either be “defined on a cartographic representation of the surveyed territory” (Gallego, 2004), in which case it is an area frame, or on a list of the features. According to (Gallego, 2004), area frames give a better representation of the population as the spatial dimension is kept.

In an area frame, sample units can be points, lines (often referred to as transects) or areas (often referred to as segments, described by Gallego, 1995). The first step is to define the geographical area for which the accuracy assessment is to be reported and the type of sample units. For the majority of cases, point samples will be used, but areas or segments may be used in specific cases such as when not only thematic accuracy needs to be reported, but also the geometry of mapped objects. Points are considered as the most appropriate unit for our purpose. Polygons have also the drawback of being specific to a single map. In case of changes, the sample may not be adapted anymore.

Sampling design refers to the protocol whereby the samples are selected. A probability sampling design is preferred for its objectivity. “Simple random, stratified random, clustered random and systematic designs are all examples of probability sampling designs” (Stehman *et al.*, 1998). Even though a simple random design is easy to implement, its main drawback is that some portions of the population may not be adequately sampled. Cluster sampling is often used to reduce the costs of the collection of reference data, but does not resolve geographic distribution problems. A systematic approach would solve this problem, yet it is not appropriate if the map contains cyclic patterns. A stratified approach consists in allocating a pre-defined number of samples per land-

cover class. As explained in (Stehman *et al.*, 1998), stratification ensures that each class is represented. For comparison see Figure 8.

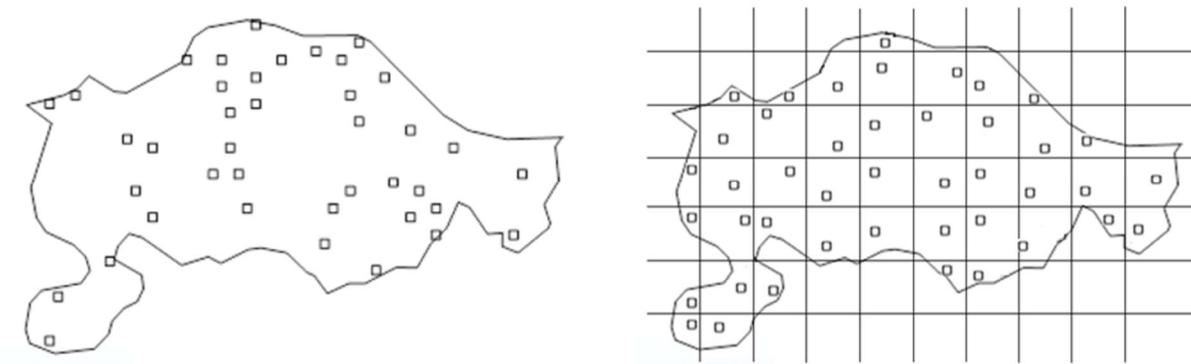


Figure 8: Simple random (left) and random systematic (right) sampling designs.

The sampling and stratification design presented below is applicable to all HRL SWF products at 100m spatial resolution.

The validation approach of the Copernicus land Lot 1 external validation contract were adopted for the internal validation. It is based on a selected sample design for thematic accuracy assessment and combines systematic and stratified approaches and benefits from the advantages of both of them. It is based on the LUCAS (Land Use/Cover Area frame statistical Survey) sampling approach. LUCAS corresponds to a grid of approximatively 1,100,000 points throughout the European Union where land cover or land use type is observed. Using LUCAS points ensures traceability and coherence between the different layers.

LUCAS points are located every 2 km on a regular grid, as illustrated below in Figure 9. A set of 81 points located on an 18x18 km square constitutes a group in which every point is associated with a number comprised between 1 and 81 (the numbers do not follow each other spatially). The same pattern with the same numbers allocation is repeated all over the grid. A replicate refers to the points with the same number selected on the whole LUCAS grid.

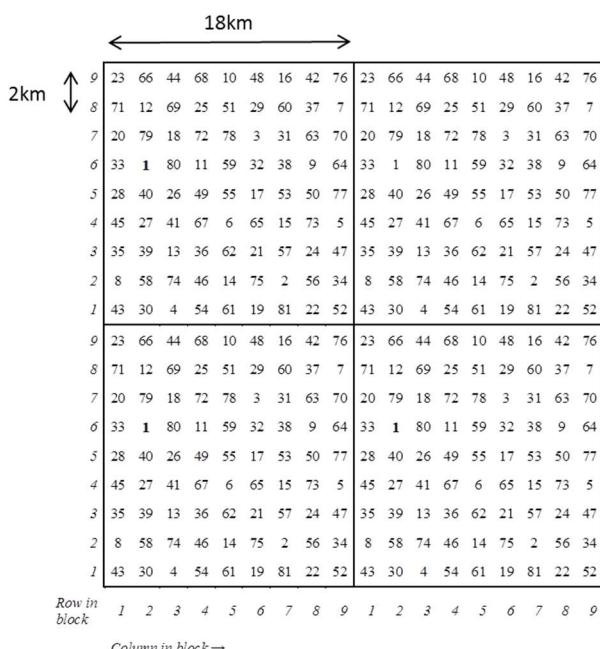


Figure 9: LUCAS points located on a regular grid.

At first, the number of samples to allocate to each stratum (or thematic class) is calculated as a function of their area. In this manner the sampling design is not only systematic but also stratified. The number of sample units per stratum is to be defined to ensure sufficient level of precision at reporting level:

The determination of the number of sample units also considers the number of thematic classes.

It is possible to estimate a suitable sample size for each stratum based on the expected acceptable error rate.

The standard error of the error rate can be calculated as follows: $\sigma_h = \sqrt{\frac{p_h(1-p_h)}{n_h}}$ (1) where n_h is the sample size for stratum h and p_h is the expected error rate. This can be reworked to express the sample size n_h as a function of p_h and desired standard error σ_h : $n_h = \frac{p_h(1-p_h)}{\sigma_h^2}$. (2)

From Figure 10 it can be seen that for an expected 50% error rate, within a stratum, 100 sample units would be required to guarantee a standard error of 5%, whereas the number of samples would need to be increased by a factor of four if the accepted standard deviation is divided by a factor of 2. On the other hand if the expected error rate is 15%, only 51 samples would be necessary with a 5% standard error. A similar approach was adopted to determine the sample size for assessing the accuracy of CLC2006 and CLC2000-2006 changes (Büttner *et al.* 2012). This works well to assess commission errors, the definition of an appropriate number of sample units for omission errors is more difficult because it depends on the expected area of the theme to be mapped.

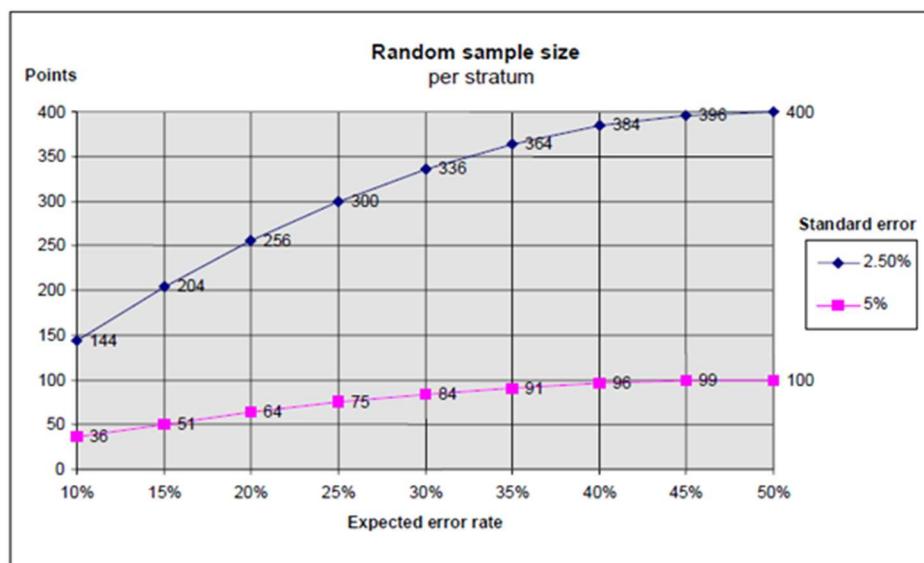


Figure 10. Number of sample points as a function of the expected error rate for two accepted standard error values (after Wack *et al.* 2012).

When using stratified sampling, the main issue to maximize the efficiency of the stratification (maximize the level of precision) is to optimize the sample allocation per strata. A simple way is the use of equal allocation. Alternatively, the Neyman allocation algorithm is also often used for that purpose:

$$n_h = n * (N_h * \sigma_h) / [\Sigma (N_i * \sigma_i)], \quad (3)$$

where n_h is the sample size for stratum h , n is the total sample size, N_h is the population size for stratum h , and σ_h is the standard deviation of stratum h . According to Stehman (2012), Neyman optimal allocation should be preferred for estimating area of change as well as overall accuracy, whereas equal allocation is effective for estimating user accuracy.

Moreover, in addition to the Neyman sample allocation, a minimum number of sample units per stratum is defined to ensure that even small strata are represented in the sample. As the expected error rate is expected to be less than 10%, a minimum of 50 sample units per reporting stratum should be sufficient.

The number of replicates to be selected for a stratum depends on its area and the number of LUCAS points intersecting the stratum.

For thematic classes covering a large proportion of the study area, 1 replicate may already exceed the defined number of samples for this class. To solve this problem, replicates are split into four sub-replicates, as illustrated by the blue numbers in Figure 11.

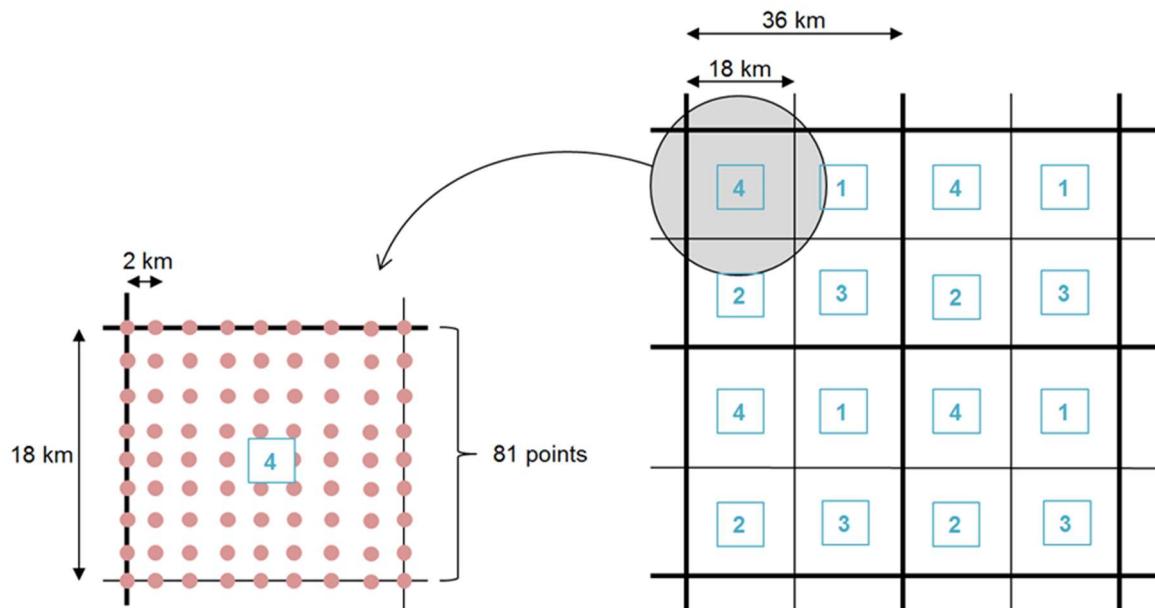


Figure 11: Replicates and sub-replicates used on LUCAS grid.

The opposite problem is encountered for land cover classes covering a small proportion of the study area: even by selecting 81 replicates (the maximum number), the intersecting area between the stratum and LUCAS points is too small to reach the required number of samples. Therefore LUCAS grid could be densified by creating one point every 200m.

3.4.2.2. Stratification approach

The selected sampling design for this validation approach combines **systematic and stratified approaches**. It is based on the **LUCAS** (Land Use/Cover Area frame statistical Survey) sampling approach, described above.

For the SWF, AWF, SWF+AWF products, based on the lessons learnt from the previous Specific Contracts, the stratification procedure was simplified to only include a series of omission/commission strata applied at pan-European level:

1. Omission strata: SWF or AWF density = 0%
2. Commission strata: SWF or AWF density > 0%

The French DOMs were sampled and stratified apart in a second step since part of the national products (and not the pan-European full mosaic).

In addition, we used the intermediate forest cover mask to exclude areas from the sampling. Thus, no samples are selected within the forest cover mask. Samples are thus only located in non-forest areas, except for errors that may occur in the forest cover mask. Also areas outside or within gaps in the SWF layers (parts of Norway) are excluded from sampling (see Figure 12; no data areas).

SWF 2015 - Availability -

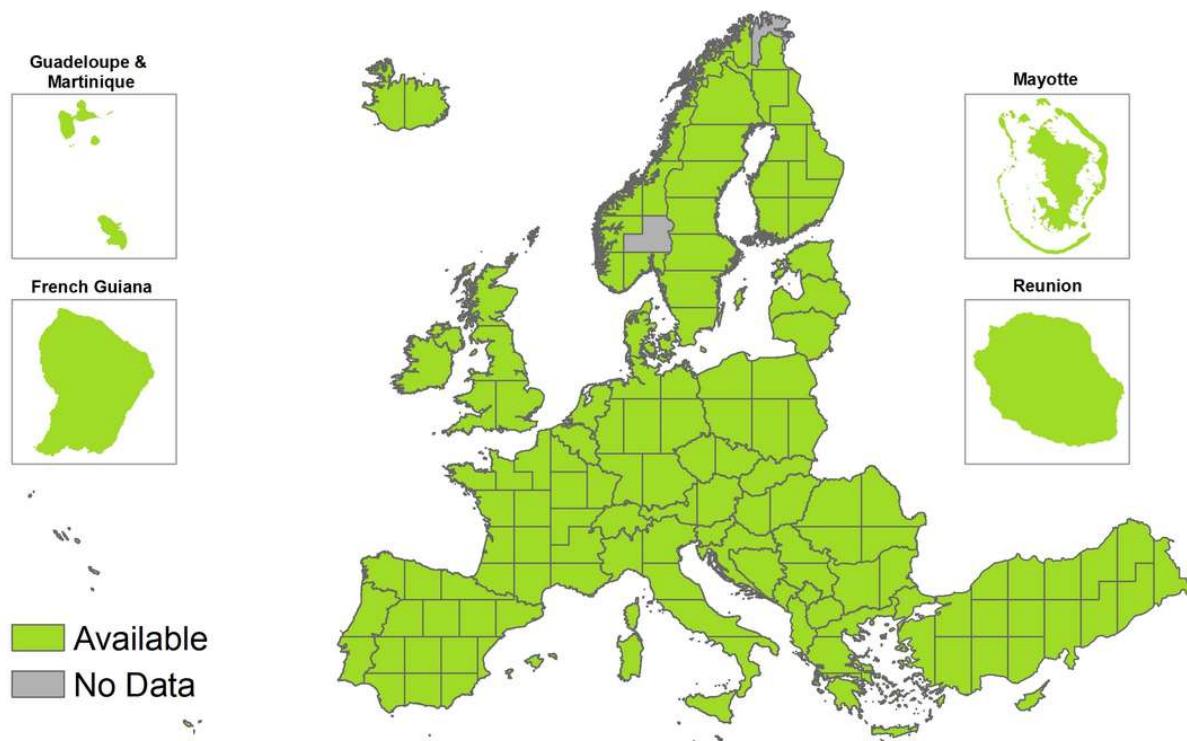


Figure 12: SWF 2015 availability and production units.

A minimum of 30-50 samples per stratum of each production unit should provide a standard error of 5-10% which should be sufficient to detect issues at production unit level and provide a high degree of precision at pan-European level.

In this approach, a sample unit (PSU) is a 100x100m square snapped to the EEA grid, representing 1 HRL 100m product pixel. For density products such as SWF, also SSUs are used during the interpretation to determine a mean density for the 1ha plot (PSU). A grid of 5x5 points (SSUs) with a 20m spacing is used for SWF layer interpretation (Figure 13), similar to e.g. the HRL Forest TCD layer validation. The analysis however, is only performed at PSU level by applying density thresholds and thus producing binary layers for which the accuracy can be assessed.

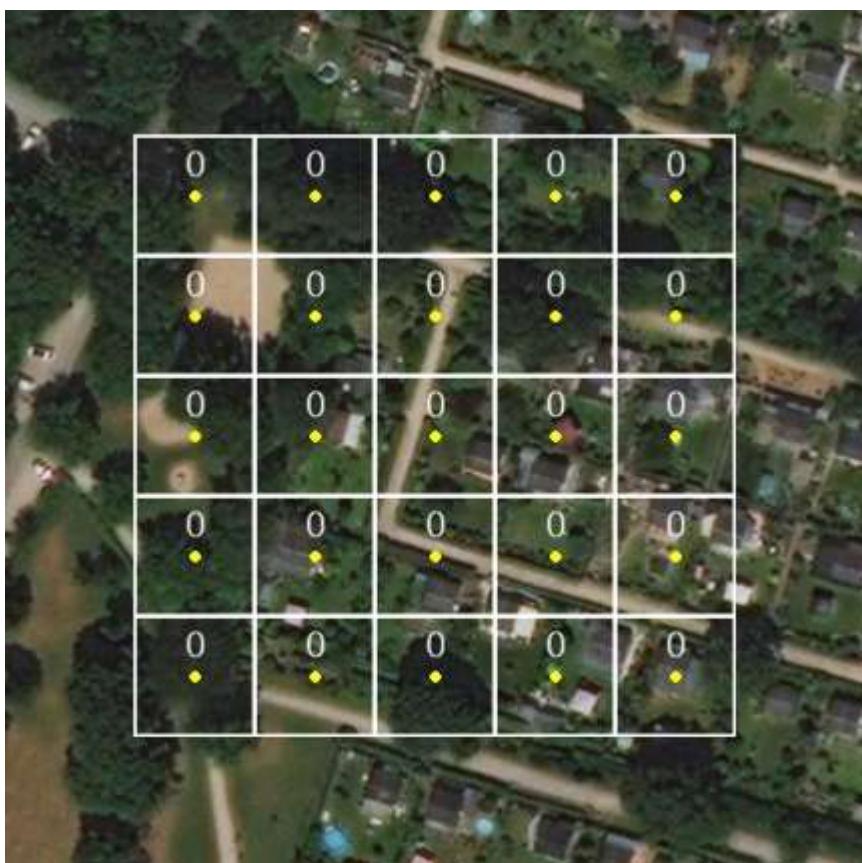


Figure 13: Example of SSUs organised in a 5x5point 20m grid.

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

In total, 25,000 sampling points were interpreted and used to validate the products (Table 3.2, Table 3.3 and Table 3.4). Densities for SWF and AWF are interpreted at each sampling plot. For the SWF+AWF analysis we add the values from SWF and AWF during final analysis. A small number of plots had to be eliminated from the final analysis as these plots had low reliability due to low quality reference imagery or significant land use changes within the valid time window.

Table 3.2: SWF sample plot distribution by level of stratification at pan-European level.

| LABEL | Commission | Omission | Total |
|---------------------|---------------|--------------|---------------|
| Pan-European level | 18,066 | 6,834 | 24,900 |
| French Guiana level | 10 | 10 | 20 |
| Guadeloupe level | 10 | 10 | 20 |
| Martinique level | 10 | 10 | 20 |
| Reunion level | 10 | 10 | 20 |
| Mayotte level | 10 | 10 | 20 |
| TOTAL | 18,116 | 6,874 | 25,000 |

Table 3.3: SWF sample plot distribution per strata and area in km² at pan-European level (excluding forest cover mask).

| EEA39 | Area (km ²) | |
|--|-------------------------|----------------|
| Zone | Omission (0) | Commission (1) |
| PAN | 2 309 037 | 873 361 |
| GF | 732 | 142 |
| GP | 299 | 398 |
| MQ | 168 | 236 |
| RE | 481 | 375 |
| YT | 48 | 52 |
| (0) omission strata = No SWF | | Total: |
| (1) commission strata = SWF+AWF | | |

Table 3.4: SWF distribution of sample units per countries or groups of countries.

| LABEL | Commission | Omission | Total |
|------------------|---------------|--------------|---------------|
| AL+ME+MK+RS+XK | 431 | 190 | 621 |
| AT + CH + LI | 320 | 116 | 436 |
| BA + HR + SI | 193 | 140 | 333 |
| BE + LU+ NL + DK | 445 | 259 | 704 |
| BG | 363 | 95 | 458 |
| CZ + SK | 348 | 150 | 498 |
| DE | 1,008 | 564 | 1,572 |
| EE + LT + LV | 419 | 183 | 602 |
| EL + CY | 416 | 174 | 590 |
| ES | 1,687 | 594 | 2,281 |
| FI | 401 | 87 | 488 |
| FR | 1,424 | 1,134 | 2,558 |
| HU | 389 | 116 | 505 |
| IE + UK | 1,152 | 789 | 1,941 |
| IS | 764 | 7 | 771 |
| IT + MT | 765 | 359 | 1,124 |
| NO | 965 | 245 | 1,210 |
| PL | 994 | 397 | 1,391 |
| PT | 316 | 129 | 445 |
| RO | 911 | 238 | 1,149 |
| SE | 745 | 204 | 949 |
| TR | 3,610 | 664 | 4,274 |
| French DOM | 50 | 50 | 100 |
| Total | 18,116 | 6,884 | 25,000 |

In order to ensure that unequal inclusion probabilities are accounted for in the construction of the error matrix, weights are applied to each stratum as shown in (

Table 3.5):

Table 3.5: SWF weight factors to be applied to each stratum and level of stratification for constructing confusion matrices.

| LABEL | Commission | Omission |
|---------------------|------------|----------|
| Pan-European level | 1.003124 | 1.003008 |
| French Guiana level | 0.574846 | 0.111668 |
| Guadeloupe level | 0.234481 | 0.312723 |
| Martinique level | 0.131666 | 0.185569 |
| Reunion level | 0.377653 | 0.294161 |
| Mayotte level | 0.037579 | 0.041181 |

The calculate the weights based on the interpreted sample units. When the number of sample units is modified in a stratum, because e.g. some samples were classified as uncertain, the weights need to be recalculated. However, if the number of samples removed is small, weight changes are minimal considering the large sample overall.

The sample units were provided to the bulk interpretation team as a shapefile, where the information on strata was removed to ensure the independence of the interpretation.

3.4.3. Response Design

3.4.3.1. Overview

The sampling design is based on the LUCAS sample frame and some sample units will coincide with some LUCAS points. However, all LUCAS points (PSUs) are re-interpreted based on available in situ data. LUCAS thematic information is not used.

The first sample interpretation is based on a blind approach, i.e. the validation is undertaken without considering the map layer information, and then a plausibility analysis was implemented for both layers considering the information of the map layer. This will contribute to provide an in-depth understanding of the causes of classification errors and to account for ambiguous map values at the class borders.

3.4.3.2. SWF and AWF Layers

The SWF and AWF layers at 100m spatial resolution were evaluated according to the following product definitions and aggregation rules (cf. SWF 2015 Product Specification Document: https://land.copernicus.eu/user-corner/technical-library/hrl_lot5_d5-1_product-specification-document_i3-4_public-1.pdf).

- Density 0-100%: to be interpreted depending on the level of detail that can be detected from the available VHR image 2015 datasets
- Interpretation follows the thematic and geometric definitions outlined in Table 1.1 and Table 1.2.
- In addition to the above mentioned geometric definitions, a 3m minimum mapping width (MMW) is applied for linear elements. This metadata can be found at the Copernicus Land website (<https://land.copernicus.eu/pan-european/high-resolution-layers/small-woody-features/small-woody-features-2015?tab=metadata>). The reason is, that smaller features cannot be detected from the VHR 2015 dataset

For a better perception of the compactness factor used for the differentiation and classification of SWFs (threshold 0.75), Figure 14 shows different compactness classes. For SWF Compactness during SWF production was calculated as follow:

$$C = \frac{(4 \cdot \pi \cdot \text{area})}{\text{perimeter}^2}$$

The 0.75 threshold was determined in order to include in linear elements the shortest/widest hedgerow allowed by the specs (i.e. hedgerow of 50m long and 30m wide).

However, this definition is not explained in the current SWF product specifications. Since there are different ways to determine the compactness, we recommend to include this information in future SWF product specification documents.

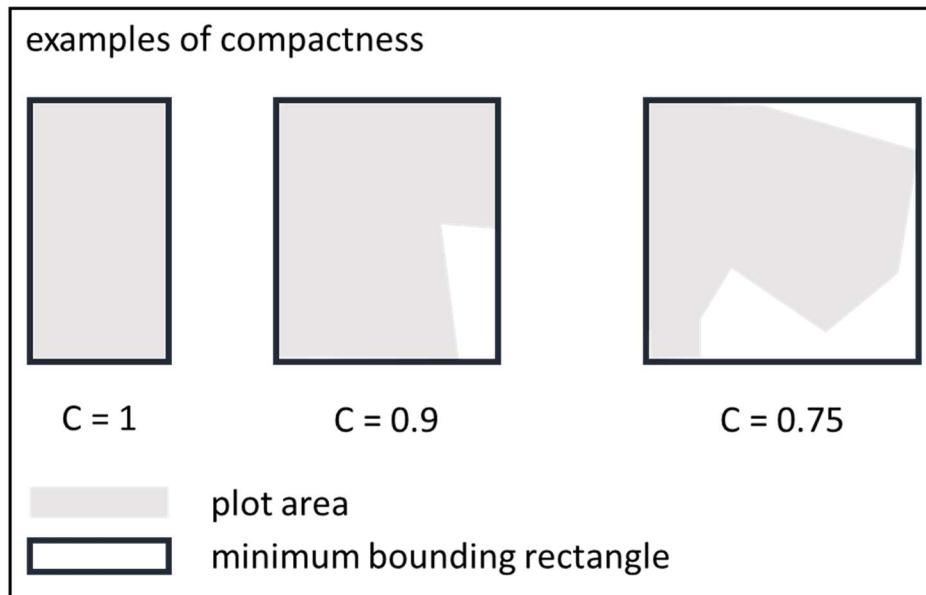


Figure 14: Examples of different compactness factors based on a bounding rectangle.

SWF are mapped in the following categories:

- Linear SWF: represent landscape features such as hedgerows or tree alignments that are defined by a compactness criterion less or equal to 0.75, up to 30m width and at least 50m length.
- Patchy SWF: represent areas of isolated and scattered patches of trees or scrubs defined by a compactness criterion greater than 0.75, at least 10m width and with an area greater than 200m² and less than 5,000m².
- Additional Woody Features: Woody features that are neither linear nor patchy SWF, but which are connected to linear or patchy SWF and isolated woody features that are not linear nor patchy SWF, but which present an area above 1500m² (linear features wider than 30m, and out-of-specifications patches). Unfortunately, no definition is currently provided of what “isolated” means in this context. A minimum distance of isolated woody features to other woody features is recommended for future product descriptions.

SWFs include: linear hedgerows and scrubs, tree rows (e.g. along field boundaries), isolated/scattered patches of trees.

SWFs exclude: 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.

In addition to above-mentioned rules, several specific mapping rules are applied to derive a homogeneous pan-European SWF database. Examples of such mapping rules are:

- Linear structures may contain feature parts wider than 30m if connecting 2 features of less than 30m width, each longer than 50m, over a distance of less than 50m – otherwise this would result in rather artificial cuts in linear structures;
- Patchy structures may have a width of less than 30m over a distance below 50m. A cut is not applied in order to keep the overall, natural characteristic of the identified patch.
- Trees are considered as green linear structure when the gaps between the trees are smaller than 5m.
- Linear and patchy features within open forest are excluded from the SWF product
- SWF and AWF layers are mutually exclusive (non-overlapping).

3.4.3.3. SWF/AWF interpretation procedure

We apply a double-blind approach, where the map information is not available during interpretation. To ensure that the reference interpretation is of higher quality than the map classification, we also use higher resolution reference imagery in addition to the production data (VHR 2015).

Used reference data, which cover the EEA39 area at a wall-to-wall basis:

- HR Image 2015 Coverage 1 CIR imagery with spatial resolution of 10-20m
- HR Image 2015 Coverage 2 CIR imagery with spatial resolution of 20m
- VHR Image 2015 with spatial resolution of 1-2.5m (JP2 pan-sharpened version made available from SIRS)
- HR Image 2012 Coverage 1 CIR imagery with spatial resolution of 25m
- HR Image 2012 Coverage 2 RGB imagery with spatial resolution of 20m
- VHR Image 2012 CIR with spatial resolution of 2.5m (Web map service provided by JRC)
- VHR Image 2012 RGB with spatial resolution of 2.5m (Web map service provided by JRC)

Used reference data of higher spatial resolution, which cover only parts of the EEA39 area:

- Bing maps (ArcGIS Basemap layer, RGB imagery with varying spatial resolution)
- Google Earth imagery (Google commercial ArcGIS plugin, RGB imagery with varying spatial resolution)
- Google StreetView
- National and regional web mapping services (RGB and/or CIR imagery with varying spatial resolution)
- Sentinel-2 data via Sentinel-Hub Playground

In addition to the reference data, also the forest cover mask from SWF production was used to eliminate forest areas from the interpretation, as some plots (1ha PSUs) partly intersect with areas within the forest cover mask. The forest cover mask has a minimum mapping unit of 5ha. Thus, forest areas < 5ha may be considered as AWF, even though this is not explicitly stated in the SWF product specifications.

These comprehensive reference data sets provided a sound basis for the interpretation. For labelling of the reference data, an html-based tool, the EEA-Validator, was implemented by Joanneum Research and is exemplarily shown in Figure 15. For estimating the SWF and AWF densities, we interpreted if the centre point is located on a woody feature for each of the 25 grid cells (SSUs). Further, for each of the PSU sampling units, the interpreter assigns a reliability estimate.

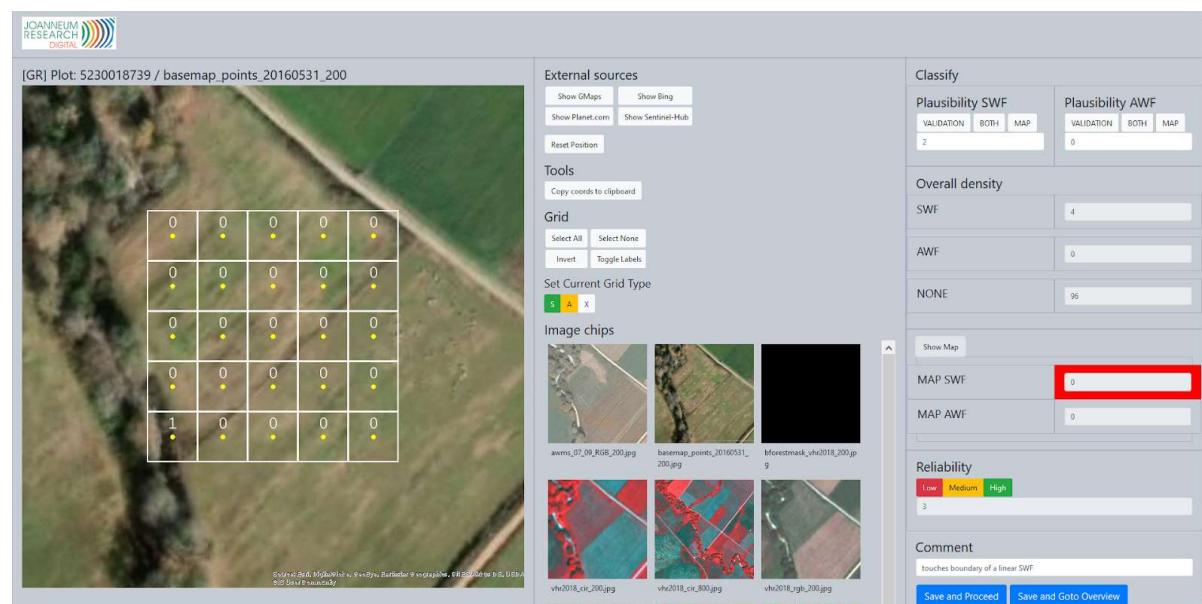


Figure 15: SWF Interpretation tool for reference data labelling (implemented by Joanneum Research). The tool provides relevant image chips, single grid cell value selection, density value summaries, and can handle a plausibility analysis. The red encircled area shows that there a discrepancy between map SWF value and interpreted SWF value.

We used the following step-by-step interpretation procedure to interpret the SWF 2015 product. We carried out the interpretation for the 100x100m PSU plot level at a secondary sampling unit consisting of 25 grid cells with a centre point. If a point is located on a SWF, a 4% density value is assigned to this SSU. Per default all 25 grid cells (SSUs) within the PSU plot are flagged as NONE, meaning no SWF or AWF are present. First, every detected woody feature is compared with the SWF geometric rules. If this feature meets SWF linear or SWF patch specifications, it is labelled as SWF. If not, it is determined if it meets the AWF specifications.

- i. Determine if small woody features are present in the PSU sampling plot, which are not located within the forest cover mask:
 - a. Are linear woody features present that have a minimum width of > 3m and a maximum width of \leq 30m (the maximum width may be wider for a maximum distance of 50m when connecting two linear woody features with a width \leq 30m and at least 50m long) and a minimum length of \geq 50m and a compactness \leq 0.75 and where the maximum distance between single trees/bushes is 5m?
 - i. If Yes: flag all SSU grid points located on these linear SWFs as SWF
 - ii. If No: proceed with b.
 - b. Are patchy woody features present that have a minimum mapping area \geq 200m² and a maximum mapping area \leq 5000m² (0.5ha) and a compactness $>$ 0.75?
 - i. If YES: flag all SSU points located on these patchy SWF as SWF
 - ii. If NO: proceed with ii.
- ii. For all grid cells that are not flagged as SWF by the above procedure (i.), determine if additional woody features are present
 - a. Are woody features present that are not located within the forest cover mask?
 - i. If YES:
 1. Does the feature belong to a forest area $>$ 5ha that is by error not included in the forest cover mask? -> flag these areas as NONE
 2. Is the feature connected to SWFs and thus enhances their connectivity? -> feature is AWF: flag all SSU grid points located on this feature as AWF
 3. Is the feature isolated and is its area $>$ 1500m² and $<$ 5ha. -> feature is AWF: flag all SSU grid points located on this feature as AWF
 - ii. If NO: All areas that are not SWF are flagged as NONE.
 - iii. In the rare case that the forest cover mask is wrong, SWF and AWF may also be present within the forest cover mask. The same rules as above apply, but the forest cover mask is ignored for the erroneous areas.
 - iv. A final check is then performed to determine if the flagged grid cells represent the overall density. It can happen that e.g. 5 grid points are located on a thin straight linear woody feature. This would result in 20% SWF for the entire plot even though the total area covered by the linear feature may not exceed 10%. In this case, the interpreter is advised to flag 2-3 points that are located on the linear SWF as NONE, so that the overall SWF density at PSU level (1ha) is correct.

Flag the quality of the data: All sampling units, to which the interpreter assigns a low reliability during interpretation, are removed from further product analysis. Reasons for the rejection of sample units are: Primarily, reference imagery of too bad quality, but also situations where no exact interpretation is possible, as for example significant land cover changes have occurred during the allowed 2-year time window. The interpretation also includes internal quality measures to ensure the quality during the validation process, e.g. interpreters may interpret a joint set of points. This allows to determine the consistency in interpretation when working with several interpreters.

After the blind interpretation, we apply a 1% threshold resulting in a binary map where 0 = no presence of SWF/AWF and 1 = presence of SWF/AWF. Sampling plots, where the interpretation and the map value disagree on the presence of SWF/AWF, were reinterpreted as part of the plausibility analysis.

During plausibility the interpreter must decide if:

1. only the blind interpretation is correct -> the interpreted reference values are not changed

2. only the map value is correct -> the interpreted reference values are changed to the map value
3. both the blind interpretation and the map value are possible to ambiguous situations (thresholds, area overlap between SWF and AWF, ...) -> the interpreted reference value is changed to the map value for the plausibility analysis, as the map is considered correct.

The subsequent plausibility analysis is based on the recoded interpretation values.

3.4.3.4. Interpretation examples

This section provides a number of examples from the interpretation for a better understanding of the interpretation procedure and the many specific mapping rules and thematic and geometric specifications that need to be considered. Some of the interpretation ambiguities are then easier to understand. Figure 16 shows a very common situation where no SWF and AWF are present in the plot. This is the easiest scenario for interpretation.



Figure 16: SWF interpretation example: no SWF and AWF are present, all grid cells (SSUs) are None (0).

Figure 17 shows an olive grove. By product definition olive trees, orchards and vineyards are not considered as SWF/AWF.

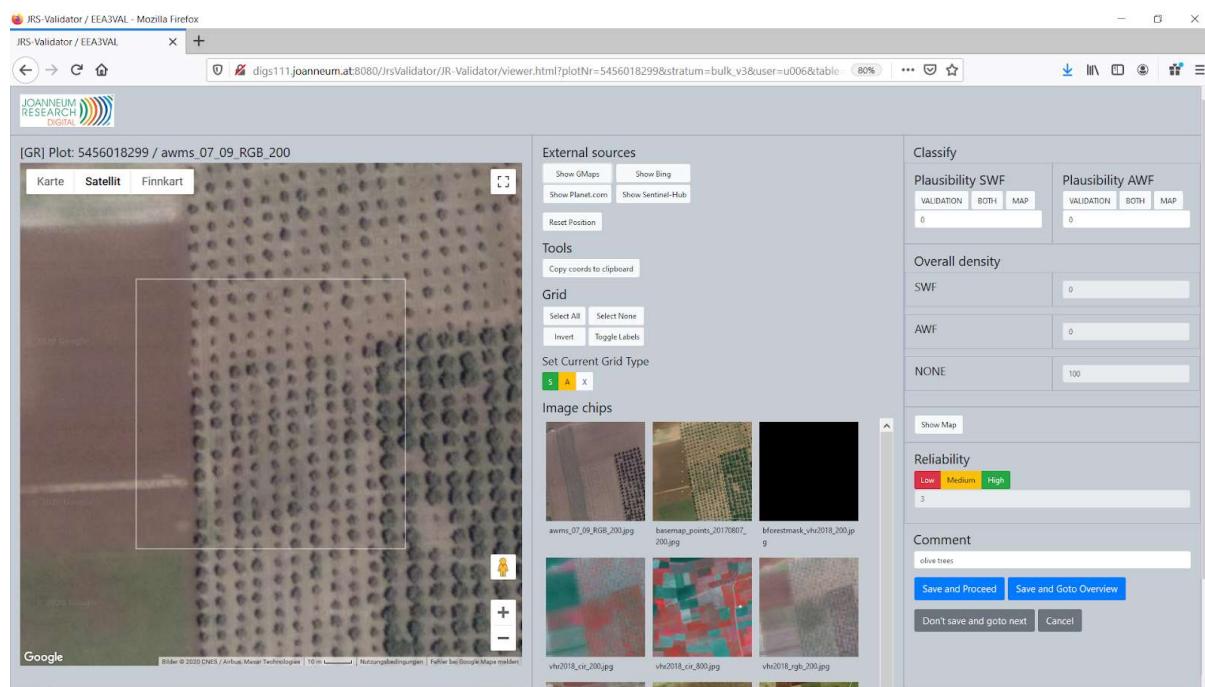


Figure 17: SWF interpretation example: olive trees or orchards are not SWF/AWF.

Figure 18 shows how the forest cover map from SWF production was integrated during the interpretation. SSUs that are masked out (green) are not considered during SWF/AWF density estimation unless the forest cover mask is obviously wrong.

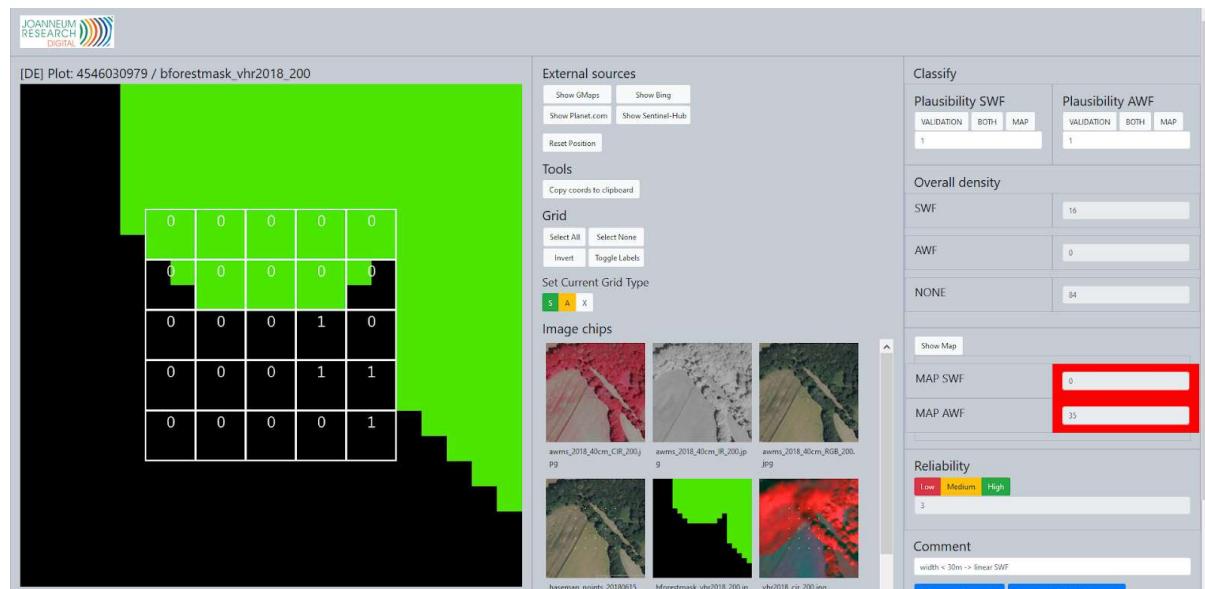


Figure 18: SWF interpretation example: areas within the forest cover mask are not considered.

In Figure 19 the forest cover mask (green area on right image) is ignored during interpretation of the plot, as it is obviously wrong for this plot.

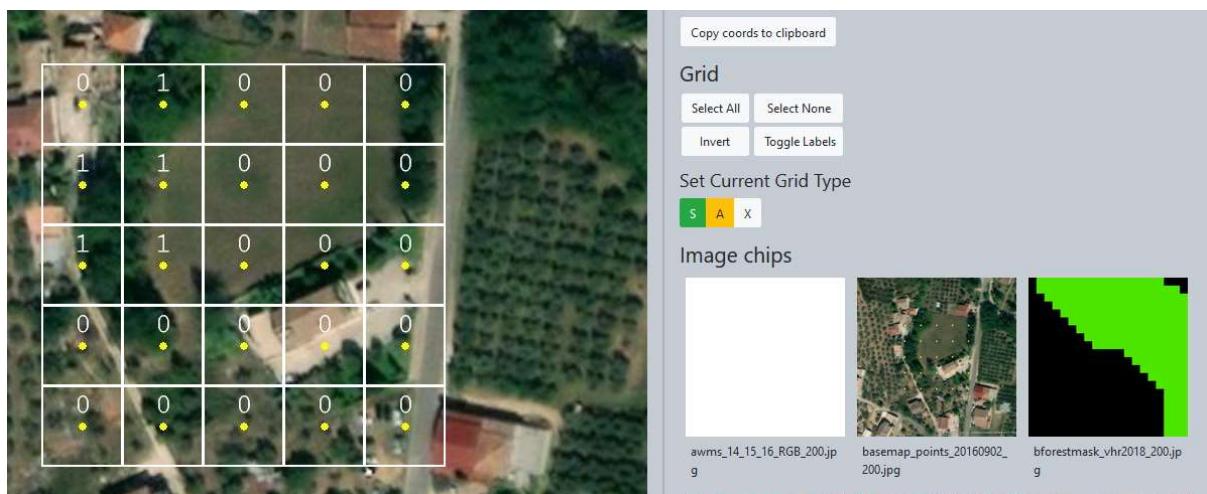


Figure 19: SWF interpretation example: For some plots, the forest cover mask was not considered due to obvious thematic errors.

Figure 20 gives an example of an open forest. Open forests are not included in the SWF or AWF layers by definition, but often they are not considered as forests in the forest cover mask (small back image has no green mask) and were only removed by manual editing at a later stage during SWF production.

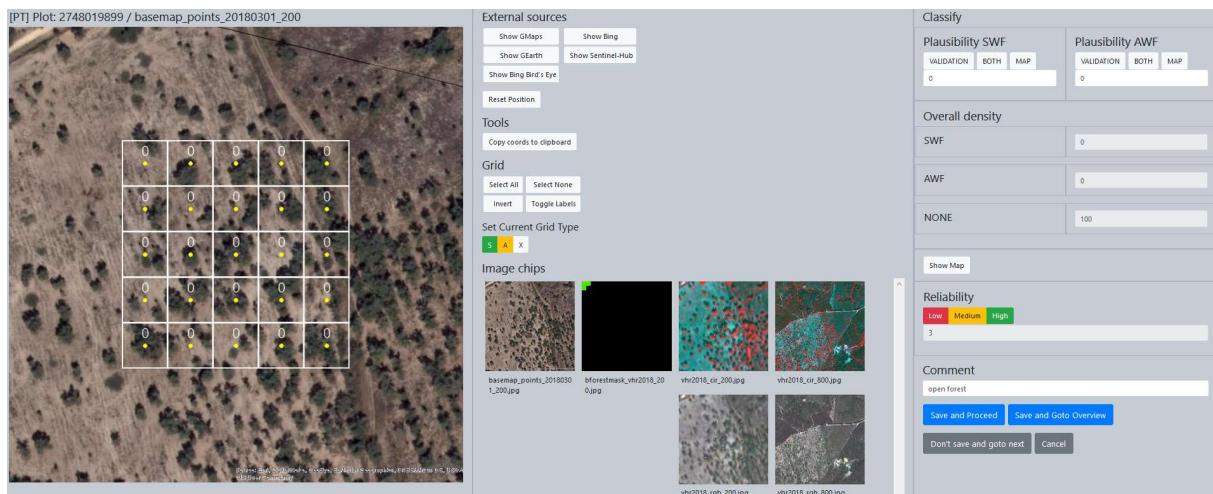


Figure 20: SWF interpretation example: open forests are not SWF or AWF even though they are not excluded by a forest cover mask at this sample plot.

Figure 21 exemplifies how the maximum mapping size is considered. This plot has a 57% SWF density value in the map, but the woody area clearly exceeds the maximum allowed size of a patchy SWF, which is 0.5ha. It must therefore be an isolated AWF.

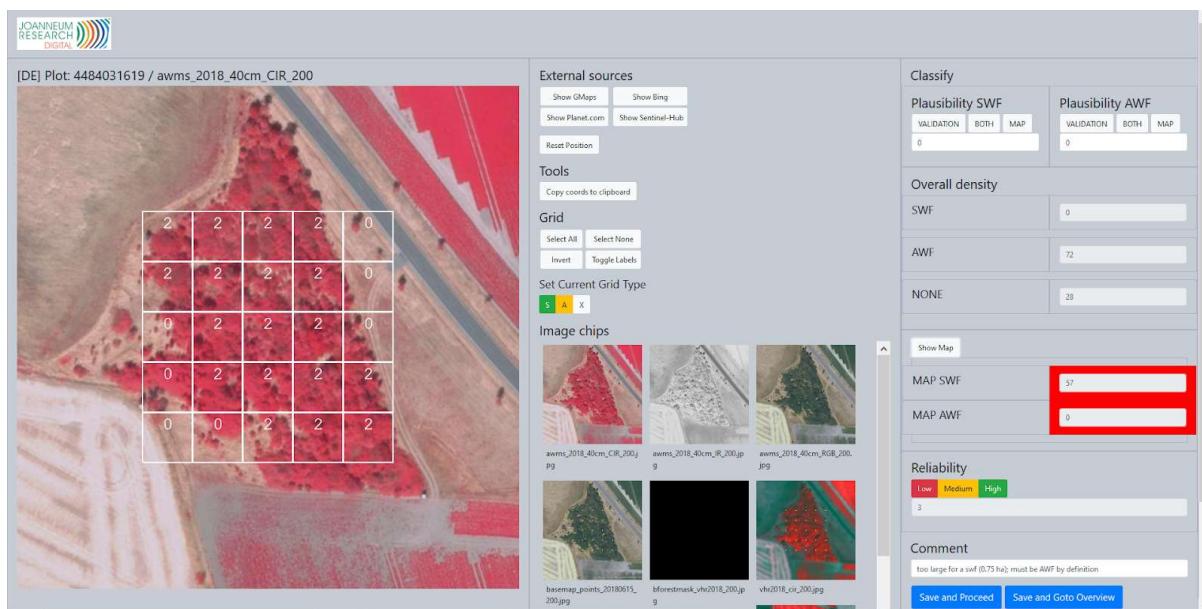


Figure 21: SWF interpretation example: Maximum mapping size for SWF exceeded.

Figure 22 is an example of how the 5m maximum distance between single trees for linear SWF is considered. Due to this rule, this plot has 0% SWF as the distances between single trees exceed 5m.

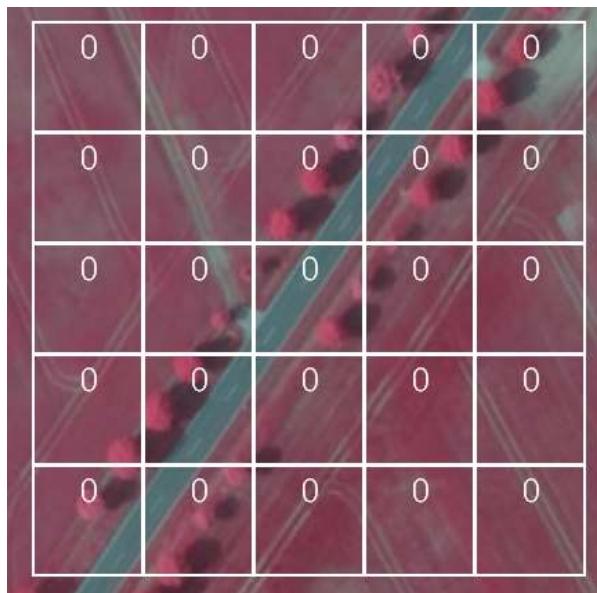


Figure 22: SWF interpretation example: example of the 5m maximum distance between single trees to be considered as linear SWF.

Figure 23 shows how the 30m threshold for separating linear SWF from AWF can be ambiguous as many woody features show a width around 30m. In this case, the width was measured to be between 30-35m, which is slightly too large for a linear SWF by product specifications. It is also not a patchy SWF. Since it connects other SWFs it is considered as an AWF in the blind interpretation. The map says only SWF are present in this plot. The plausibility assessment therefore says both blind interpretation and map are possible and the plot is considered correctly classified by the map in the plausibility analysis.

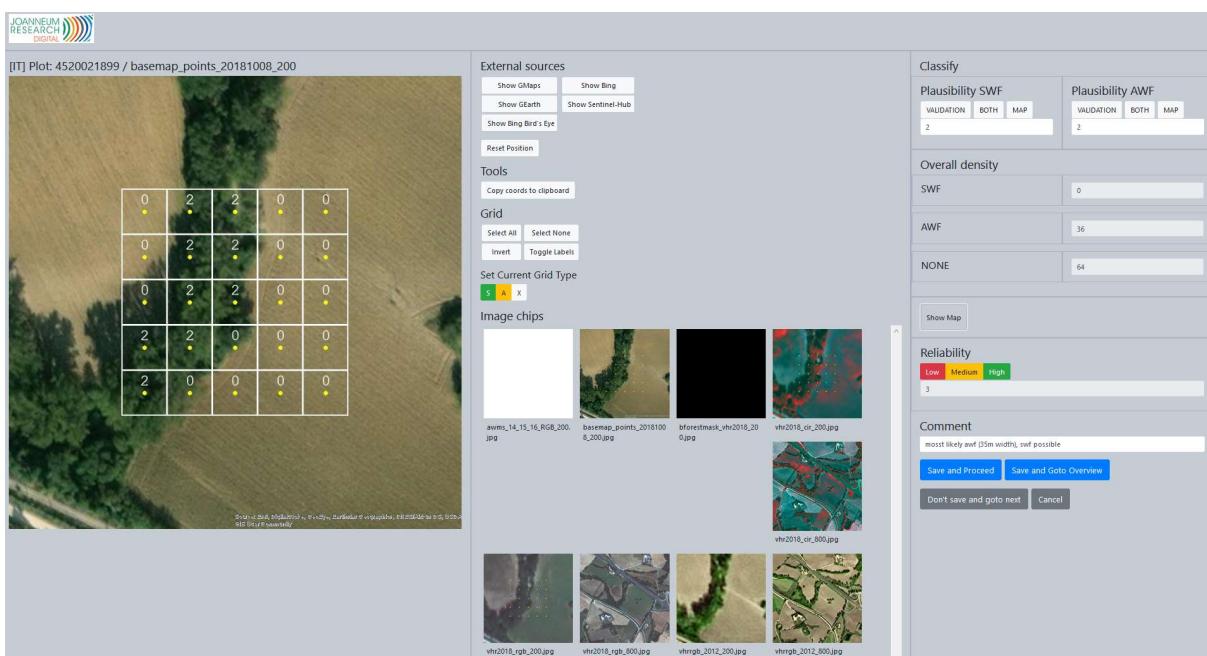


Figure 23: SWF interpretation example: ambiguous interpretations between SWF and AWF are possible due to the 30m width threshold applied to linear SWF.

Another ambiguous situation is shown in Figure 24. Here it is the 3m minimum mapping width for linear SWF, which is difficult to determine exactly in the reference imagery. The plot could thus have 0% SWF or 4% SWF, which makes a big difference during the analysis where only presence and absence of SWFs are considered. The many ambiguous situations encountered with the SWF product are a result of the large number of absolute thresholds used for the product specification and lead to rather large differences between blind and plausibility results.

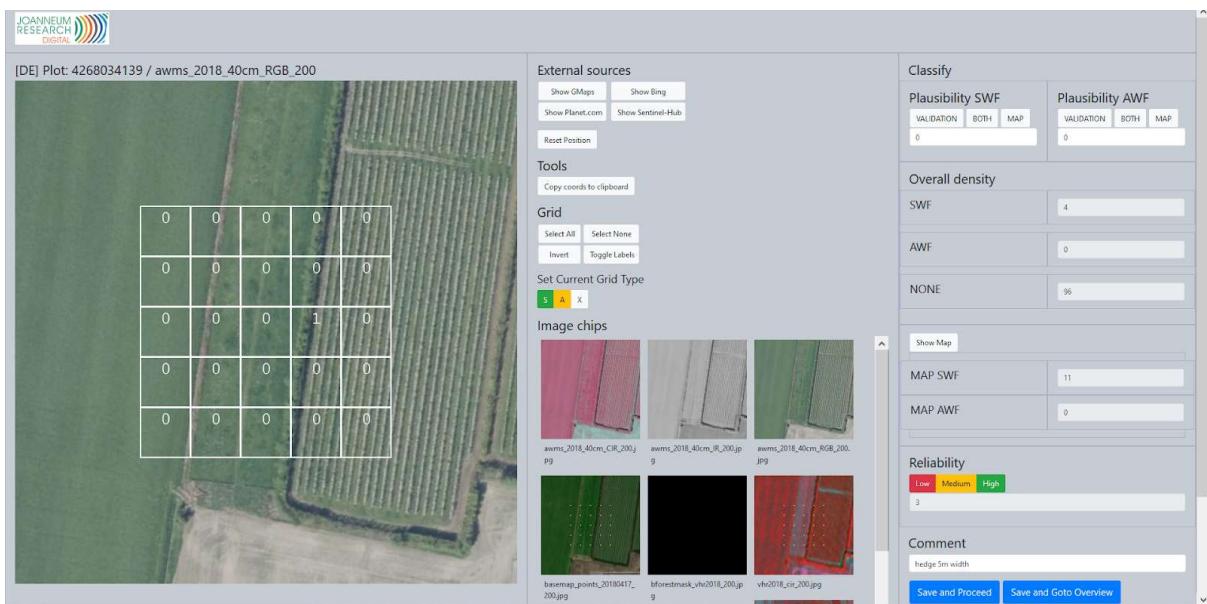


Figure 24: SWF interpretation example: ambiguous interpretations are possible also for the 3m minimum mapping width of linear SWF.

Figure 25 shows an example of mountain dwarf pines. These are not specifically mentioned in the thematic specifications of the SWF and AWF products. We considered them as either patchy SWF or AWF in the blind interpretation, but allowed for a different interpretation during the plausibility assessment if the map disagreed. We recommend including shrubs and dwarf pines in future thematic product specifications.

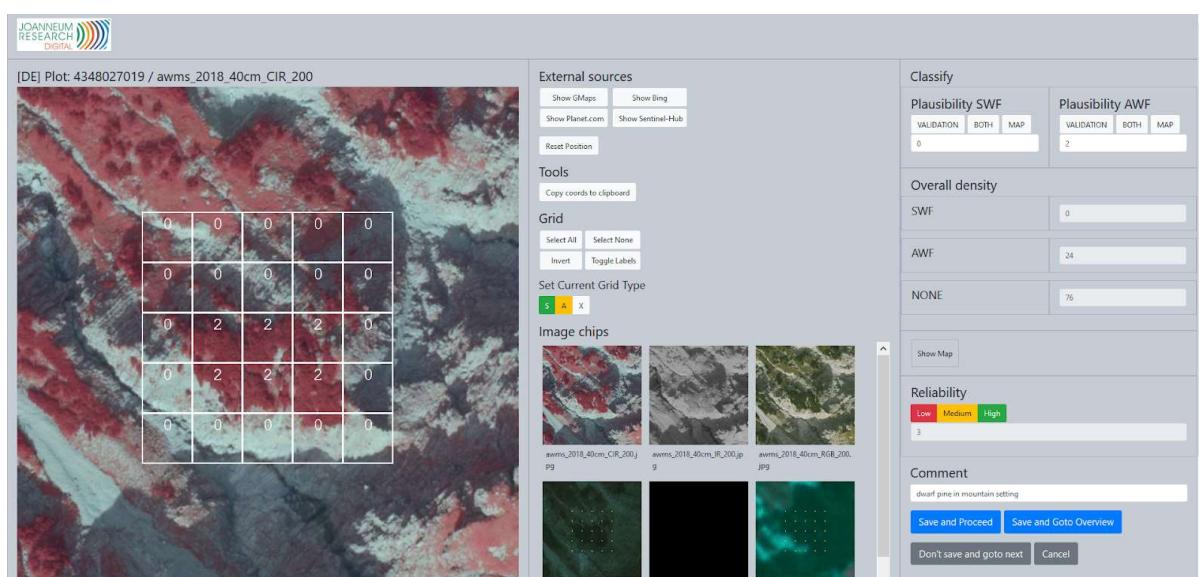


Figure 25: SWF interpretation example: mountain dwarf pines are currently not considered in the product specifications.

Figure 26 shows an example where the plot is within a forest area that is > 5ha, but the forest cover map shows no forest cover (small black image). Since forest areas > 5ha are supposedly removed from the product, this area should be forest and not SWF or AWF. The plot is thus interpreted as having 0% SWF/AWF density during blind interpretation. However, the product specifications of SWF/AWF do not explicitly mention the 5ha threshold, which is why the plausibility assessment also considers the map values to be correct.

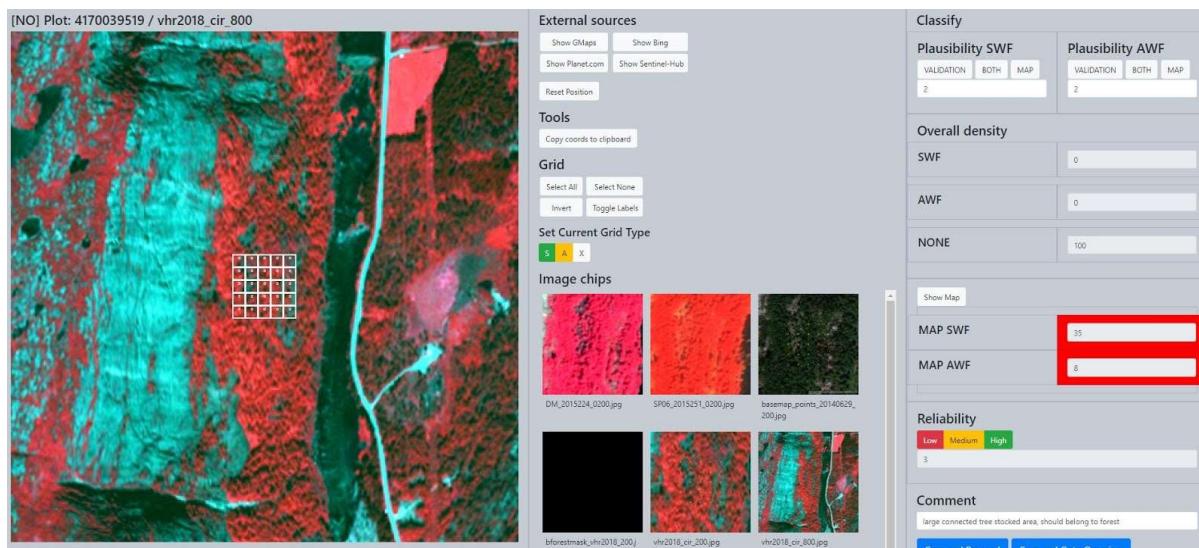


Figure 26: SWF interpretation example: Plot within a forest area > 5ha, but not removed by the forest cover mask.

Some features may look like linear woody features in the reference imagery although they are non-woody vegetation such as reed (Figure 27). Street View data is very helpful for such cases.



Figure 27: SWF interpretation example: A linear vegetation element of reed. It is not woody and thus not considered as SWF / AWF.

Often a linear SWF, but also patchy SWF or even AWF only barely touch the edges of the PSU polygon and only cover 1-3% of the polygon area (Figure 28). Such plots are ambiguous, as a small geometric shift in the VHR 2015 imagery could result a 0% density. During blind interpretation the most likely density value is chosen based on the better reference imagery. If the map value differs at the 1% threshold, during plausibility assessment, also a 0% density is accepted.

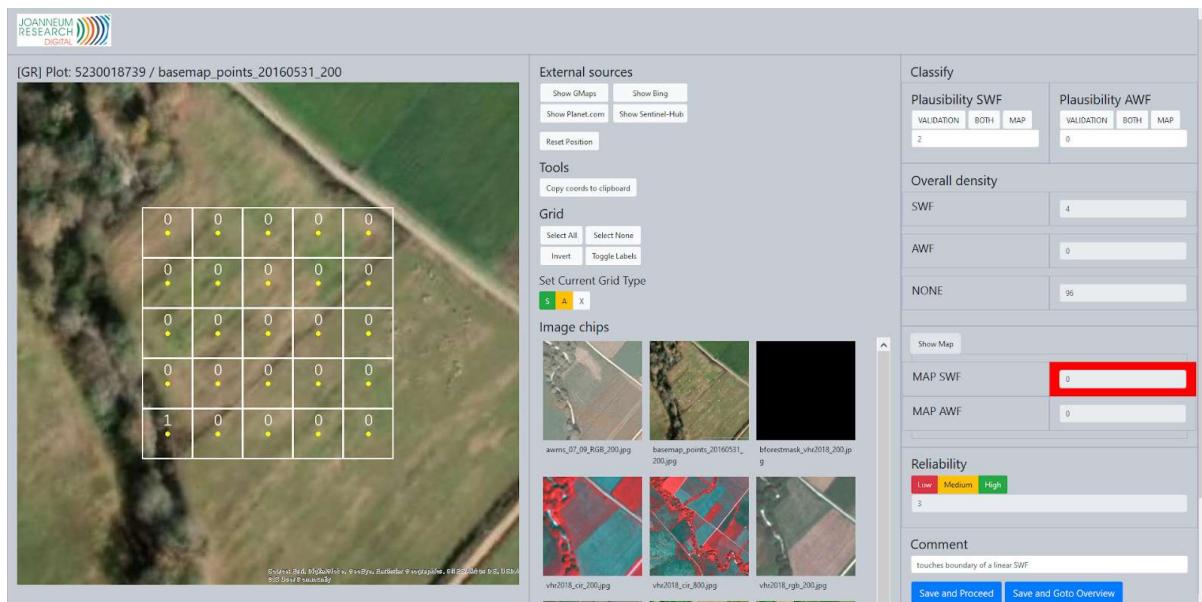


Figure 28: SWF interpretation example: a linear SWF barely touches the PSU grid.

3.4.5. Estimation and analyses procedures

3.4.5.1. Analysis of density values

As described above, density values from the reference data are not directly assessed, but generated from secondary sampling units (SSUs). Therefore, these suffer from sampling error which needs to be considered in the analysis. This makes the use of correlation coefficient difficult to set a suitable threshold above which the correlation is deemed acceptable. If we had a complete information on the cell for our reference data, a reasonable measure of the commission φ and omission ψ errors would be:

$$\varphi = \frac{\sum_i pos(m_i - r_i)}{\sum_i m_i} \quad \psi = \frac{\sum_i pos(r_i - m_i)}{\sum_i r_i} \quad (4)$$

where $pos(x)$ is the positive part, i.e. $pos(x) = x$ if $x > 0$ and $pos(x) = 0$ if $x \leq 0$.

If the map reports a proportion m_i and the reference data give a proportion r_i ,

For each sampling unit of 100m we have a quantitative value in the map (estimated % in the satellite image classification) and a reference value that is an estimation obtained from a sample of 25 points. The number of forest points that we are using as reference value has a probability distribution due to the within-cell sampling. If the within-sampling is random, the number of points follows a binomial $B(25, p)$. In our case the sampling scheme is systematic, but we use anyhow the binomial as an approximation.

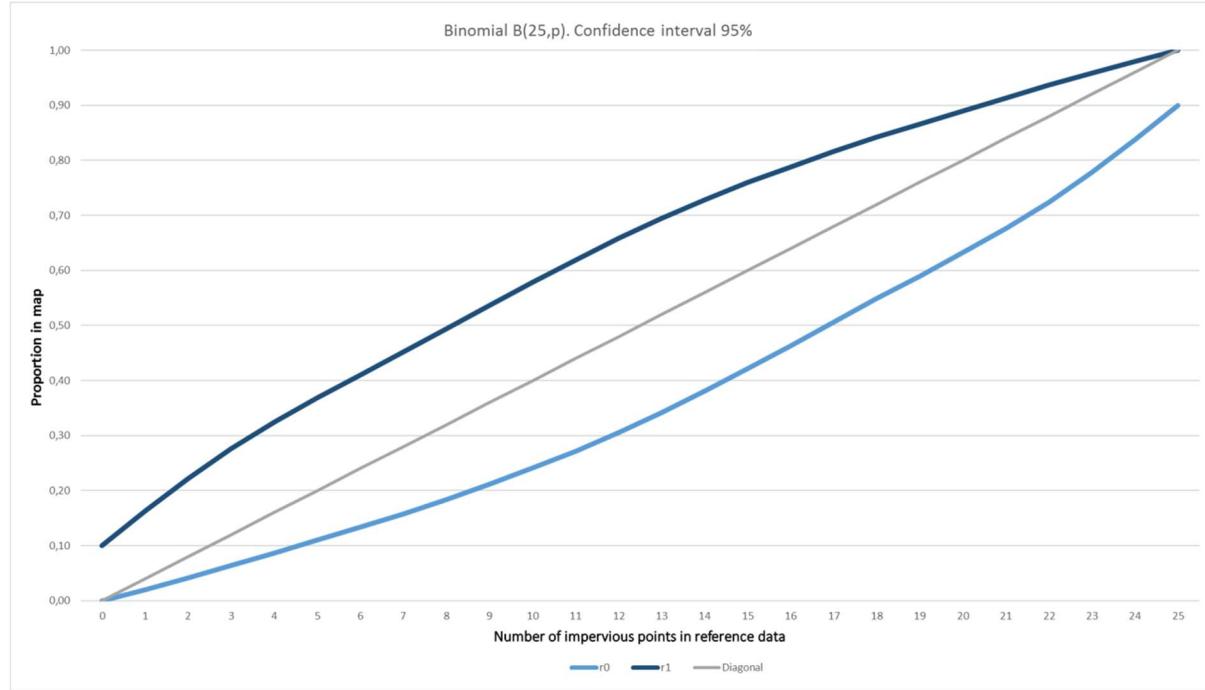


Figure 29: Representation of the behaviour of the 95% confidence interval for a 5x5point SSU grid over the whole range of density values (example from imperviousness analysis).

Therefore, we cannot say that there is any significant disagreement if m_i lays within (r_{0i}, r_{1i}) a confidence interval corresponding to $B(25, r_i)$. Figure 29 represents the behaviour of the 95% confidence interval for $B(25, r_i)$. Notice that only for proportions close to 0.5 we can apply the usual Gaussian approximation that leads to an interval approximately $(r_i \pm 2s_i)$, while for proportions close to 0 or to 1 the intervals are strongly asymmetric.

A possible adaptation of the formulas (4) above for the commission φ and omission ψ would be:

$$\varphi = \frac{\sum_i pos(m_i - r_{1i})}{\sum_i m_i} \quad \psi = \frac{\sum_i pos(r_{0i} - m_i)}{\sum_i r_i} \quad (5)$$

3.4.5.1. Thematic accuracy

The last step consists in analyzing the samples in order to draw conclusions for the product. This will require applying a threshold to the SWF density products in order to convert the continuous density product to a binary mask. For SWF, a 1% threshold is applied which relates to the presence or absence of SWF / AWF features within the PSU.

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.

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 .

3.5. Temporal Quality

Temporal quality is evaluated by providing an indication of the closeness of the acquired image data to the reference year, e.g. the percentage area covered outside the accepted reference period as defined in the tender/product specification i.e. 2015 +/- 1-2 year(s).

3.6. Usability

Usability relates to the appropriateness of the metadata description and accompanying documentation to describe the processes and workflows involved in the production of the data. Although it is difficult to describe usability in quantitative terms, it provides a clear evaluation based on objective criteria of any limitation in the intended use of the data.

3.7. INSPIRE compliant metadata

Presence of INSPIRE compliant metadata should be verified.

4. Validation check list

4.1. Small Woody Features (SWF)

Preliminary validation exercises are usually performed on the final semantic check report for the full pan-European 100m mosaic integrating some of the comments from the semantic check reports on lot deliveries (now QC Tool during upload). We recommend to openly publish the QC Tool check reports together with the map products to offer more details on quality issues to the users. QC report from: 2020-02-25 14:47:54; QC Tool version: 1.5.9; file name: swf_2015_100m_eu_3035_full.zip

| PRODUCT: | | | | Small Woody Features (SWF) @ 100m | | | |
|--|----------------------------|-----------------------|----------------------|--|--|------------------------|------------------------|
| VALIDATION LEVEL: | | | | Pan-European | | | |
| SERVICE PROVIDER: SIRS | | | SERVICE USER: | ISSUE/REVISION:1.0 | | | |
| VALIDATION DATE: 16/07/2020 | | | REVIEW DATE: | | | | |
| CONDUCTED BY: Joanneum Research | | | REVIEWED BY: | APPROVED BY: | | | |
| No. | Data Quality Sub- | Data Quality Measure | Data Quality Result | COMMENTS BY AUDIT TEAM | | Draft Audit Conclusion | Final Audit Conclusion |
| 1 | Completeness | | | | | | |
| 1.1 | Commission | Rate of excess items | OK | | | OK | |
| 1.2 | Omission | Rate of missing items | OK? | Tiles NO116 (Southeast Norway) and NO119 (Northeast Norway) have been discontinued and were not published due to the lack of suitable satellite data | | OK | |
| 2 | Logical Consistency | | | | | | |
| 2.1 | Format consistency | File format/readable | OK | GeoTiff Readable in ArcGIS 10.x and QGIS. OK in QC Tool report | | OK | |

| PRODUCT: | | Small Woody Features (SWF) @ 100m | | | | |
|--|------------------------|---|---------------------|---|--|------------------------|
| VALIDATION LEVEL: | | Pan-European | | | | |
| SERVICE PROVIDER: SIRS | | SERVICE USER: | | ISSUE/REVISION:1.0 | | |
| VALIDATION DATE: 16/07/2020 | | REVIEW DATE: | | | | |
| CONDUCTED BY: Joanneum Research | | REVIEWED BY: | | APPROVED BY: | | |
| No. | Data Quality Sub- | Data Quality Measure | Data Quality Result | Comments by Audit Team | | Draft Audit Conclusion |
| 2.2 | Conceptual consistency | File naming conventions | OK | Corresponds to convention for internal deliveries. OK in QC Tool report | | OK |
| 2.3 | | Attributes naming conventions | n/a | n/a | | n/a |
| 2.4 | | Attributes types | n/a | n/a | | n/a |
| 2.5 | Domaine consistency | Coordinate reference system | OK | European ETRS89 LAEA projection (EPSG 3035), OK in QC Tool report | | OK |
| 2.6 | | Pixel size and origin | OK | Corresponds to 20m LAEA grid, OK in QC Tool report | | OK |
| 2.7 | | Compliance between 5m and 100m product type | OK | OK | | n/a |
| 2.9 | | Additional attributes, symbology | n/a | n/a | | n/a |
| 2.10 | Domaine consistency | Valid Codes | OK | 0: no SWF/AWF density 1-100: SWF/AWF density 254: unclassifiable 255: outside area | | OK |

| PRODUCT: | | Small Woody Features (SWF) @ 100m | | | | | |
|--|-------------------------------|-----------------------------------|---------------------|--|-----|------------------------|------------------------|
| VALIDATION LEVEL: | | Pan-European | | | | | |
| SERVICE PROVIDER: SIRS | | SERVICE USER: | | ISSUE/REVISION:1.0 | | | |
| VALIDATION DATE: 16/07/2020 | | REVIEW DATE: | | | | | |
| CONDUCTED BY: Joanneum Research | | REVIEWED BY: | | APPROVED BY: | | | |
| No. | Data Quality Sub- | Data Quality Measure | Data Quality Result | Comments by Audit Team | | Draft Audit Conclusion | Final Audit Conclusion |
| 3 | POSITIONAL ACCURACY | | | | | | |
| 3.1 | Absolute or external accuracy | RMSEP | n/a | This is dependent on the assessment of the VHR 2015 dataset | n/a | | |
| 3.2 | Relative or internal accuracy | RMSEP | n/a | Positional accuracy of SWF/AWF layers was checked against Google Earth imagery and Bing Maps. No large positional error was found. | n/a | | |
| 4 | THEMATIC ACCURACY | | | | | | |
| 4.1 | Classification correctness | Overall accuracy | OK | See Separate Section 4 below | OK | | |
| | | Min. producer accuracy | OK | See Separate Section 4 below | OK | | |
| 4.2 | | Min. user accuracy | OK | See Separate Section 4 below | OK | | |

| PRODUCT: | | Small Woody Features (SWF) @ 100m | | | | | |
|--|----------------------------|--|---------------------|--|--|------------------------|------------------------|
| VALIDATION LEVEL: | | Pan-European | | | | | |
| SERVICE PROVIDER: SIRS | | SERVICE USER: | | ISSUE/REVISION:1.0 | | | |
| VALIDATION DATE: 16/07/2020 | | REVIEW DATE: | | | | | |
| CONDUCTED BY: Joanneum Research | | REVIEWED BY: | | APPROVED BY: | | | |
| No. | Data Quality Sub- | Data Quality Measure | Data Quality Result | Comments by Audit Team | | Draft Audit Conclusion | Final Audit Conclusion |
| 4.4 | | Discrepancies along borderlines | n/a | n/a | | n/a | |
| 5 | TEMPORAL QUALITY | | | | | | |
| 5.1 | Temporal quality | Closeness of the acquired image data to the reference year | OK | The imagery used spans several years (2015 +/- one year) | | OK | |
| 6 | USABILITY | | | | | | |
| 6.1 | Usability | Usability description | OK | The product is in line with the specification and is based on a clear definition | | OK | |
| 7 | METADATA | | | | | | |
| 7.1 | INSPIRE compliant metadata | Presence | OK | OK in QC Tool report | | OK | |
| 7.2 | | File format | OK | OK in QC Tool report | | OK | |
| 7.3 | | File name | OK | OK in QC Tool report | | OK | |

| PRODUCT: | | Small Woody Features (SWF) @ 100m | | | | |
|--|-------------------|-----------------------------------|---------------------------|------------------------|------------------------|------------------------|
| VALIDATION LEVEL: | | Pan-European | | | | |
| SERVICE PROVIDER: SIRS | | SERVICE USER: | ISSUE/REVISION:1.0 | | | |
| VALIDATION DATE: 16/07/2020 | | REVIEW DATE: | | | | |
| CONDUCTED BY: Joanneum Research | | REVIEWED BY: | APPROVED BY: | | | |
| No. | Data Quality Sub- | Data Quality Measure | Data Quality Result | Comments by Audit Team | Draft Audit Conclusion | Final Audit Conclusion |
| 7.4 | | INSPIRE compliance | OK | OK in QC Tool report | OK | |

5. Thematic accuracy

5.1. Small Woody Features (SWF)

As indicated in the Request for Services for this contract, all SWF density products shall be assessed in line with agreed validation concepts for HRL 2015 Lot 5. For density products, a threshold based thematic validation via a binary mask and an analysis of scatterplots of density values are applied. After the removal of all unreliable plots the remaining 24578 plots are used for accuracy assessment.

Scatterplots are presented for all sample units. Sampling sizes for each stratum are big enough to calculate scatterplots (with exception of the DOMs) even after removal of unreliable plots. However, the distribution of sample units is strongly centred at 0%/0% density, much more than for similar density products such as TCD. More than 70% of the SWF sample plots and more than 85% of the AWF sample plots have density values of 0% in the reference (see Figure 34 and Figure 32). The mean SWF density over all plots is only 2.6%. The mean AWF density over all plots is only 1.92%. Annex 5 lists confusion matrices for SWF and AWF products. For AWF 21,822 of 24,578 valid sample plots (88.8%) have density values of 0% in both map and reference (Table 7.2). Of the remaining 2,756 plots only 1,434 plots (52.0%) have density values $\geq 1\%$ in both map and reference. 1,322 plots (48.0%) only have density values $\geq 1\%$ in either map or reference. These 1322 plots strongly influence the R^2 value shown in the scatterplots.

For SWF 17,525 of 24,578 valid sample plots (71.3%) have density values of 0% in both map and reference (Table 7.4). Of the remaining 7,053 plots only 4,424 plots (62.7%) have density values $\geq 1\%$ in both map and reference. 2,629 plots (37.3%) only have density values $\geq 1\%$ in either map or reference. These 2,629 plots strongly influence the R^2 value in the scatterplots, though not to the same extent as for AWF, as density values are generally lower for SWF and thus the differences between map value and interpreted value are smaller.

SWF values rarely exceed 40%, which is consistent with the geometric definitions and AWF values rarely exceed 60% in the reference (see Figure 31 and Figure 37). A number of AWF plots $> 60\%$ density is however present in the AWF map layer. Often these plots are interpreted as forest during the bulk interpretation (when they belong to a forest area $> 5\text{ha}$). They are thus assigned a density value of 0% in the reference (see confusion Matrix Table 7.2). This leads to sometimes very large differences in density values that strongly affect statistical analysis parameters such as R^2 . This makes the interpretation of regression coefficients, slope and intercept rather difficult for the SWF and AWF products. The scatterplots for all country zones and biogeographic regions are provided in the Annex 1, Annex 2, Annex 3 and Annex 4.

For a quantitative summary of the results displayed in the scatterplots, a linear regression analysis is performed to estimate the relationships between the reference and product information. The analysis produces a coefficient of determination (R^2) which provides information about the goodness of fit of the estimated regression model. Coefficients of determination closer to 1 represent a better fit. It is also useful to consider the slope and intercept of the estimated regression model. The slope should approach 1 and the intercept should be close to 0 for the required relations. Deviations from the expected values give an indication of the overall correspondence of the reference density values and the map values. A summary of the results is given in Table 5.1. Slope and intercept show good to excellent values, while R^2 for reasons already mentioned above is rather low. R^2 values are slightly higher for AWF than for SWF. Whether R^2 is a suitable parameter for estimating the SWF product quality seems questionable due to the distribution of density values. The intercept parameter shows good values close to 0. As 70%/85% (SWF/AWF) of sample plots are located at the 0%/0% position in the scatterplot, intercept values near 0 are to be expected. Some countries and regions show values of slightly over 1, which is a result of commission errors (positive intercept) or omission errors (negative intercept) for the associated bio-geographical region or countries.

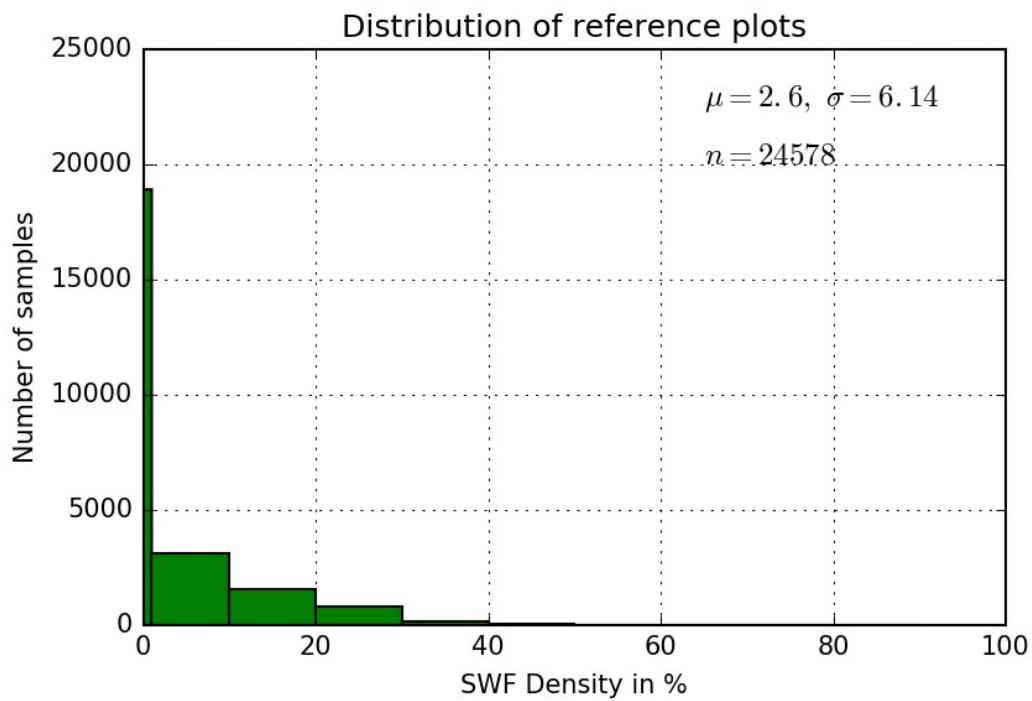


Figure 30: Distribution of sampling units by reference SWF density values.

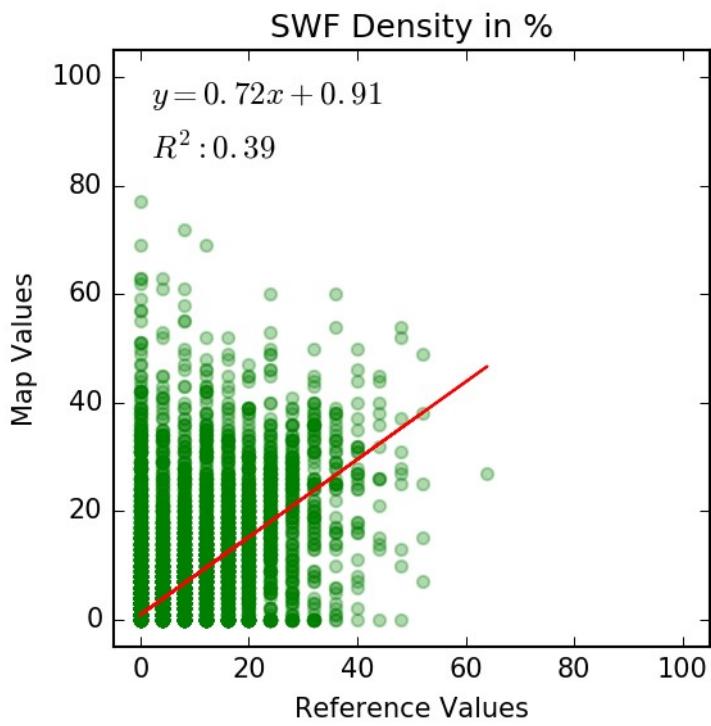


Figure 31: Scatterplot of all SWF sample units darker green indicates a greater number of observations.

Figure 31 shows that both reference and map density values for SWF only rarely exceed 40%. There is a large number of plots with reference values > 0%, but with map values of 0%, which leads to a slope value smaller than 1. The intercept is close to 0, which is to be expected considering that > 70% of plots have map and reference values of 0% density. The R^2 value is 0.39, which shows only medium agreement between reference and map.

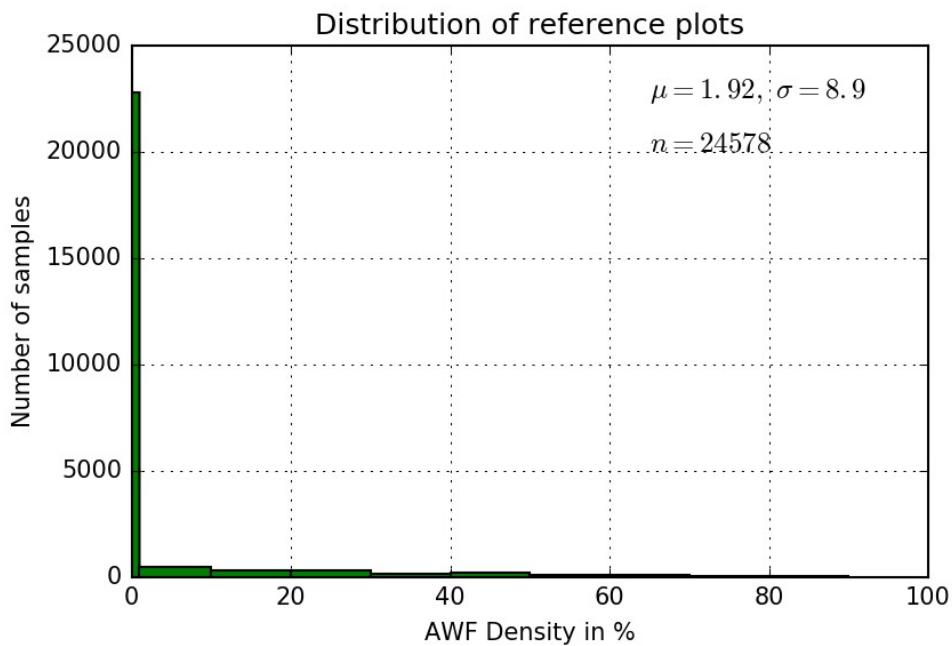


Figure 32: Distribution of sampling units by reference AWF density values.

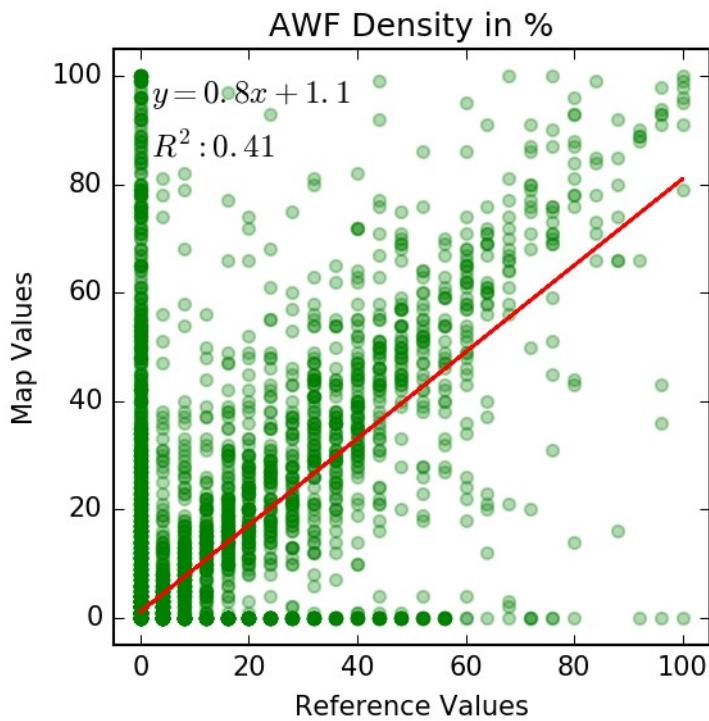


Figure 33: Scatterplot of all AWF sample units darker green indicates a greater number of observations.

Figure 32 shows the distribution of density values for AWF. In Figure 33 it can be seen that there are many plots distributed along an ideal regression line there but there is also quite a lot of confusion between plots that have 0% density in the map and > 0% in the reference and vice versa, so omission and commission errors.

Table 5.1: Details of SWF/AWF regression line parameter and coefficients and for the whole of Europe cover, per lot, bio-geographical regions and country and group of countries greater than 90,000km², n corresponds to the number of observations. Values in green are closest to optimal values (excellent when ≥ 0.8 and okay when ≥ 0.5 for R², close to 1 slope and close to 0 for intercept).

| | n | SWF | | | AWF | | |
|------------------------------|--------|------|-------|-----------|------|-------|-----------|
| | | R2 | Slope | Intercept | R2 | Slope | Intercept |
| EEA39 | 24,578 | 0.39 | 0.72 | 0.91 | 0.41 | 0.8 | 1.1 |
| Biogeographic regions | | | | | | | |
| ALP | 2,067 | 0.28 | 0.76 | 0.99 | 0.34 | 0.76 | 1.02 |
| ANA | 2,867 | 0.23 | 0.84 | 0.67 | 0.11 | 0.71 | 0.88 |
| ARC | 766 | 0.5 | 0.76 | -0.04 | 0.03 | 0.63 | 0.13 |
| ATL | 4,757 | 0.41 | 0.71 | 1.2 | 0.54 | 0.88 | 1.25 |
| BLS | 319 | 0.3 | 0.65 | 1.4 | 0.33 | 0.83 | 1.62 |
| BOR | 1,614 | 0.43 | 0.77 | 0.73 | 0.69 | 0.84 | 0.36 |
| CON | 5,970 | 0.42 | 0.67 | 0.81 | 0.53 | 0.76 | 0.85 |
| MAC | 56 | 0.31 | 0.69 | 0.84 | 0.05 | 0.91 | 0.91 |
| MED | 5,008 | 0.3 | 0.7 | 1.22 | 0.13 | 0.63 | 1.93 |
| PAN | 878 | 0.6 | 0.87 | 0.32 | 0.68 | 0.87 | 0.39 |
| STE | 276 | 0.61 | 0.77 | 0.06 | 0.73 | 1.92 | 0.34 |
| Country zones | | | | | | | |
| AL_MK_ME_RS_KS | 620 | 0.37 | 0.76 | 0.85 | 0.37 | 0.71 | 1.68 |
| AT_CH_LI | 424 | 0.46 | 0.71 | 0.72 | 0.76 | 0.99 | 0.59 |
| BE_NL_LU_DK | 704 | 0.45 | 0.8 | 0.58 | 0.6 | 0.86 | 0.76 |
| BG | 456 | 0.5 | 0.74 | 0.47 | 0.75 | 0.85 | 0.25 |
| CZ_SK | 492 | 0.46 | 0.73 | 0.95 | 0.34 | 0.78 | 0.84 |
| DE | 1,549 | 0.51 | 0.79 | 0.64 | 0.62 | 0.88 | 0.64 |
| EE_LT_LV | 602 | 0.48 | 0.86 | 0.86 | 0.77 | 0.91 | 0.4 |
| ES | 2,249 | 0.31 | 0.67 | 0.97 | 0.17 | 0.73 | 2.43 |
| FI | 475 | 0.59 | 0.78 | 0.06 | 0.71 | 0.88 | 0.23 |
| FR_AD | 2,544 | 0.37 | 0.71 | 1.67 | 0.48 | 0.88 | 1.59 |
| GB_IE | 1,940 | 0.41 | 0.64 | 1.02 | 0.5 | 0.82 | 1.21 |
| GR_CY | 572 | 0.32 | 0.68 | 1.38 | 0.2 | 0.5 | 1.29 |
| HR_BA_SI | 332 | 0.47 | 0.76 | 0.84 | 0.5 | 1.04 | 2.16 |
| HU | 491 | 0.57 | 0.82 | 0.32 | 0.82 | 0.91 | 0.09 |
| IS | 763 | 0.5 | 0.76 | -0.04 | 0.03 | 0.63 | 0.13 |
| IT_MT | 1,102 | 0.27 | 0.63 | 1.62 | 0.38 | 0.8 | 0.76 |
| NO | 1,198 | 0.27 | 0.84 | 1.41 | 0.25 | 0.74 | 1.67 |
| PL | 1,389 | 0.4 | 0.53 | 0.62 | 0.54 | 0.7 | 0.82 |
| PT | 438 | 0.29 | 0.81 | 1.28 | 0.13 | 0.6 | 2.07 |
| RO | 1,147 | 0.45 | 0.8 | 0.47 | 0.47 | 0.6 | 0.46 |
| SE | 929 | 0.44 | 0.77 | 0.74 | 0.59 | 0.74 | 0.39 |
| TR | 4,162 | 0.28 | 0.76 | 0.78 | 0.16 | 0.68 | 0.96 |

Most of the **regression coefficients** do not exceed the 0.5 threshold. Whether R² is a suitable parameter for estimating the SWF product quality seems questionable due to the distribution of density values. Regression coefficients for SWF are higher for regions like the Boreal, Pannonic, Atlantic, Continental and Steppic with a R² above 0.5. At country level, the highest R² are related to central and eastern countries such as Bulgaria, Germany, Hungary and also for Iceland and Finland with a R² greater than 0.5. Concerning the AWF layer the regression coefficients are higher for the Arctic, Pannonic and Steppic biogeographic regions. At country level, the highest R² are related to central and eastern countries such as Hungary, Estonia - Lithuania – Latvia, Austria - Switzerland – Liechtenstein, Bulgaria, Germany, Poland, Belgium - Netherlands - Luxembourg - Denmark and also for Finland and Sweden. The lowest values are related to regions or countries with very low tree cover or complex landscapes (e.g. Arctic bio-geographical region with Iceland, Mediterranean bio-geographical region with Portugal, Spain, Greece or the Anatolian bio-geographical region with Turkey). In general R² values are slightly higher for AWF than for SWF.

Regression **slopes** regarding SWF density values are consistently relatively close to 1 with very few variability in the Arctic bio-geographical region (mainly related to Iceland) and the Alpine region related to Italy – Malta, Austria - Switzerland – Liechtenstein and Sweden. Slope values for AWF density values are also relatively close to 1 with very few variability in the Steppic bio-geographical region (mainly related to Turkey). Better slope values can be observed for the Black Sea (Bulgaria), Boreal (Estonia - Lithuania - Latvia, Finland), Pannonic (Hungary), Atlantic (Germany, France, Belgium - Netherlands - Luxembourg - Denmark) and the Macaronesia bio-geographical Region. All slope values are smaller than 1.0 which is mainly a result of commission errors.

The **intercept** parameter shows good values close to 0. As 70%/85% (SWF/AWF) of sample plots are located at the 0%/0% position in the scatterplot, intercept values near 0 are to be expected. For SWF shows good values close to 0 with some of them slightly over 1 which is a result of the commission errors (positive intercept) or omission errors (negative intercept) for the associated bio-geographical region or countries. While optimal value should be close to 0 with a slope close to 1, some results present an intercept around/above 1.5. The greatest values are related to the Atlantic, Mediterranean and the Black Sea bio-geographical regions well-known to be especially prone to omission and commissions errors due to a very complex landscape. The Continental, Steppic, Boreal, Arctic, Anatolian and Pannonic regions show the better values close to 0. For AWF more values over 1 can be observed compared to SWF, which is a result of the commission errors (positive intercept) for Black Sea region and Mediterranean region (Greece, France, Spain, Albania - Republic of North Macedonia - Montenegro - Serbia – Kosovo, Bosnia and Herzegovina - Croatia - Slovenia). Values close to 0 are related to the Arctic (Iceland), Steppic (Turkey), Boreal (Finland, Sweden), Pannonic (Hungary), Continental (Germany, Poland, Czech Republic - Slovakia), Anatolian (Turkey) and the Macaronesia bio-geographical regions.

The DOM's are excluded in the regression analysis due to the low number of sampling points.

Additionally to the analysis of scatterplots also the thematic validation is also conducted with confusion matrices of threshold based binary masks. Therefore, SWF/ AWF /SWF+AWF values were regrouped based on the 1% threshold (presence or absence of small or additional woody features in the 100m x 100m PSU) to perform a thematic accuracy assessment:

- SWF values >= 1%
- AWF values >= 1%
- SWF+AWF values >= 1%

Table 5.2 shows the results at European level and include Overall Accuracy (OA), Producer Accuracy (PA) and User Accuracy (UA). Accuracy figures were calculated following the procedure described in 3.4.4 taking the weights of

Table 3.5 into account. At European level, for both blind and plausibility interpretation, the overall accuracy meets the target accuracy requirements (80%). Concerning the SWF layer the User and Producer accuracy do not meet the defined target accuracy of 80% in the blind analysis, but it in the plausibility analysis. For the combined product, the suggested 80% target accuracy would be met for all accuracy measures. Nevertheless, the Producer Accuracy for the AWF layer would not meet the suggested 80% target accuracy which indicated that the product slightly overestimates the AWF density values. This might be related to the intermediate forest cover mask, which is based on the HRL TCD 2015 product, since missing forest patches are classified as AWF.

Table 5.2: SWF / AWF / SWF+AWF layers thematic accuracy results at the 1% threshold for the pan-European area + DOMs. Values in green indicate accuracies greater than the 80% threshold (blind/plaus), yellow greater than 80% threshold when the 95% CI is considered, red when less than 80% even when considering the 95% CI (Prod. = Producer).

| | | BLIND | | | | | | PLAUS | | | | | | No. of plots |
|-------|---------|-------|----------|-------|----------|-------|----------|-------|----------|-------|----------|-------|----------|--------------|
| | | OA | Conf-95% | PA | Conf-95% | UA | Conf-95% | OA | Conf-95% | PA | Conf-95% | UA | Conf-95% | |
| EEA39 | SWF | 89.3 | 0.38 | 77.97 | 0.95 | 76.24 | 1.1 | 94.13 | 0.29 | 88.59 | 0.77 | 86.27 | 0.89 | 24578 |
| | AWF | 94.62 | 0.24 | 80.16 | 1.75 | 59.73 | 1.96 | 96.48 | 0.21 | 89.59 | 1.31 | 72.43 | 1.79 | |
| | SWF+AWF | 89.82 | 0.37 | 81.88 | 0.84 | 79.94 | 0.96 | 94.32 | 0.29 | 90.68 | 0.67 | 88.01 | 0.78 | |

The width of the 95% confidence intervals depends on (i) the number of sample units selected, (ii) the accuracy reached and (iii) the efficiency of the stratification. The stratification appears to be very efficient with 95% CI less than 1% achieved at pan-European for assessing commission and omission errors. Indeed, based on the recommendations made in the delivery report for the 2012 Forest products¹, the stratification was simplified to reach this efficiency. There is more variability at bio-geographical region and country/group of countries level.

The results for biogeographical regions are shown in Table 5.3. Here the picture is more diverse. While all country zones and regions fulfil the overall accuracy for both the blind and plausibility analysis, the User and Producer threshold of 80% are not always met. In terms of User Accuracies, especially Anatolia, the Alpine region and the Mediterranean region show low accuracy values. The reason could be that these areas are characterized by Mediterranean shrubs (Macchia, Garrigue) and alpine shrubs and that the forest cover mask is less accurate in these regions. The HRL Forest validation report² also reported lowest TCD accuracies for the Anatolian and Mediterranean region. This could simply be a propagation of that error. Producer Accuracy values are low for the Arctic region, but here only very few sample plots with density values > 0% are available and confidence intervals are thus huge. In summary, the best results are obtained for the Steppic and Pannonic regions.

The results for countries/groups of countries are shown in Table 5.4. The target overall accuracy is met for all countries and groups of countries (80%). In addition, the Producer Accuracy is met for almost all countries and country groups. However, some countries do not reach the target User Accuracy requirements (80%). These countries are: Spain, Greece and Cyprus, Portugal, Italy, Turkey and Norway. Most of these countries are part of the Mediterranean and Anatolian region where also the TCD 2015 product showed deficiencies. This could be the same issue described above for the biogeographic regions. In Norway the low density forests and possibly image data quality issues during production (terrain shadows) could be possible reasons for the lower accuracies. The DOMs show good accuracies, but confidence intervals are huge, due to the small sample size.

Table 5.3: SWF / AWF / SWF+AWF product thematic accuracy results for the bio-geographical regions. Values in green indicate accuracies $\geq 80\%$. Values in yellow indicate accuracies $\geq 80\%$ when adding the confidence values.

| Biogeographic regions | | BLIND | | | | | | PLAUS | | | | | | No. of plots |
|-----------------------|---------|-------|----------|-------|----------|-------|----------|-------|----------|-------|----------|-------|----------|--------------|
| | | OA | Conf-95% | PA | Conf-95% | UA | Conf-95% | OA | Conf-95% | PA | Conf-95% | UA | Conf-95% | |
| ALP | SWF | 91.58 | 1.08 | 77.69 | 4.67 | 63.52 | 5.3 | 95.5 | 0.83 | 92.13 | 3.12 | 77.36 | 4.61 | 2067 |
| | AWF | 95.89 | 0.72 | 80 | 7.31 | 56.76 | 8.01 | 97.63 | 0.59 | 91.6 | 4.83 | 73.65 | 7.12 | |
| | SWF+AWF | 91.29 | 1.11 | 81.06 | 4.11 | 66.49 | 4.84 | 95.36 | 0.85 | 93.29 | 2.68 | 79.56 | 4.13 | |

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

² see <https://land.copernicus.eu/user-corner/technical-library/hrl-forest-2015-final-validation-report>

| | | | | | | | | | | | | | | |
|-----|---------|-------|------|-------|-------|-------|-------|-------|------|-------|-------|-------|-------|------|
| ANA | SWF | 94.11 | 0.7 | 80.66 | 5.58 | 52.14 | 5.86 | 95.81 | 0.64 | 88.83 | 4.15 | 65.36 | 5.58 | 2867 |
| | AWF | 96.76 | 0.3 | 91.3 | 12.91 | 18.75 | 7.26 | 97 | 0.32 | 96.43 | 7.12 | 24.11 | 7.96 | |
| | SWF+AWF | 94.18 | 0.69 | 88.6 | 4.4 | 54.11 | 5.5 | 95.78 | 0.62 | 94.93 | 2.88 | 65.19 | 5.26 | |
| ARC | SWF | 98.83 | 0.76 | 30.77 | 13.83 | 100 | 0 | 99.48 | 0.51 | 50 | 24.44 | 100 | 0 | 766 |
| | AWF | 99.61 | 0.44 | 50 | 38.7 | 66.67 | 65.33 | 99.74 | 0.36 | 60 | 33.22 | 100 | 0 | |
| | SWF+AWF | 98.83 | 0.76 | 38.46 | 17.3 | 83.33 | 32.67 | 99.48 | 0.51 | 60 | 23.46 | 100 | 0 | |
| ATL | SWF | 86.15 | 0.98 | 79.51 | 1.59 | 83.25 | 1.76 | 92.83 | 0.73 | 89.27 | 1.33 | 91.19 | 1.34 | 4757 |
| | AWF | 93.65 | 0.63 | 85.35 | 2.79 | 68.11 | 3.47 | 96.24 | 0.51 | 92.55 | 2.02 | 80.66 | 2.94 | |
| | SWF+AWF | 87.32 | 0.95 | 83.27 | 1.41 | 86.55 | 1.51 | 93.32 | 0.71 | 91.15 | 1.16 | 92.74 | 1.15 | |
| BLS | SWF | 87.77 | 3.48 | 77.92 | 8.22 | 73.17 | 9.65 | 92.48 | 2.87 | 85.36 | 6.98 | 85.37 | 7.7 | 319 |
| | AWF | 93.1 | 2.03 | 81.25 | 19.38 | 40.62 | 17.29 | 93.73 | 2.06 | 83.33 | 17.06 | 46.87 | 17.57 | |
| | SWF+AWF | 88.4 | 3.45 | 80 | 7.55 | 77.27 | 8.81 | 91.85 | 2.97 | 86.05 | 6.71 | 84.09 | 7.69 | |
| BOR | SWF | 91.7 | 1.32 | 82.29 | 3.52 | 81.4 | 3.96 | 96.65 | 0.87 | 92.49 | 2.55 | 92.99 | 2.6 | 1614 |
| | AWF | 94.92 | 1.04 | 76.32 | 5.36 | 79.67 | 5.86 | 98.39 | 0.61 | 91.49 | 3.8 | 94.51 | 3.32 | |
| | SWF+AWF | 93.87 | 1.17 | 86.27 | 2.86 | 91.67 | 2.61 | 97.4 | 0.78 | 93.53 | 2.18 | 96.99 | 1.61 | |
| CON | SWF | 88.02 | 0.82 | 76.5 | 1.68 | 82.17 | 1.84 | 93.43 | 0.63 | 86.48 | 1.46 | 90.54 | 1.41 | 5970 |
| | AWF | 94.02 | 0.54 | 79.52 | 3.03 | 66.29 | 3.5 | 96.4 | 0.45 | 88.85 | 2.34 | 79.37 | 2.99 | |
| | SWF+AWF | 89.23 | 0.79 | 80.97 | 1.47 | 86.92 | 1.51 | 94.34 | 0.59 | 89.43 | 1.24 | 93.43 | 1.11 | |
| MAC | SWF | 87.5 | 8.82 | 66.66 | 19.02 | 83.33 | 22.02 | 96.43 | 4.95 | 91.67 | 14.86 | 91.67 | 16.33 | 56 |
| | AWF | 94.64 | 4.95 | 50 | 68.97 | 33.33 | 65.33 | 94.64 | 4.95 | 50 | 68.97 | 33.33 | 65.33 | |
| | SWF+AWF | 87.5 | 8.82 | 66.66 | 19.02 | 83.33 | 22.02 | 96.43 | 4.95 | 91.67 | 14.86 | 91.67 | 16.33 | |
| MED | SWF | 87.12 | 0.86 | 77.04 | 2.36 | 65.04 | 2.71 | 92.63 | 0.69 | 89.63 | 1.76 | 77.93 | 2.36 | 5008 |
| | AWF | 93.17 | 0.52 | 71.49 | 5.76 | 38.79 | 4.53 | 94.75 | 0.49 | 82.79 | 4.32 | 51.79 | 4.64 | |
| | SWF+AWF | 86.22 | 0.89 | 78.95 | 2.18 | 65.66 | 2.55 | 91.79 | 0.72 | 90.07 | 1.64 | 77.69 | 2.24 | |
| PAN | SWF | 91.8 | 1.76 | 75.16 | 6 | 77.18 | 6.76 | 95.33 | 1.38 | 85.06 | 5.16 | 87.92 | 5.25 | 878 |
| | AWF | 96.7 | 1.1 | 82.81 | 8.6 | 74.65 | 10.19 | 97.95 | 0.9 | 90.77 | 6.74 | 83.1 | 8.78 | |
| | SWF+AWF | 92.82 | 1.68 | 82.63 | 4.86 | 83.96 | 5.27 | 95.9 | 1.3 | 89.95 | 4.03 | 90.91 | 4.13 | |
| STE | SWF | 97.83 | 1.71 | 76.47 | 17.82 | 86.67 | 17.81 | 99.28 | 1 | 88.23 | 14.33 | 100 | 0 | 276 |
| | AWF | 98.55 | 1.07 | 100 | 0 | 50 | 37.04 | 98.91 | 1.04 | 100 | 0 | 62.5 | 35.86 | |
| | SWF+AWF | 98.19 | 1.56 | 83.33 | 15.8 | 88.24 | 15.79 | 99.64 | 0.71 | 94.44 | 10.27 | 100 | 0 | |

Table 5.4: SWF / AWF / SWF+AWF product thematic accuracy results for countries and group of countries greater than 90,000km². Values in green indicate accuracies ≥ 80%. Values in yellow indicate accuracies ≥ 80% when adding the confidence values.

| | | BLIND | | | | | | PLAUS | | | | | | No. of plots |
|----------------|---------|-------|----------|-------|----------|-------|----------|-------|----------|-------|----------|-------|----------|--------------|
| | | OA | Conf-95% | PA | Conf-95% | UA | Conf-95% | OA | Conf-95% | PA | Conf-95% | UA | Conf-95% | |
| Country Zones | | | | | | | | | | | | | | |
| AL_MK_ME_RS_KS | SWF | 86.77 | 2.62 | 74.07 | 5.79 | 75 | 6.73 | 95.97 | 1.55 | 90.42 | 4.18 | 94.38 | 3.58 | 620 |
| | AWF | 93.06 | 1.66 | 87.27 | 8.56 | 57.14 | 10.65 | 97.1 | 1.24 | 95.83 | 4.53 | 82.14 | 8.24 | |
| | SWF+AWF | 87.58 | 2.56 | 80.43 | 5.07 | 78.31 | 5.89 | 96.45 | 1.46 | 93.26 | 3.37 | 95.24 | 3.04 | |
| AT_CH_LI | SWF | 91.04 | 2.64 | 82.61 | 7.04 | 77.55 | 8.3 | 94.1 | 2.18 | 90.11 | 5.8 | 83.67 | 7.36 | 424 |
| | AWF | 95.76 | 1.63 | 88.89 | 11.52 | 61.54 | 15.47 | 96.7 | 1.54 | 90.32 | 10.04 | 71.79 | 14.31 | |
| | SWF+AWF | 91.04 | 2.64 | 85.44 | 6.27 | 79.28 | 7.57 | 93.87 | 2.22 | 92.08 | 5.03 | 83.78 | 6.89 | |
| BE_NL_LU_DK | SWF | 86.93 | 2.49 | 77.42 | 4.4 | 84.21 | 4.74 | 93.61 | 1.81 | 88.61 | 3.72 | 92.11 | 3.51 | 704 |
| | AWF | 95.17 | 1.53 | 82.02 | 7.29 | 80.22 | 8.23 | 98.58 | 0.85 | 97.65 | 3.18 | 91.21 | 5.85 | |
| | SWF+AWF | 88.35 | 2.37 | 81.27 | 3.9 | 88.8 | 3.85 | 94.03 | 1.75 | 90.04 | 3.29 | 94.21 | 2.85 | |
| BG | SWF | 92.11 | 2.45 | 75.55 | 7.69 | 82.93 | 8.19 | 96.27 | 1.73 | 88.23 | 6.38 | 91.46 | 6.09 | 456 |
| | AWF | 96.49 | 1.59 | 77.42 | 13.39 | 72.73 | 15.43 | 98.9 | 0.94 | 93.75 | 8.13 | 90.91 | 9.96 | |
| | SWF+AWF | 91.67 | 2.53 | 76.19 | 7 | 86.02 | 7.09 | 96.27 | 1.74 | 88 | 5.9 | 94.62 | 4.61 | |
| CZ_SK | SWF | 90.04 | 2.63 | 79.85 | 5.84 | 84.09 | 6.26 | 93.5 | 2.18 | 86.23 | 5.24 | 90.15 | 5.1 | 492 |
| | AWF | 95.94 | 1.6 | 75 | 13.67 | 66.67 | 15.62 | 96.54 | 1.45 | 82.76 | 12.97 | 66.67 | 15.62 | |
| | SWF+AWF | 90.45 | 2.6 | 81.93 | 5.33 | 86.99 | 5.48 | 93.9 | 2.12 | 88.16 | 4.72 | 91.78 | 4.47 | |
| DE | SWF | 89.61 | 1.52 | 80.85 | 2.95 | 87.08 | 3 | 93.54 | 1.22 | 86.68 | 2.65 | 93.54 | 2.2 | 1549 |
| | AWF | 95.03 | 0.98 | 85.53 | 5.29 | 70.27 | 6.6 | 97.16 | 0.79 | 91.72 | 3.99 | 83.78 | 5.33 | |
| | SWF+AWF | 90.57 | 1.46 | 83.67 | 2.61 | 91.03 | 2.4 | 94.64 | 1.12 | 89.17 | 2.29 | 96.52 | 1.54 | |
| EE_LT_LV | SWF | 91.03 | 2.2 | 87.5 | 5.04 | 77.78 | 6.42 | 97.34 | 1.27 | 97.4 | 2.47 | 92.59 | 4.05 | 602 |

| | | | | | | | | | | | | | | |
|--------------|----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|
| | AWF | 95.02 | 1.65 | 82.35 | 8.36 | 75.68 | 9.84 | 98.17 | 1.06 | 93.15 | 5.58 | 91.89 | 6.26 | |
| | SWF+AWF | 93.02 | 2.01 | 90.28 | 4.12 | 86.34 | 4.99 | 97.51 | 1.24 | 97.19 | 2.38 | 94.54 | 3.3 | |
| ES | SWF | 87.51 | 1.24 | 76.1 | 3.9 | 60.79 | 4.36 | 93.11 | 0.97 | 91.18 | 2.68 | 75.1 | 3.86 | |
| | AWF | 92.84 | 0.75 | 77.14 | 8.27 | 37.16 | 6.43 | 94 | 0.74 | 85.47 | 6.37 | 45.87 | 6.63 | 2249 |
| | SWF+AWF | 86.31 | 1.29 | 79.86 | 3.48 | 61.34 | 4 | 91.91 | 1.04 | 92.53 | 2.33 | 73.99 | 3.61 | |
| | | | | | | | | | | | | | | |
| FI | SWF | 94.95 | 1.96 | 75.36 | 8.79 | 88.14 | 8.32 | 97.26 | 1.47 | 84.85 | 7.91 | 94.92 | 5.65 | |
| | AWF | 97.26 | 1.45 | 82.22 | 10.16 | 88.1 | 9.91 | 98.95 | 0.91 | 95.12 | 6.43 | 92.86 | 7.88 | 475 |
| | SWF+AWF | 96 | 1.77 | 82.98 | 6.8 | 96.3 | 4.14 | 97.89 | 1.29 | 89.89 | 5.87 | 98.77 | 2.42 | |
| FR_AD | SWF | 85.22 | 1.37 | 81.94 | 2.08 | 80.65 | 2.43 | 91.63 | 1.07 | 90.49 | 1.7 | 88.25 | 1.98 | |
| | AWF | 93 | 0.88 | 85.96 | 3.8 | 64.69 | 4.76 | 95.13 | 0.78 | 90.49 | 3.05 | 76.03 | 4.25 | 2544 |
| | SWF+AWF | 86.44 | 1.33 | 85.07 | 1.84 | 84.09 | 2.14 | 92.3 | 1.03 | 92.42 | 1.47 | 89.96 | 1.76 | |
| GB_IE | SWF | 85.26 | 1.58 | 77.07 | 2.47 | 84.52 | 2.67 | 92.84 | 1.15 | 88.43 | 2.11 | 92.33 | 1.97 | |
| | AWF | 93.45 | 1 | 82.19 | 4.74 | 67.16 | 5.63 | 96.13 | 0.81 | 90.38 | 3.57 | 80.6 | 4.74 | 1940 |
| | SWF+AWF | 86.55 | 1.52 | 80.53 | 2.22 | 88.2 | 2.25 | 93.25 | 1.12 | 89.72 | 1.89 | 94.16 | 1.64 | |
| GR_CY | SWF | 84.97 | 2.8 | 74.1 | 6.39 | 67.32 | 7.46 | 91.43 | 2.23 | 86.62 | 5.18 | 80.39 | 6.31 | |
| | AWF | 93.01 | 1.68 | 59.37 | 16.34 | 41.3 | 14.39 | 95.28 | 1.51 | 77.14 | 13.08 | 58.7 | 14.39 | 572 |
| | SWF+AWF | 84.62 | 2.88 | 73.72 | 5.95 | 70.99 | 7.01 | 91.78 | 2.22 | 86.16 | 4.91 | 84.57 | 5.58 | |
| HR_BA_SI | SWF | 87.05 | 3.57 | 79.41 | 6.87 | 78.64 | 7.95 | 93.98 | 2.56 | 89.52 | 5.44 | 91.26 | 5.48 | |
| | AWF | 90.66 | 2.7 | 94.12 | 6.34 | 63.16 | 10.92 | 94.88 | 2.22 | 96.83 | 4.26 | 80.26 | 9.01 | 332 |
| | SWF+AWF | 87.65 | 3.48 | 89.52 | 5 | 79.86 | 6.69 | 93.67 | 2.6 | 94.7 | 3.66 | 89.93 | 5.02 | |
| HU | SWF | 91.04 | 2.45 | 72.62 | 8.28 | 74.39 | 9.51 | 94.5 | 2 | 81.61 | 7.32 | 86.59 | 7.42 | |
| | AWF | 96.74 | 1.51 | 83.72 | 10.22 | 80 | 11.82 | 97.76 | 1.28 | 88.64 | 8.86 | 86.67 | 10.04 | 491 |
| | SWF+AWF | 92.46 | 2.31 | 82.14 | 6.37 | 84.4 | 6.84 | 95.72 | 1.79 | 88.6 | 5.42 | 92.66 | 4.92 | |
| IS | SWF | 98.82 | 0.77 | 30.77 | 13.83 | 100 | 0 | 99.48 | 0.51 | 50 | 24.44 | 100 | 0 | |
| | AWF | 99.61 | 0.44 | 50 | 38.7 | 66.67 | 65.33 | 99.74 | 0.36 | 60 | 33.22 | 100 | 0 | 763 |
| | SWF+AWF | 98.82 | 0.77 | 38.46 | 17.3 | 83.33 | 32.67 | 99.48 | 0.51 | 60 | 23.46 | 100 | 0 | |
| IT_MT | SWF | 86.3 | 1.96 | 78.69 | 4.18 | 72.01 | 4.94 | 91.47 | 1.62 | 86.84 | 3.5 | 83.02 | 4.13 | |
| | AWF | 95.19 | 0.97 | 78.26 | 11.79 | 45.57 | 11.05 | 96.37 | 0.93 | 85.45 | 9 | 59.49 | 10.89 | 1102 |
| | SWF+AWF | 86.12 | 1.97 | 80.45 | 3.97 | 72.65 | 4.75 | 91.38 | 1.63 | 87.92 | 3.29 | 83.53 | 3.95 | |
| NO | SWF | 90.23 | 1.51 | 76.25 | 6.08 | 60.7 | 6.77 | 94.91 | 1.13 | 94.3 | 3.53 | 74.13 | 6.07 | |
| | AWF | 93.99 | 1.13 | 70.67 | 9.71 | 51.46 | 9.7 | 97.08 | 0.85 | 92.5 | 5.62 | 71.84 | 8.73 | 1198 |
| | SWF+AWF | 90.07 | 1.55 | 80.1 | 5.23 | 65.38 | 6.11 | 95.16 | 1.13 | 95.83 | 2.77 | 78.63 | 5.26 | |
| PL | SWF | 87.55 | 1.72 | 69.03 | 3.97 | 79.15 | 4.55 | 93.59 | 1.29 | 81.68 | 3.62 | 91.53 | 3.12 | |
| | AWF | 92.15 | 1.25 | 76.35 | 6.35 | 60.43 | 7.03 | 95.25 | 1.06 | 85.8 | 4.93 | 77.54 | 6 | 1389 |
| | SWF+AWF | 90.21 | 1.56 | 79.72 | 3.27 | 87.82 | 3.23 | 95.25 | 1.12 | 88.32 | 2.79 | 95.94 | 1.95 | |
| PT | SWF | 87.44 | 3.02 | 73.27 | 7.46 | 72.55 | 8.7 | 93.38 | 2.24 | 90.11 | 5.82 | 80.39 | 7.74 | |
| | AWF | 91.1 | 1.91 | 70.83 | 18.94 | 34.69 | 13.47 | 93.84 | 1.81 | 86.67 | 11.92 | 53.06 | 14.12 | 438 |
| | SWF+AWF | 84.93 | 3.21 | 75.67 | 7.04 | 68.29 | 8.26 | 91.32 | 2.51 | 90.48 | 5.33 | 77.24 | 7.44 | |
| RO | SWF | 92.68 | 1.49 | 78.51 | 4.7 | 83.64 | 4.97 | 96.95 | 0.99 | 90.87 | 3.61 | 92.99 | 3.43 | |
| | AWF | 96.34 | 1.04 | 70.51 | 8.83 | 74.32 | 10.02 | 97.91 | 0.79 | 85.71 | 7.68 | 81.08 | 8.98 | 1147 |
| | SWF+AWF | 92.85 | 1.48 | 80.31 | 4.33 | 86.44 | 4.38 | 96.69 | 1.03 | 91.6 | 3.35 | 92.37 | 3.39 | |
| SE | SWF | 92.57 | 1.64 | 80.7 | 5.33 | 79.31 | 6.04 | 96.34 | 1.19 | 90.7 | 4.11 | 89.66 | 4.54 | |
| | AWF | 95.05 | 1.35 | 69.32 | 8.27 | 76.25 | 9.38 | 98.39 | 0.81 | 88.24 | 6.42 | 93.75 | 5.34 | 929 |
| | SWF+AWF | 94.3 | 1.49 | 83.96 | 4.46 | 90.36 | 4.13 | 97.31 | 1.04 | 91.75 | 3.56 | 95.94 | 2.76 | |
| TR | SWF | 92.34 | 0.71 | 77.03 | 3.77 | 59.08 | 4.13 | 94.71 | 0.62 | 86.52 | 3.01 | 70.64 | 3.83 | |
| | AWF | 96.3 | 0.38 | 77.65 | 9.41 | 32.84 | 6.51 | 96.59 | 0.37 | 82.42 | 8.07 | 37.31 | 6.7 | 4162 |
| | SWF+AWF | 92.41 | 0.71 | 81.28 | 3.38 | 61.5 | 3.9 | 94.52 | 0.62 | 89.07 | 2.69 | 70.67 | 3.65 | |
| Dom's | | | | | | | | | | | | | | |
| GF | SWF | 95.04 | 4.74 | 100 | 0 | 62.5 | 35.86 | 96.69 | 4.24 | 100 | 0 | 75 | 32.08 | |
| | AWF | 96.69 | 3.97 | 100 | 0 | 60 | 48.01 | 98.35 | 3.24 | 100 | 0 | 80 | 39.2 | 19 |
| | SWF+AWF | 96.69 | 4.29 | 100 | 0 | 77.78 | 28.81 | 98.35 | 3.24 | 100 | 0 | 88.89 | 21.78 | |
| GP | SWF | 84.29 | 17.1 | 90.32 | 16.91 | 77.78 | 28.81 | 95.71 | 8.54 | 92.31 | 13.49 | 100 | 0 | |
| | AWF | 90 | 14.26 | 86.96 | 22.85 | 83.33 | 32.67 | 95.71 | 8.82 | 88.89 | 19.58 | 100 | 0 | 20 |
| | SWF+AWF | 84.29 | 17.13 | 91.43 | 14.79 | 80 | 26.13 | 95.71 | 8.4 | 93.02 | 12.06 | 100 | 0 | |
| MQ | SWF | 75.85 | 19.92 | 76.68 | 18.66 | 77.78 | 28.81 | 85.85 | 16 | 84.93 | 16.64 | 88.89 | 21.78 | |
| | AWF | 88.3 | 15.28 | 83.33 | 23.64 | 83.33 | 32.67 | 94.15 | 10.1 | 85.71 | 20.38 | 100 | 0 | 20 |
| | SWF+AWF | 81.7 | 16.91 | 80.87 | 14.93 | 90 | 19.6 | 85.85 | 15.78 | 86.38 | 14.77 | 90 | 19.6 | |
| RE | SWF | 81.24 | 17.12 | 84.5 | 25.07 | 70 | 29.94 | 90 | 13.97 | 87.52 | 20.44 | 90 | 19.6 | 20 |

| | | | | | | | | | | | | | | |
|----|----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----|
| | AWF | 91.24 | 12.97 | 75 | 26.75 | 100 | 0 | 95.62 | 9.47 | 85.71 | 25.53 | 100 | 0 | |
| | SWF+AWF | 90 | 13.97 | 87.52 | 20.44 | 90 | 19.6 | 94.38 | 11.02 | 88.62 | 18.75 | 100 | 0 | |
| YT | SWF | 84.77 | 16.5 | 88.47 | 19.53 | 77.78 | 28.81 | 94.77 | 10.25 | 100 | 0 | 88.89 | 21.78 | |
| | AWF | 100 | 0 | 100 | 0 | 100 | 0 | 100 | 0 | 100 | 0 | 100 | 0 | 20 |
| | SWF+AWF | 84.77 | 16.56 | 89.76 | 17.35 | 80 | 26.13 | 94.77 | 10.25 | 100 | 0 | 90 | 19.6 | |
| | | | | | | | | | | | | | | |

6. Conclusions and recommendations

The thematic accuracy of the **HRL SWF 100m product (SWF, AWF and combined SWF+AWF Layer)** is determined using a binary product that is based on a 1% density threshold - i.e. presence or absence of small or additional woody features within a 1ha reference polygon (Primary Sampling Unit). In addition, density scatterplots at pan-European level and for biogeographic regions are provided to visualize the density distribution and to estimate linear regression coefficients. Regression coefficients better show the overall agreement between interpretation and map values, while accuracy results for the chosen binary mask show the agreement on the presence or absence of small woody features for the PSUs.

Regarding the thematic accuracy assessment, results show that the HRL 2015 SWF product layers (SWF, AWF and SWF+AWF) meet most thematic classification target accuracy requirements (80%) based on a 1% threshold of SWF, AWF and SWF+AWF density values. Nevertheless, there is variability from country to country or based on biogeographical regions and some country zones and regions do not fulfil the target accuracies.

The main findings and recommendations for the HRL SWF / AWF / SWF+AWF products can be summarised as follows:

1. The SWF layer exceeds the overall target accuracy at pan-European level at the 1% density threshold for both blind and plausibility interpretation. The target Producer and User Accuracies are met after plausibility checks.
2. The AWF layer exceeds the overall target accuracy at pan-European level at the 1% density threshold for both blind and plausibility interpretation. The target Producer Accuracy is neither met for blind nor plausibility interpretation (59.7% and 72.43%) suggesting a high commission error. This might be related to the intermediate forest cover mask, which is based on the HRL TCD 2015 product. It is noticeable that accuracy issues for AWF (and partly SWF) occur in the same biogeographic regions and country zones than for the TCD 2015 product.
3. The combined SWF+AWF layer exceeds the target accuracy at pan-European level for overall accuracy and both omission and commission errors at the 1% density threshold for blind & plausibility validation.
4. Results are provided at lot, bio-geographical regions and country/group of country level and should provide a sound basis for further improving the product for future updates in areas that's show accuracy deficiencies.
5. Based on the lessons learnt from the previous Specific Contract, the chosen stratification procedure was effective for estimating both commission and omission errors with mostly low confidence intervals. However, for the analysis of the scatterplots the sampling does not seem to be ideal as the vast majority of plots has 0% density for both SWF and AWF. A more stratified sampling that includes a larger number of higher density plots is recommended for future validations.
6. There is some confusion between SWF and AWF and also at the 1% threshold (AWF/None, SWF/None).
7. The large number of geometric thresholds used during the production leads to a large number of ambiguous plots during interpretation, especially since the area specifications for patchy SWF and AWF partly overlap, meaning some woody features can indeed be both SWF or AWF.
8. In general, the applied geometric rules are very difficult to determine by visual interpretation (e.g. compactness of a feature) which results in a large number of ambiguous plots. This is reflected by a strong increase of the thematic accuracy during plausibility assessment.
9. The AWF product is closely linked to the SWF product (may not overlap with SWF, enhance connectivity of SWFs, only in areas that are neither SWF nor included in the forest cover mask). It is thus questionable if a blind product validation without knowledge of the presence and location of SWFs is very useful. Checking if the connectivity between SWFs is really enhanced by AWFs requires knowledge on the presence and location of both the SWFs and AWFs in the plot. This cannot be done properly in a blind interpretation.
10. The scatterplots for SWF and AWF mostly show low to medium quality regression coefficients (0.03-0.77). This can be explained by the very large number of plots that disagree at the 0/1% threshold compared to those that are > 0% in both map and reference. Density value are also not scattered evenly from 0-100% but tend to centre around 0% with 70-80 of plots showing density value of 0% in SWF

and/or AWF. Thus R^2 is likely to be lower than for e.g. TCD where also many plots with 80-100% density occur. Due to the large number of plots at or around 0%, the scatterplots are also very difficult to read and to interpret. We recommend to increase the number of sample plots at higher density values for future analyses, to obtain more meaningful scatterplots.

11. A better compatibility of the SWF product with other HRL layers would be beneficial to the users. The product is closely linked to the HRL Forest. Currently the products partly overlaps with TCD and FTY, which is not intuitive to the users. However, we do understand the problems that arise due to different definitions used and a full compatibility is currently not possible. Whether HRL layer definitions can be adapted to make products more compatible, should be part of a separate analysis.
12. The intermediate forest cover mask that was used during production seems to be an integral part of the product (exclusion areas) and we recommend to make this SWF forest cover mask available as a support layer to SWF. In some areas it may not be clear to the users why certain woody features are excluded from the different SWF Layers.
13. In the course of this evaluation, only the 100m raster product is validated and therefore the results probably do not reflect the quality of the 5m raster product or even the vector product, which also map "smaller features".

7. References

- Gallego, J. (1995). *Sampling Frames of Square Segments*. Official Publication of the European Communities, EUR 16317, 72 pp.
- Gallego, J. (2004). Area Frames for Land Cover Estimation: Improving the European LUCAS Survey. *Third International Conference on Agricultural Statistics*. November 2-4, 2004, Mexico.
- Sannier, Ch. (2020): Thematic Geospatial Product Validation Guidelines – Final Report. Report within the framework of *Validation Services for the geospatial products of the Copernicus land Continental and local components including in-situ data (lot 1)*. Contract EEA/MDI/14/010. Near completion.
- Stehman, S. V. (2009). Sampling designs for accuracy assessment of land cover. *International Journal of Remote Sensing*, 30(20), 5243-5272.

Annex 1. Small Woody Features (SWF) Scatterplot for bio-geographical regions

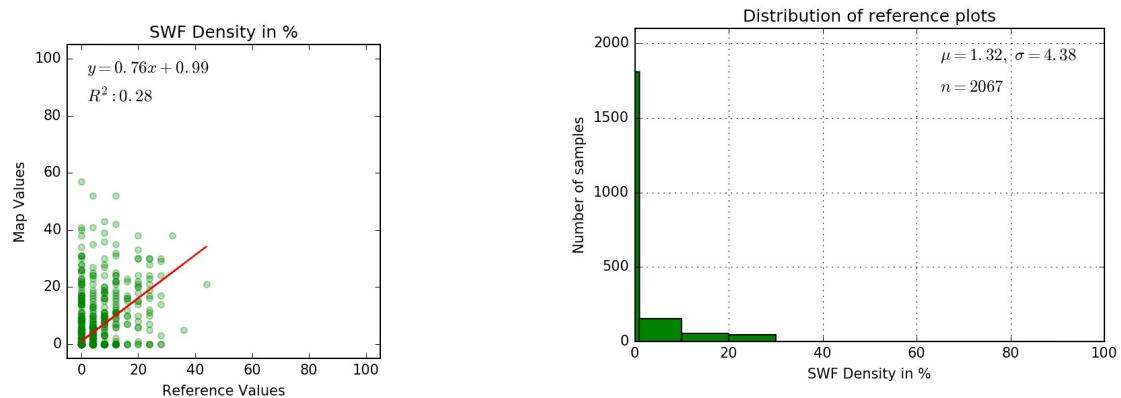


Figure 34: SWF Layer - Alpine bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values.

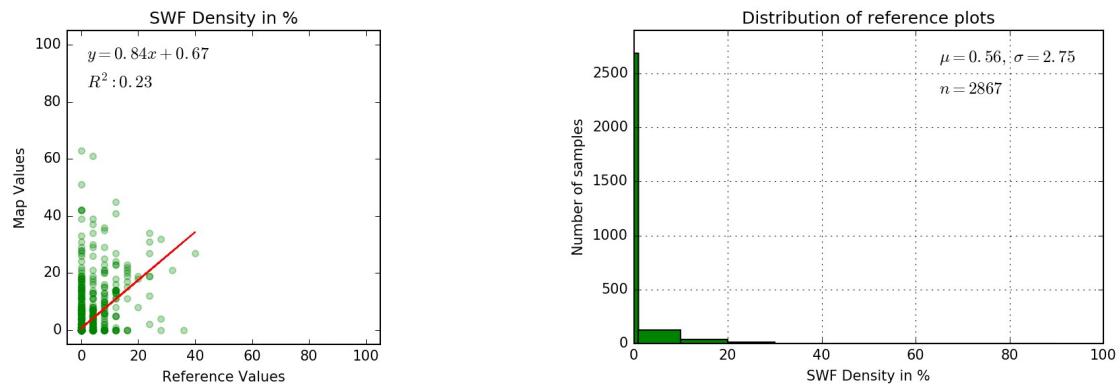


Figure 35: SWF Layer - Anatolian bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values.

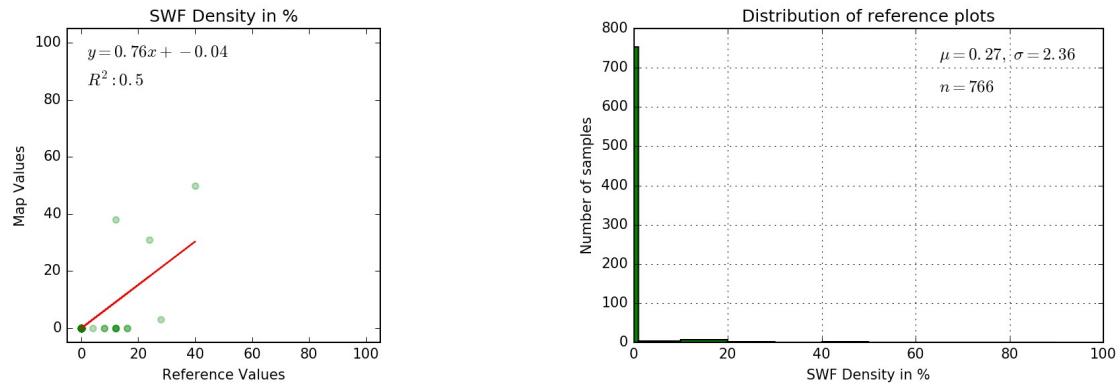


Figure 36: SWF Layer - Arctic bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values.

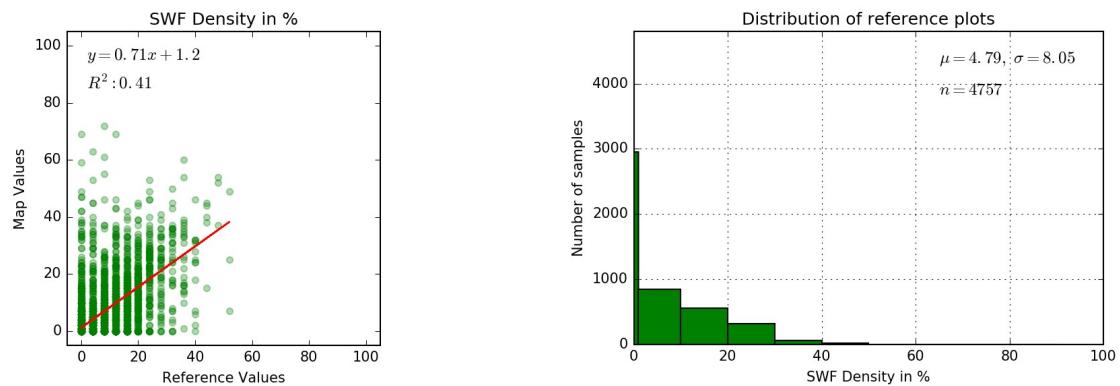


Figure 37: SWF Layer - the Atlantic bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values.

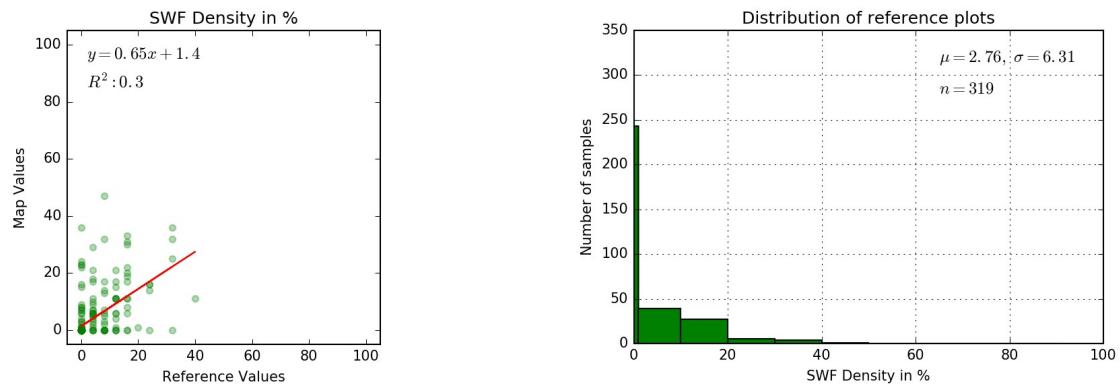


Figure 38: SWF Layer - Black Sea bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values.

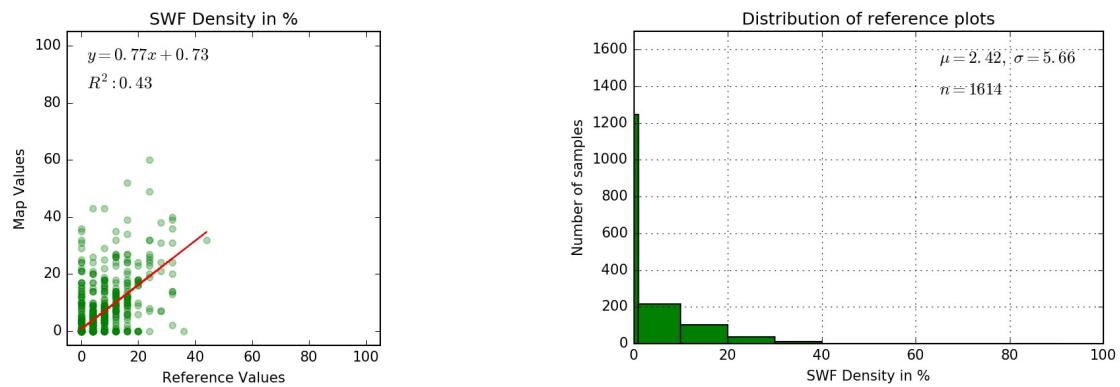


Figure 39: SWF Layer - Boreal bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values.

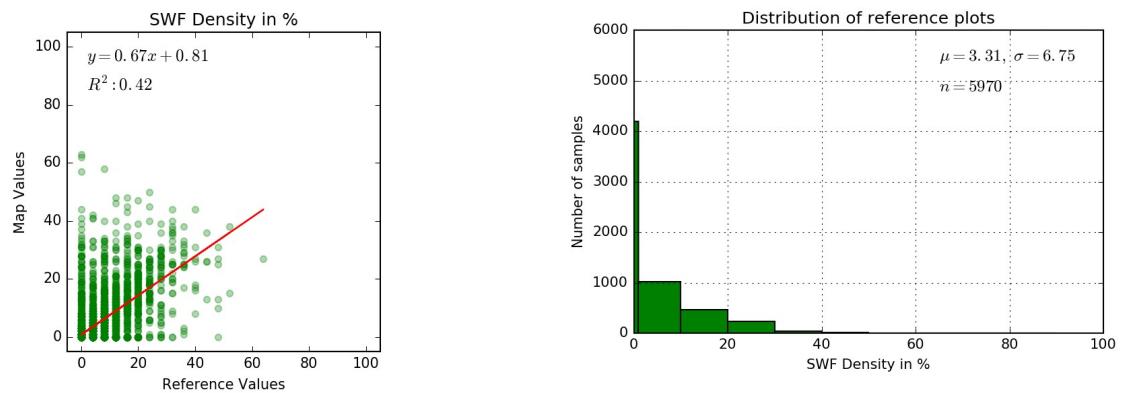


Figure 40: SWF Layer - Continental bio-geographical: (left) Scatterplot of all sample units. (Right) Distribution of density values.

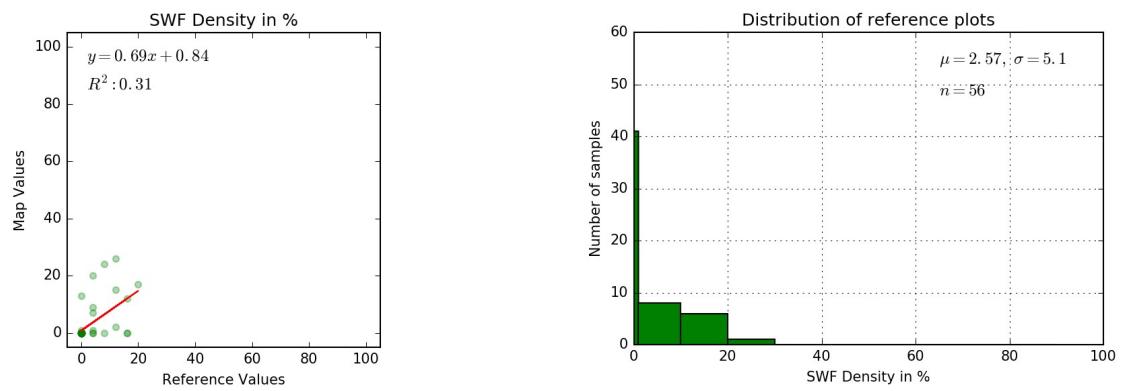


Figure 41: SWF Layer - Macaronesian bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values.

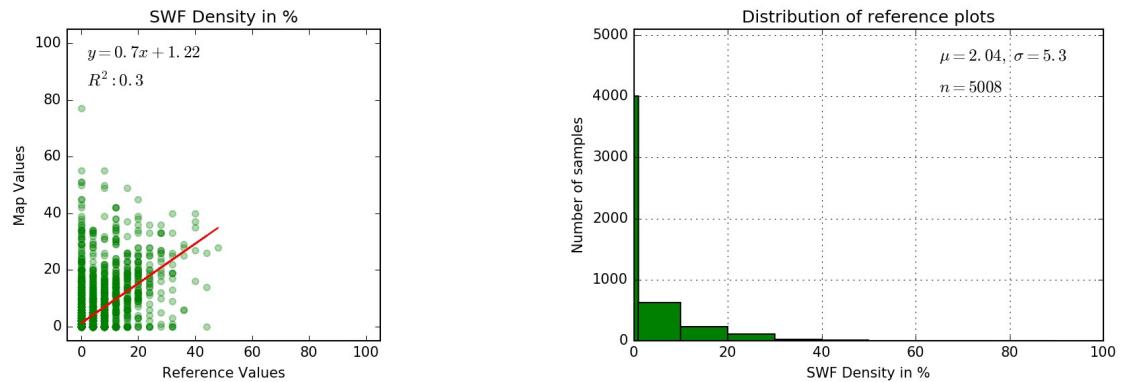


Figure 42: SWF Layer - Mediterranean bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values.

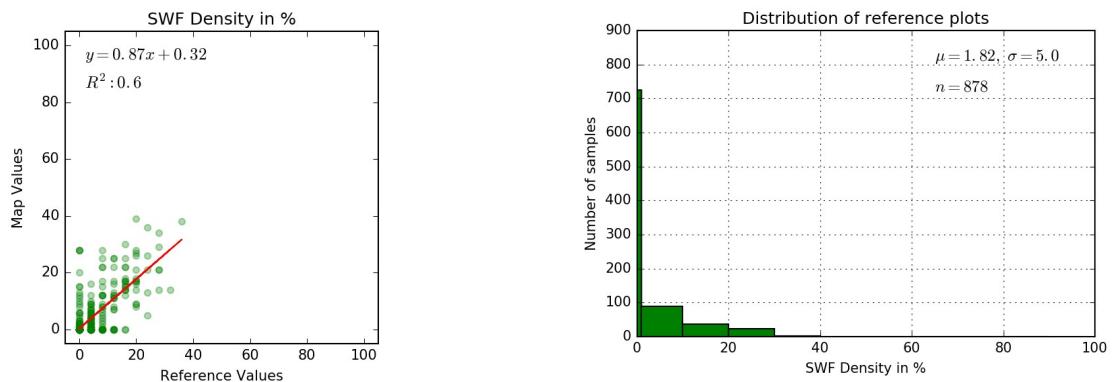


Figure 43: SWF Layer - Pannonian bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values.

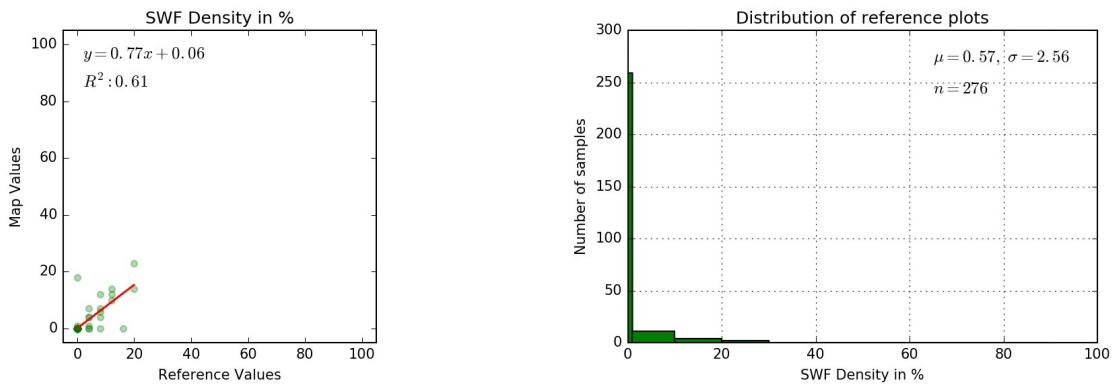


Figure 44: SWF Layer - Steppic bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values.

Annex 2. Additional Woody Features (AWF) Scatterplot for bio-geographical regions

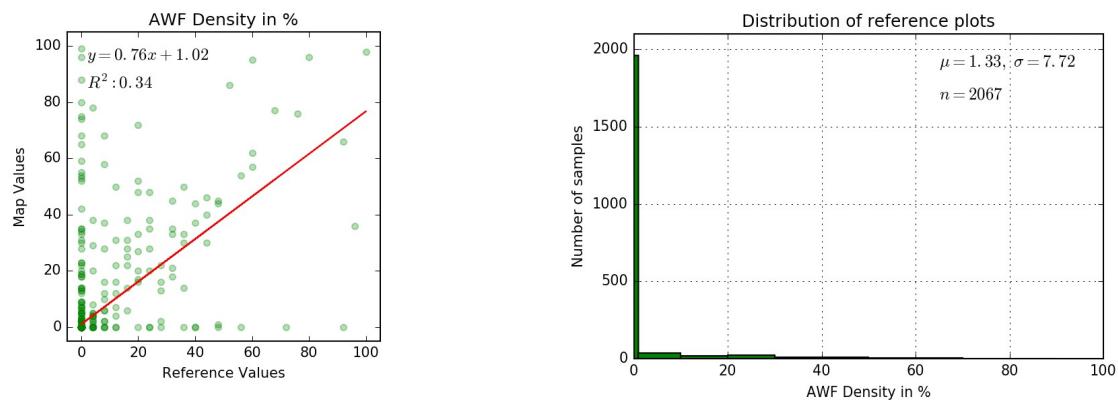


Figure 45: AWF Layer - Alpine bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values.

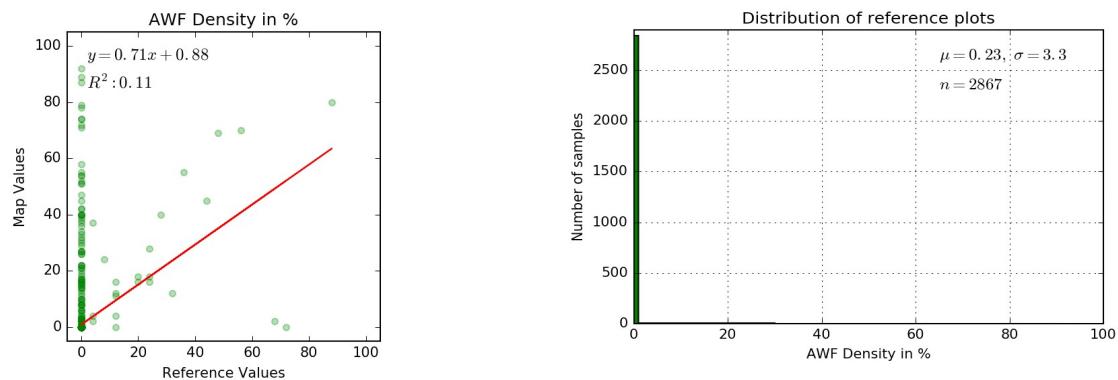


Figure 46: AWF Layer - Anatolian bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values.

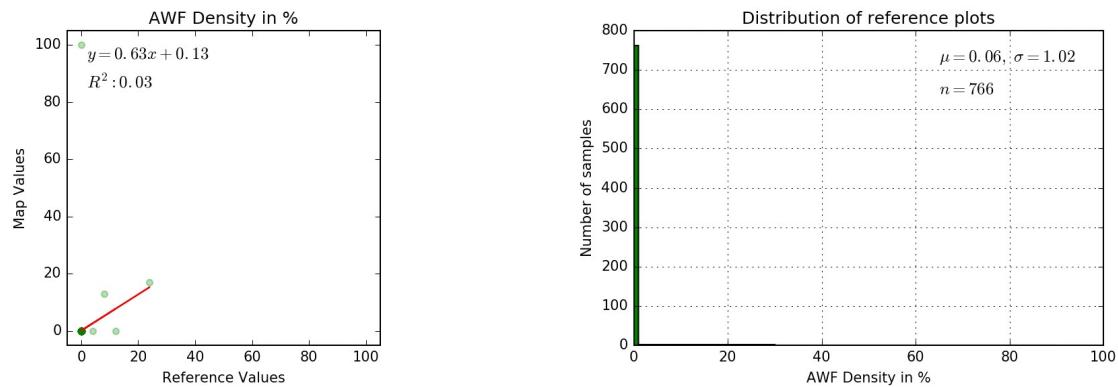


Figure 47: AWF Layer - Arctic bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values.

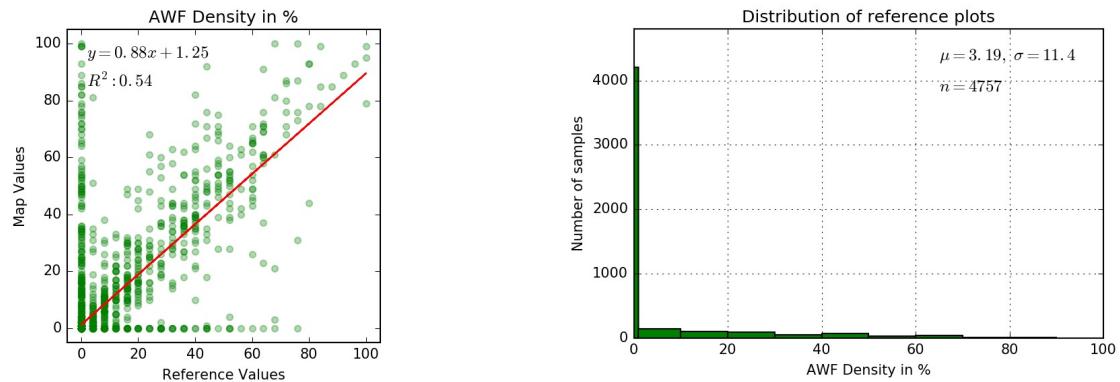


Figure 48: AWF Layer - the Atlantic bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values.

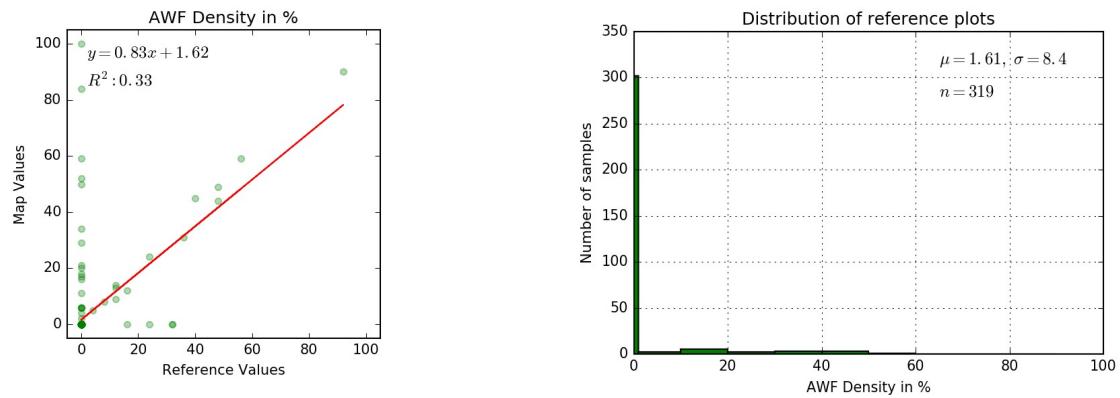


Figure 49: AWF Layer - Black Sea bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values.

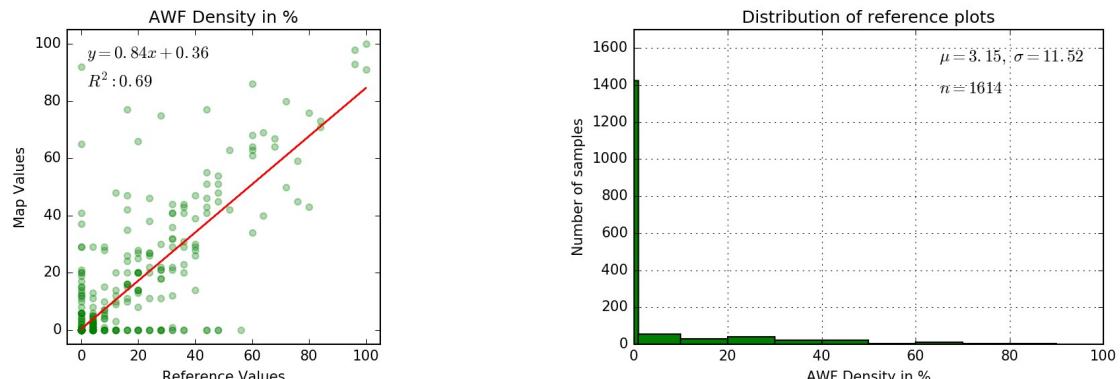


Figure 50: AWF Layer - Boreal bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values.

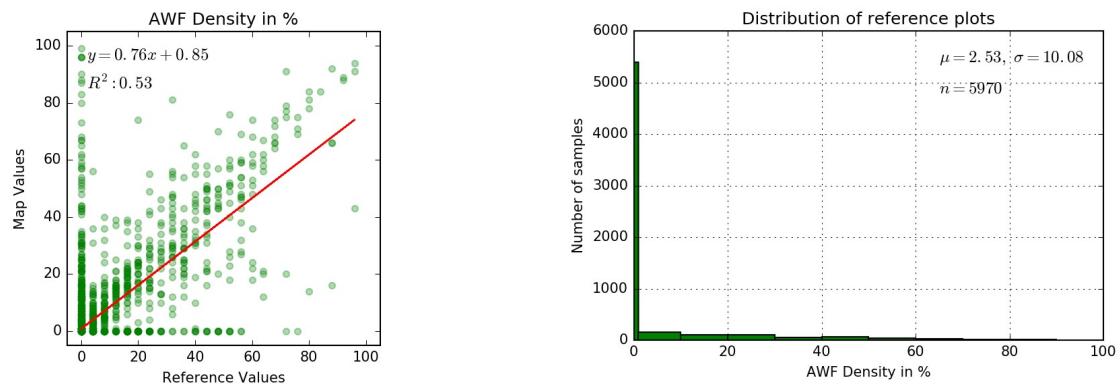


Figure 51: AWF Layer - Continental bio-geographical: (left) Scatterplot of all sample units. (Right) Distribution of density values.

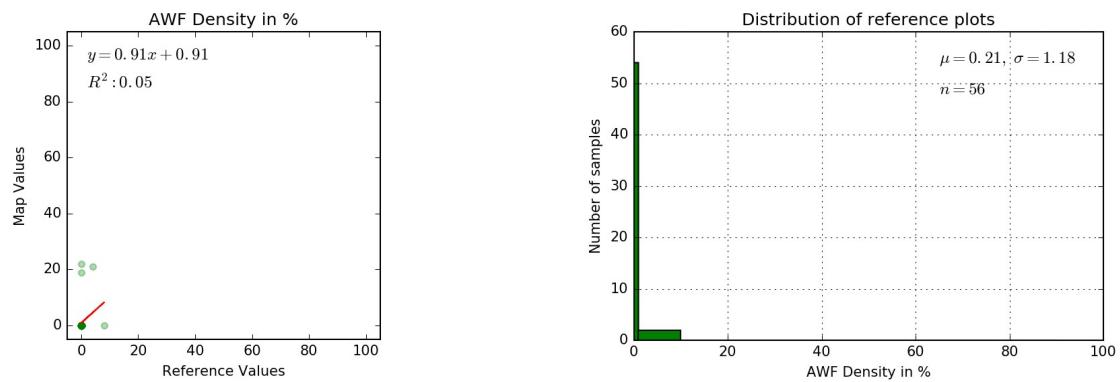


Figure 52: AWF Layer - Macaronesian bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values.

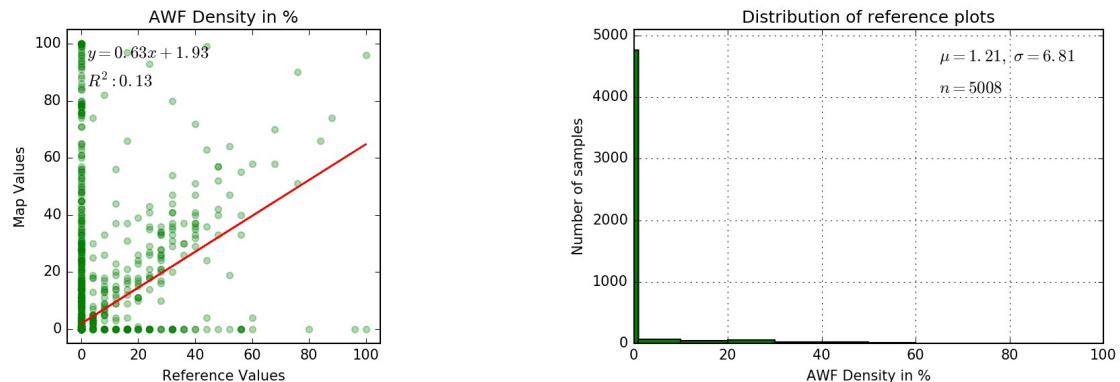


Figure 53: AWF Layer - Mediterranean bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values

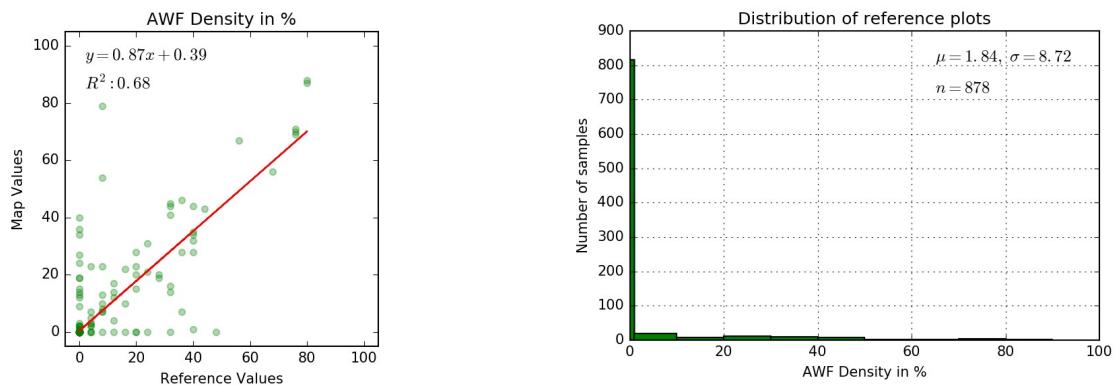


Figure 54: AWF Layer - Pannonian bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values.

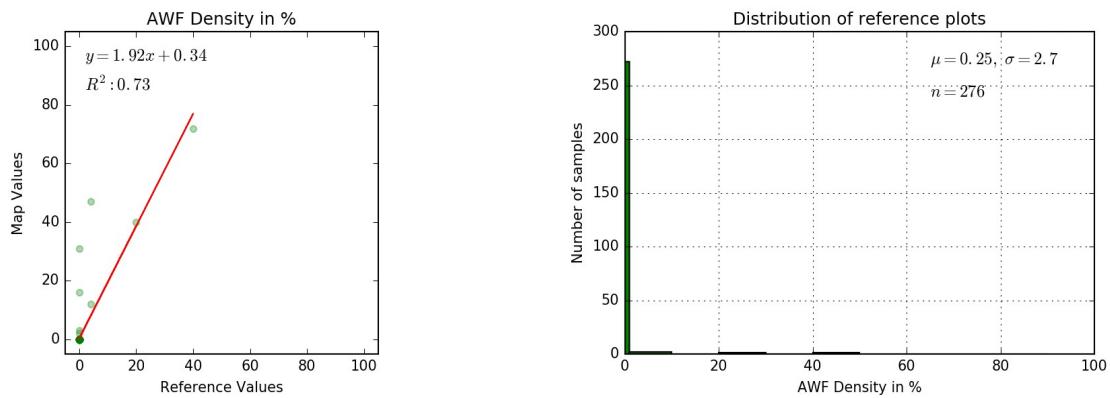


Figure 55: AWF Layer - Steppic bio-geographical region: (left) Scatterplot of all sample units. (Right) Distribution of density values.

Annex 3. Small Woody Features (SWF) Scatterplot for countries and group of countries greater than 90,000km²

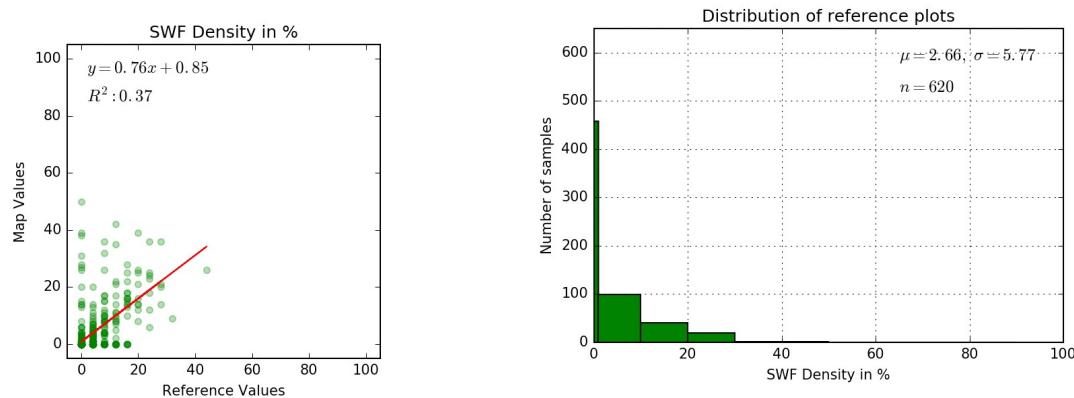


Figure 56: SWF Layer – AL_MK_ME_RS_KS country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

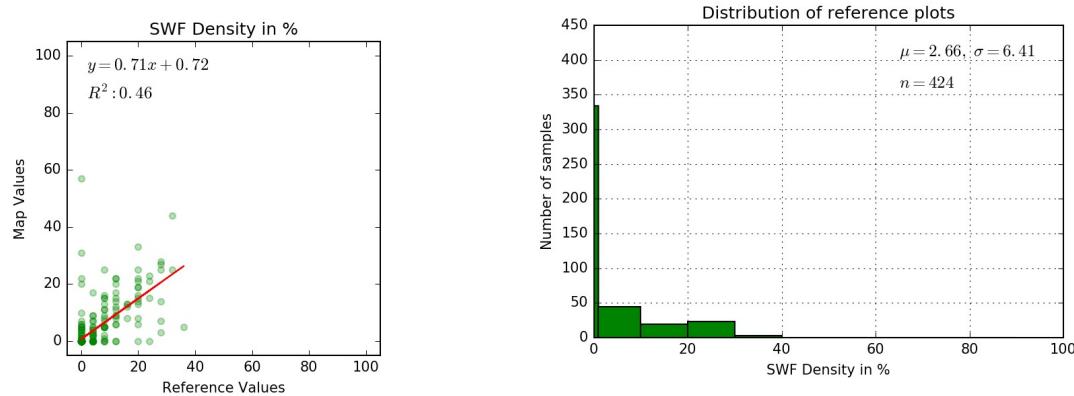


Figure 57: SWF Layer – AT_CH_LI country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

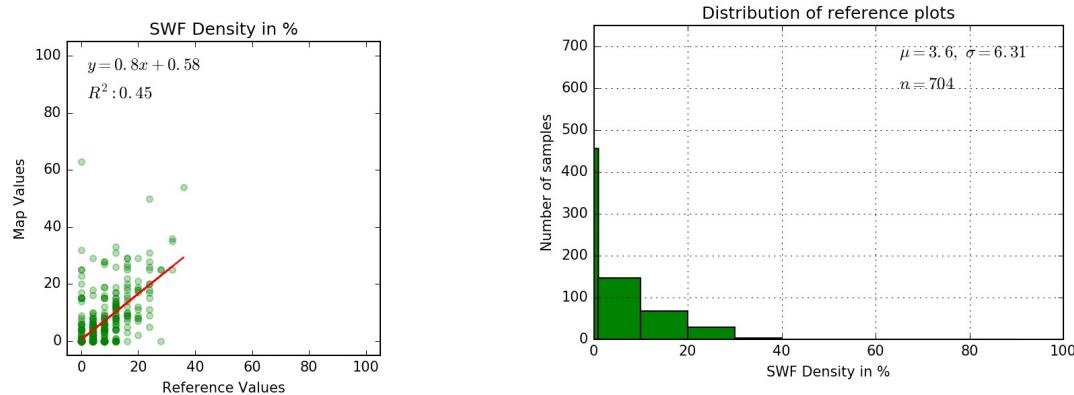


Figure 58: SWF Layer – BE_NL_LU_DK country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

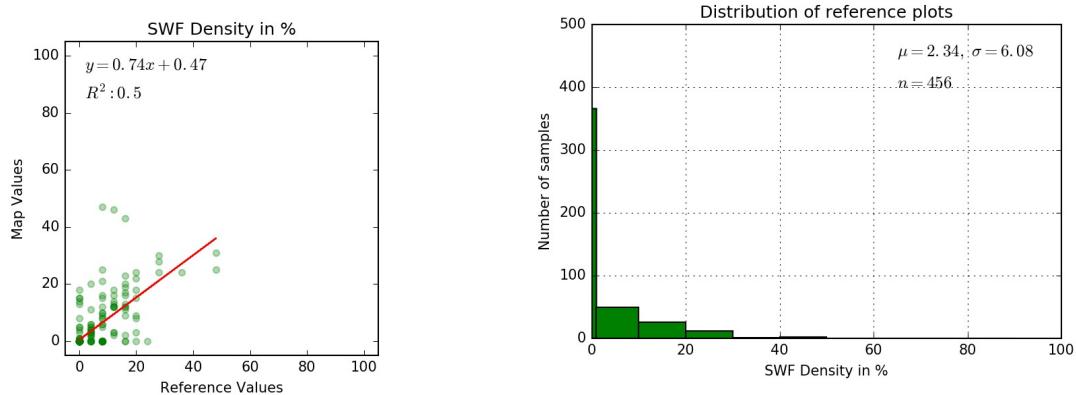


Figure 59: SWF Layer – BG country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

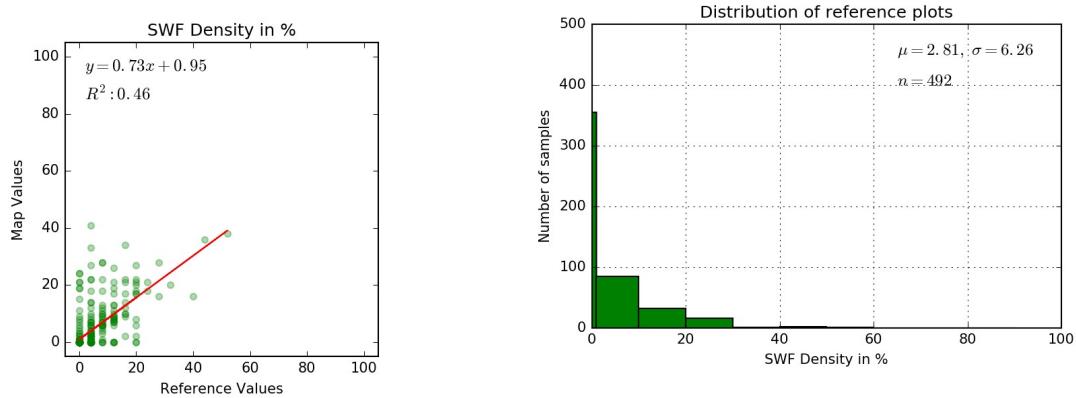


Figure 60: SWF Layer – CZ_SK country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

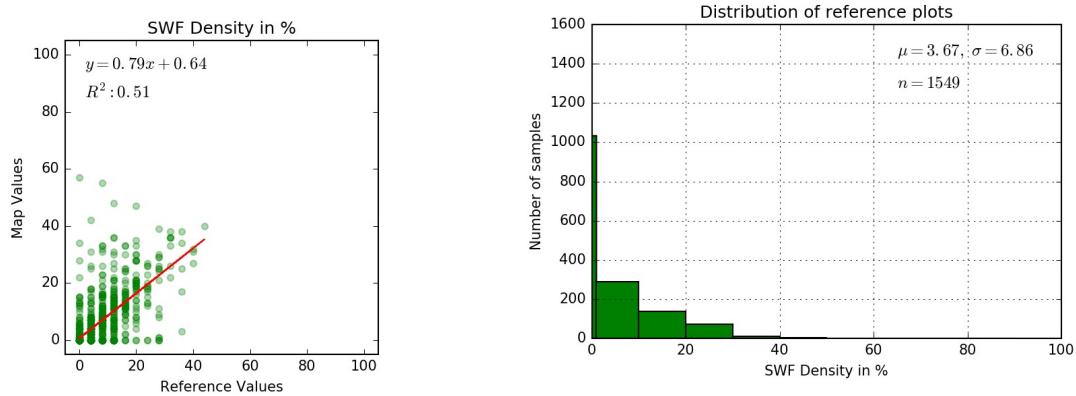


Figure 61: SWF Layer – DE country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

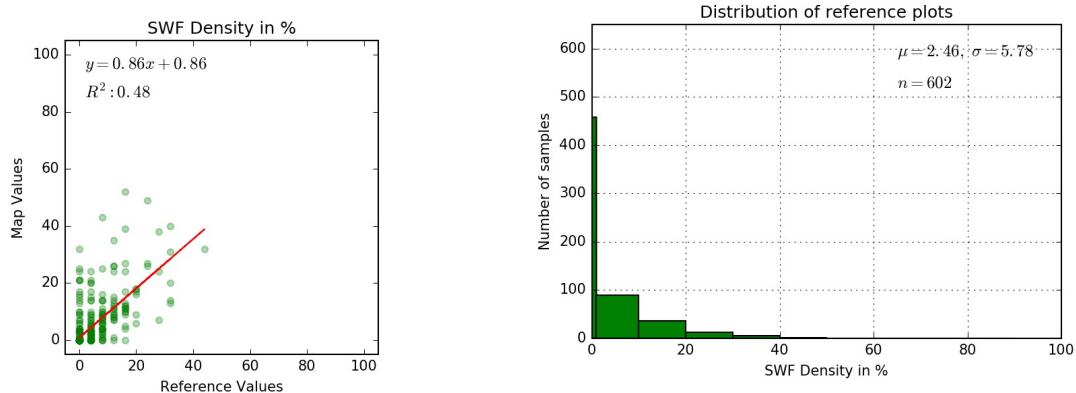


Figure 62: SWF Layer – EE_LT_LV country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

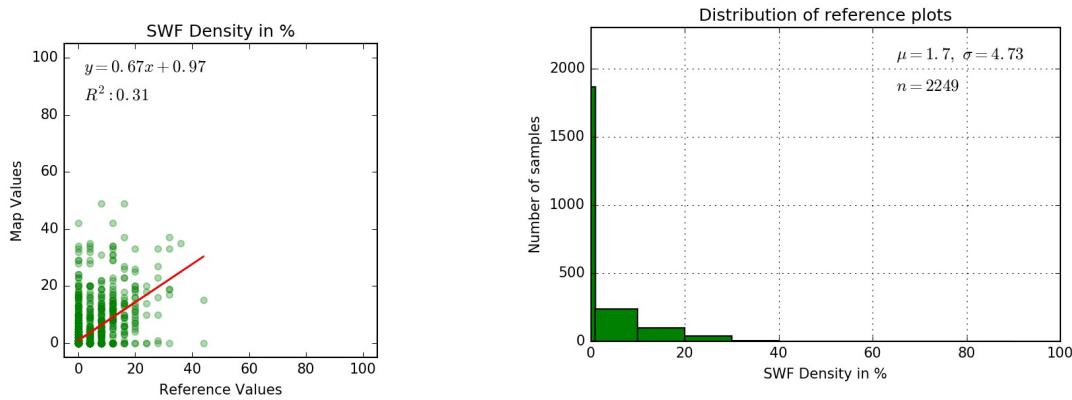


Figure 63: SWF Layer – ES country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values

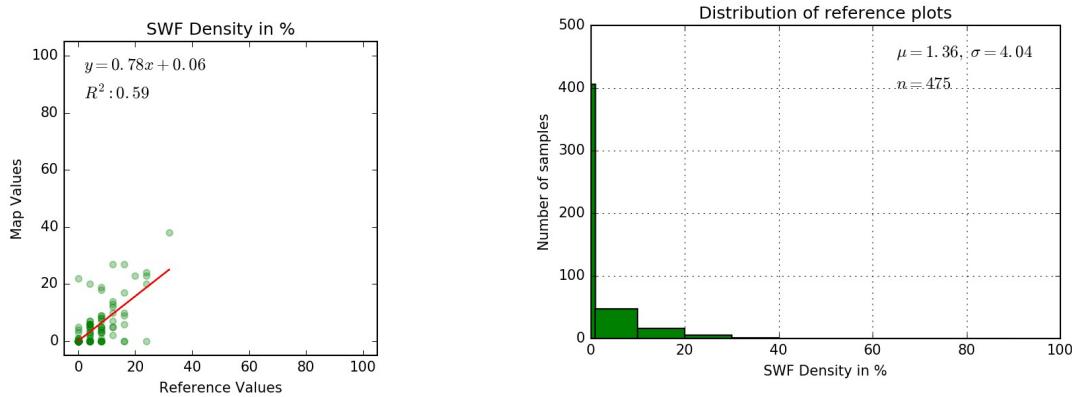


Figure 64: SWF Layer – FI country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

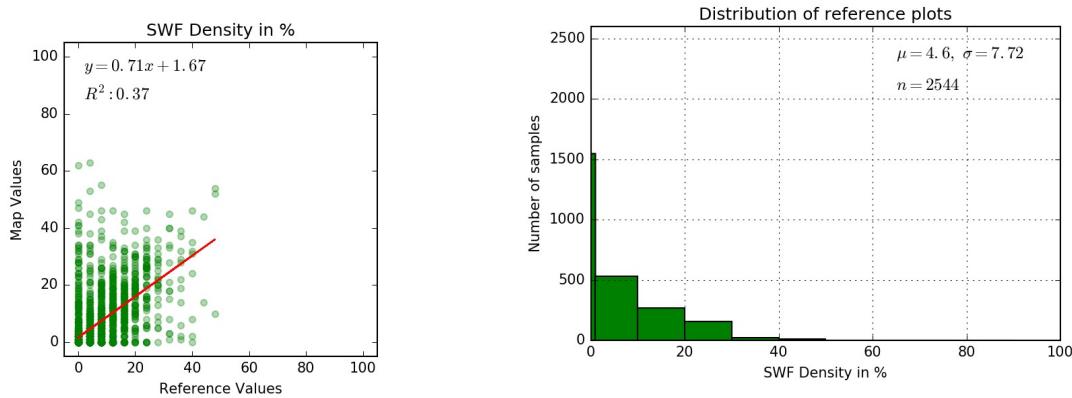


Figure 65: SWF Layer – FR_AD country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

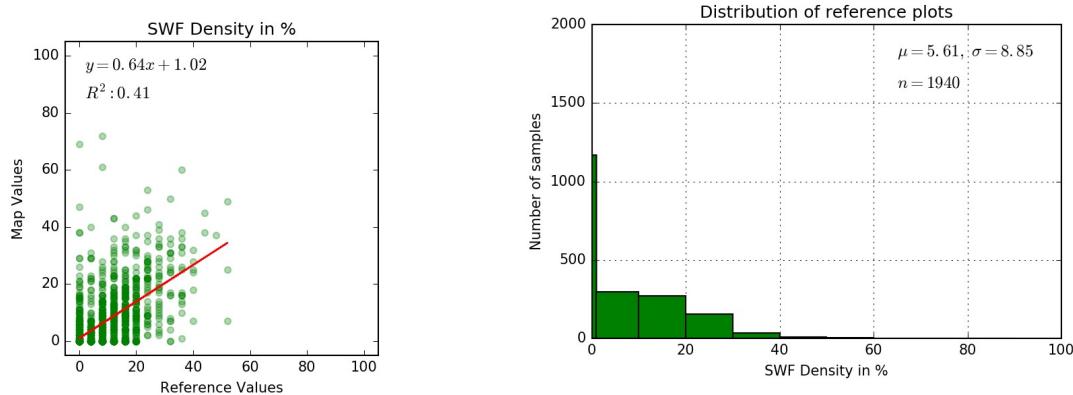


Figure 66: SWF Layer – GB_IE country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values

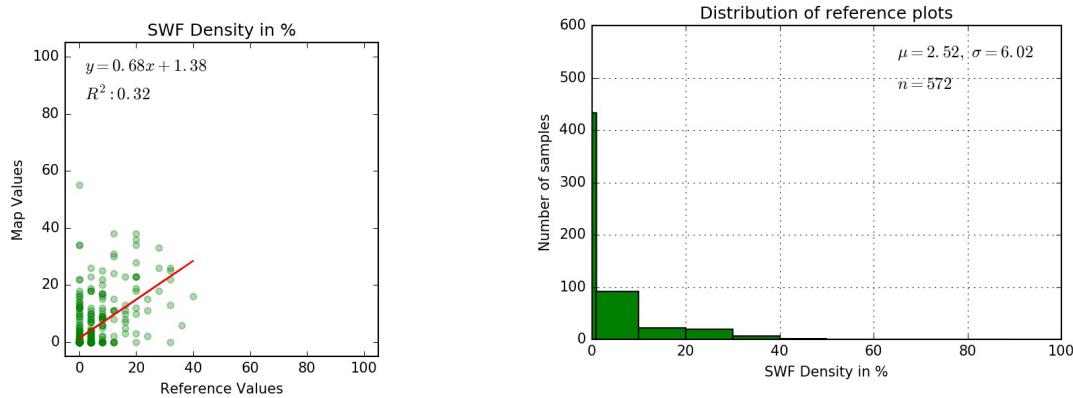


Figure 67: SWF Layer – GR_CY country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

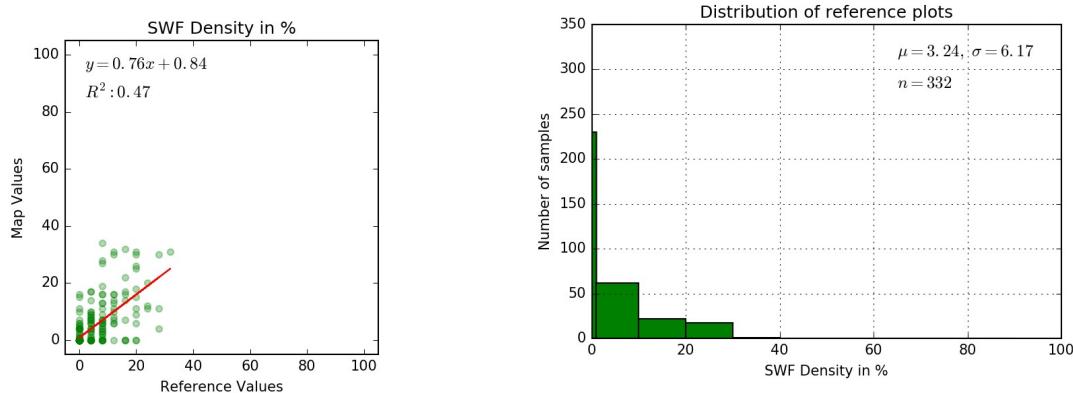


Figure 68: SWF Layer – HR_BA_SI country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

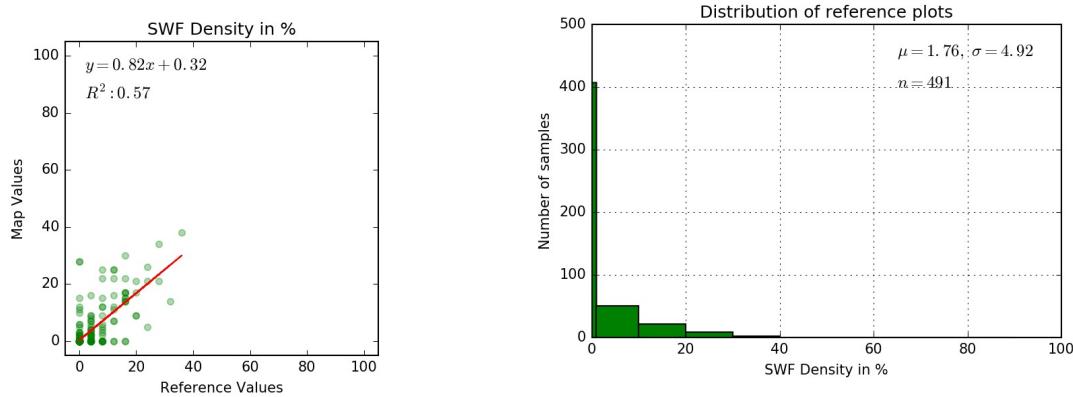


Figure 69: SWF Layer – HU country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

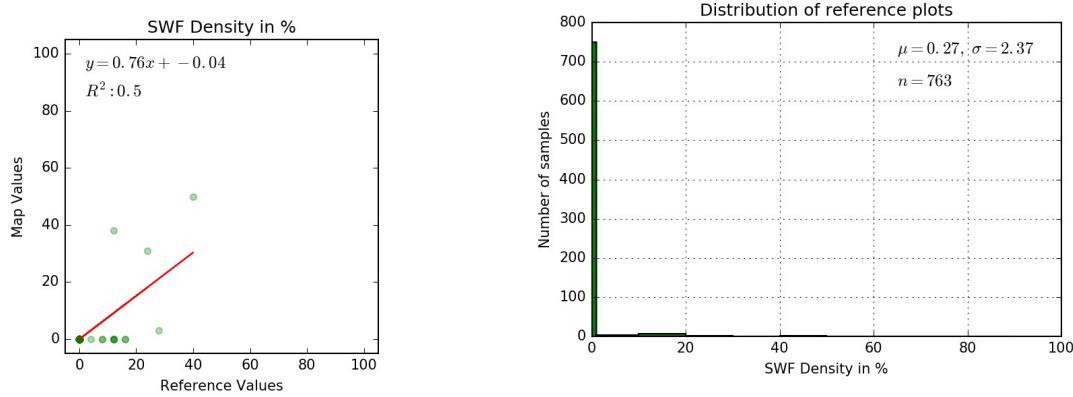


Figure 70: SWF Layer – IS country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

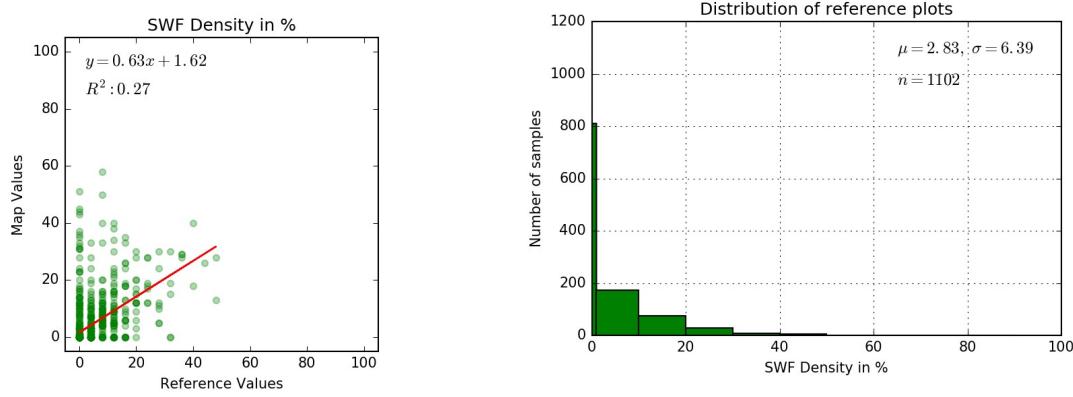


Figure 71: SWF Layer – IT_MT country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

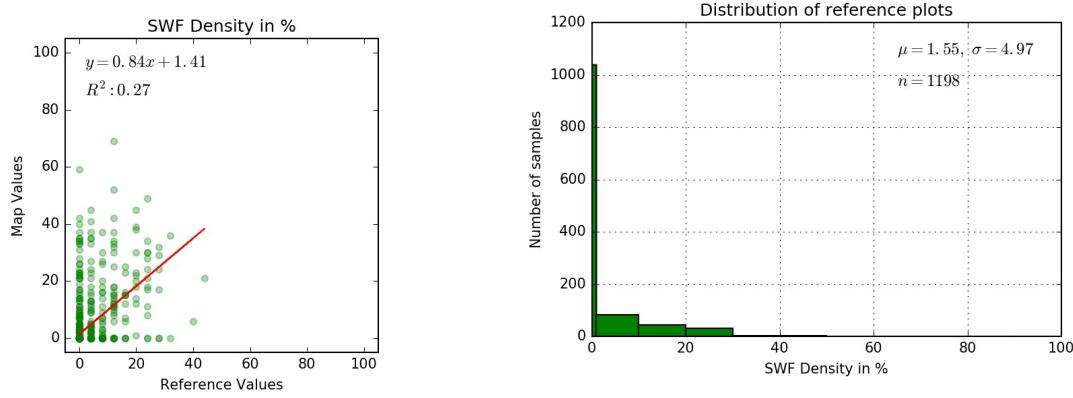


Figure 72: SWF Layer – NO country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

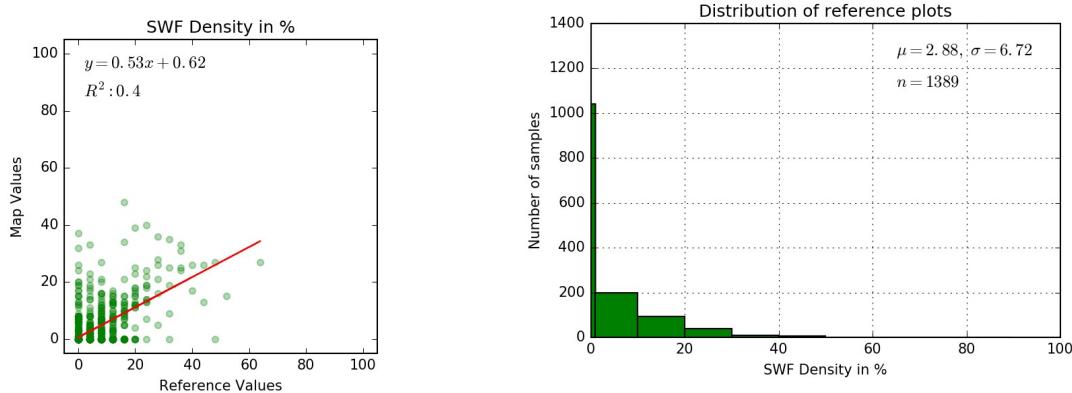


Figure 73: SWF Layer – PL country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

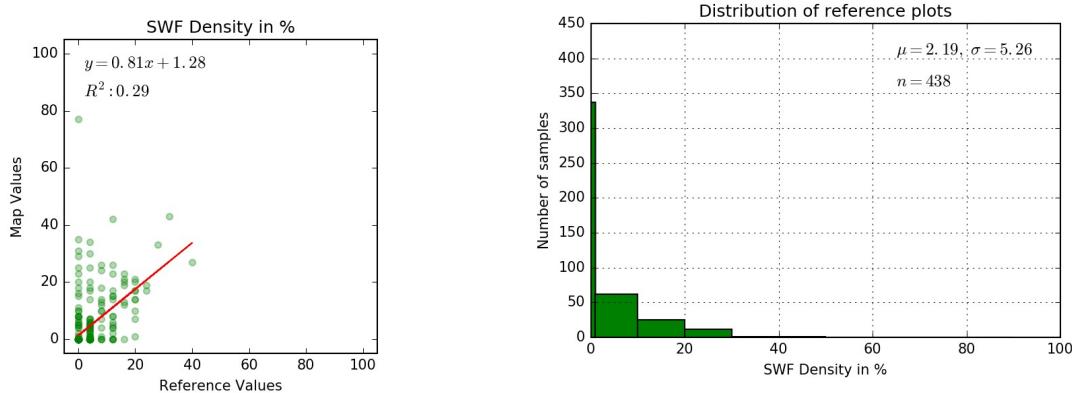


Figure 74: SWF Layer – PT country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

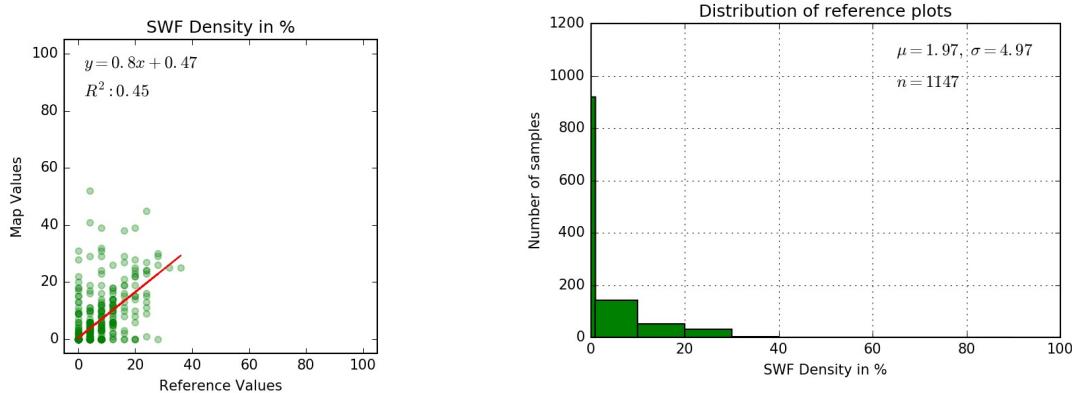


Figure 75: SWF Layer – RO country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values

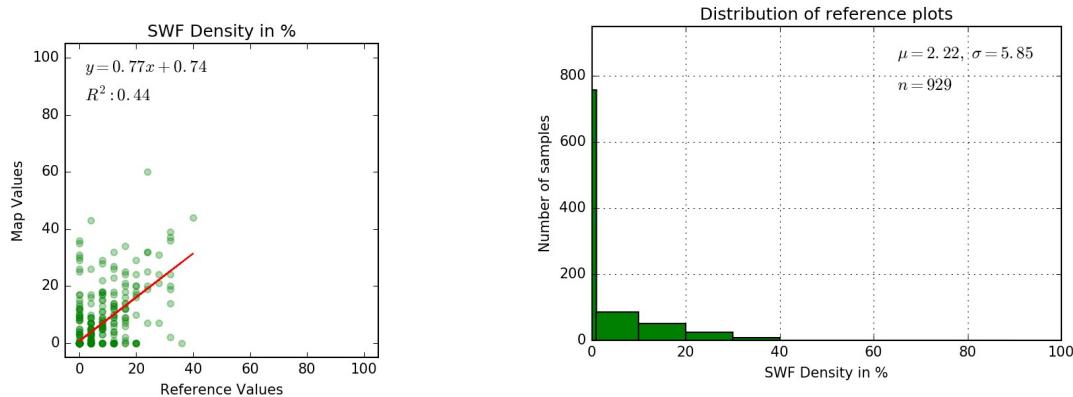


Figure 76: SWF Layer – SE country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

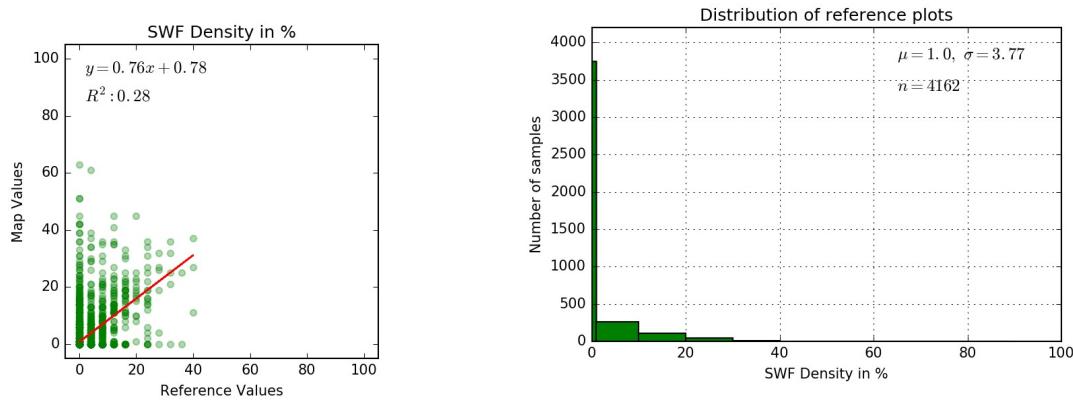


Figure 77: SWF Layer – TR country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

Annex 4. Additional Woody Features (AWF) Scatterplot for countries and group of countries greater than 90,000km²

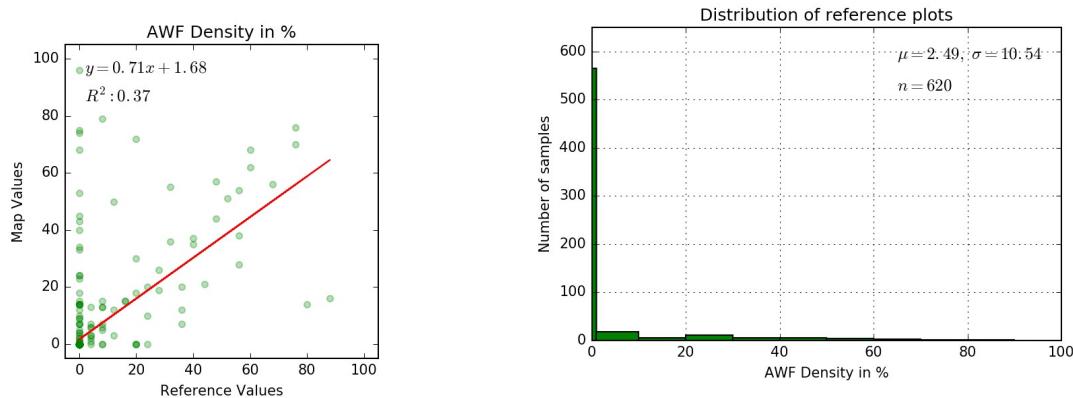


Figure 78: AWF Layer – AL_MK_ME_RS_KS country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

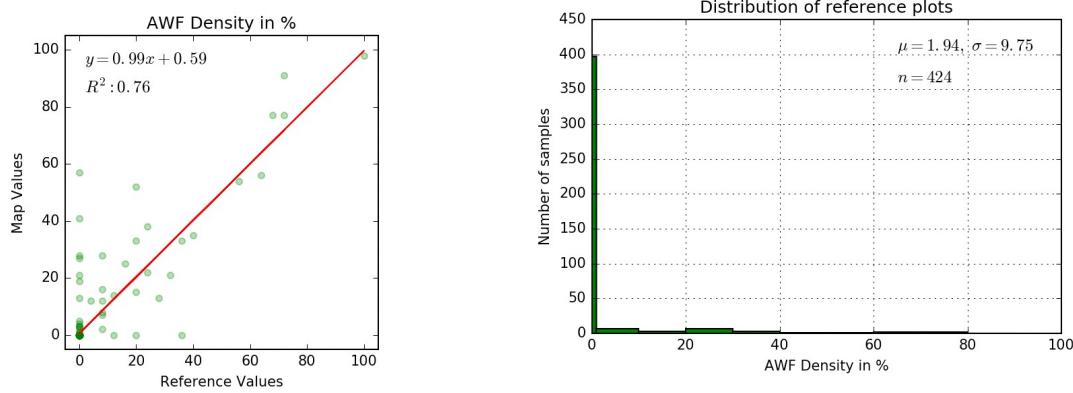


Figure 79: AWF Layer – AT_CH_LI country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

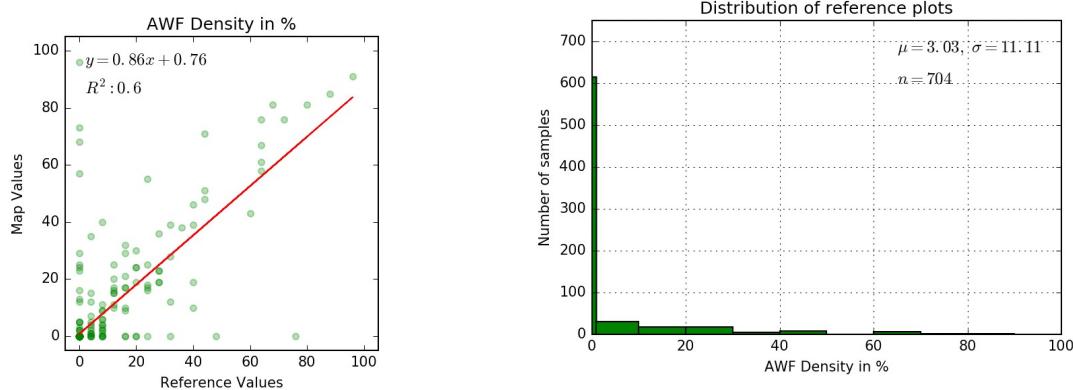


Figure 80: AWF Layer – BE_NL_LU_DK country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

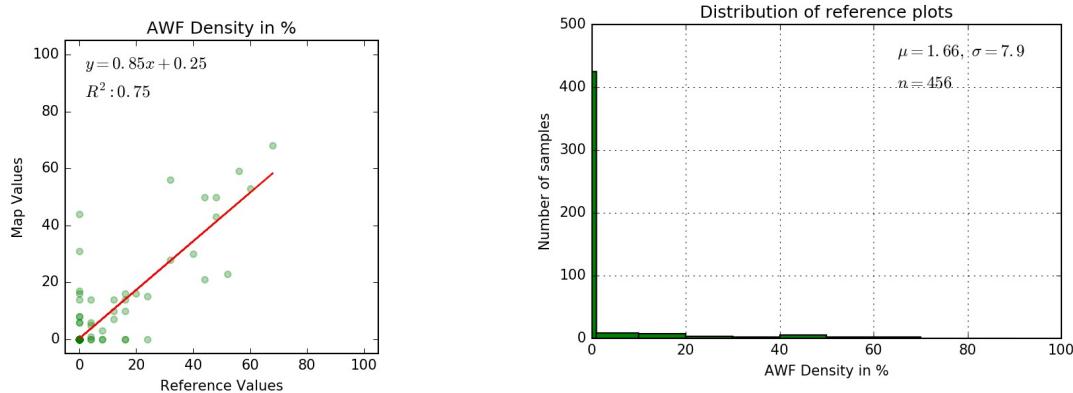


Figure 81: AWF Layer – BG country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

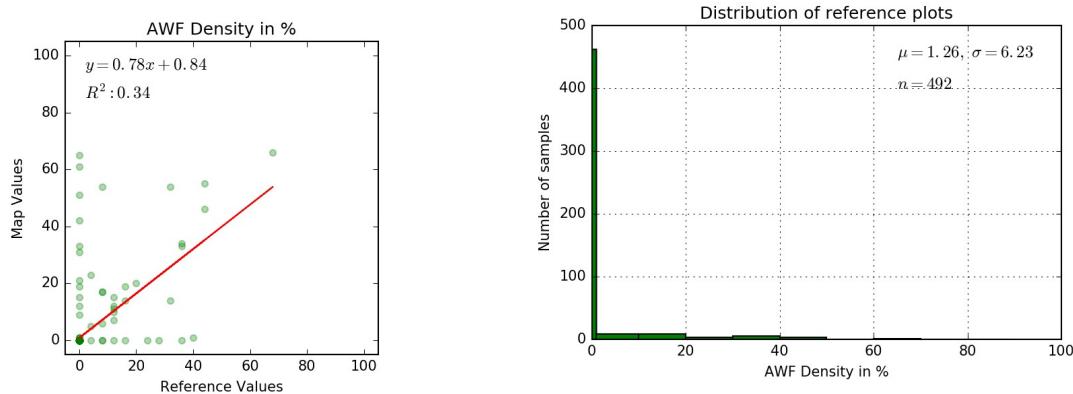


Figure 82: AWF Layer – CZ_SK country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

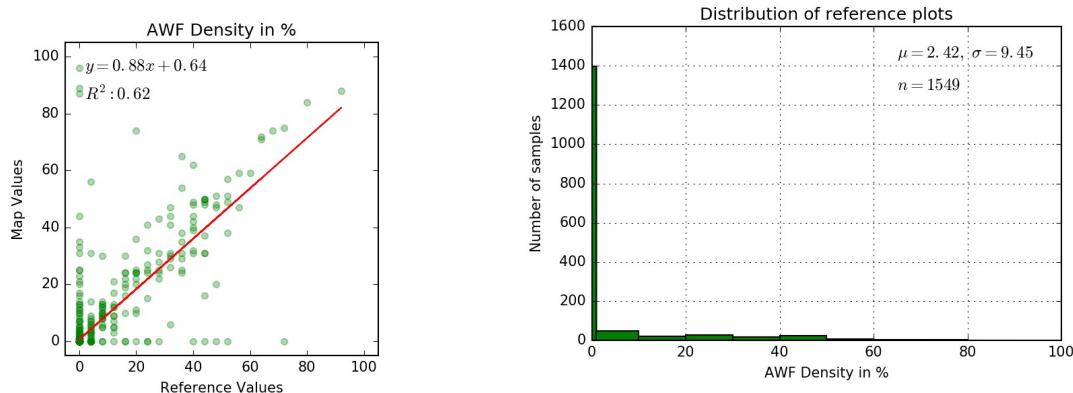


Figure 83: AWF Layer – DE country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

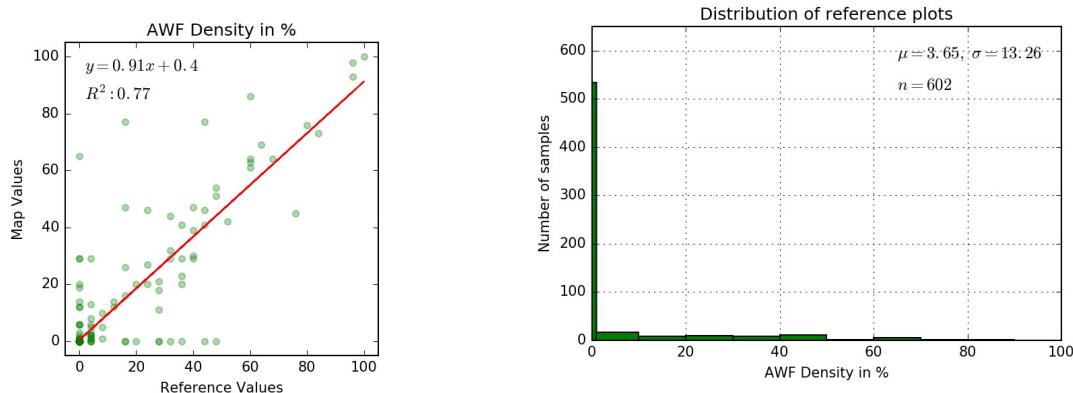


Figure 84: AWF Layer – EE_LT_LV country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values

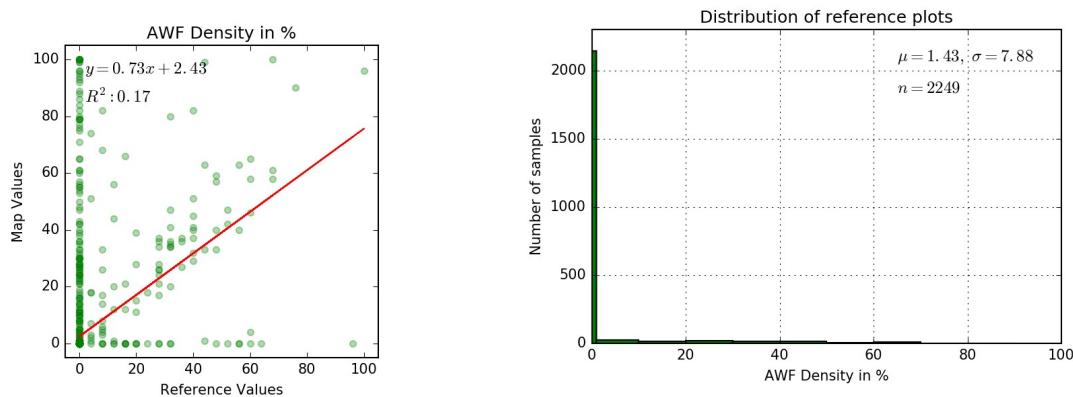


Figure 85: AWF Layer – ES country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

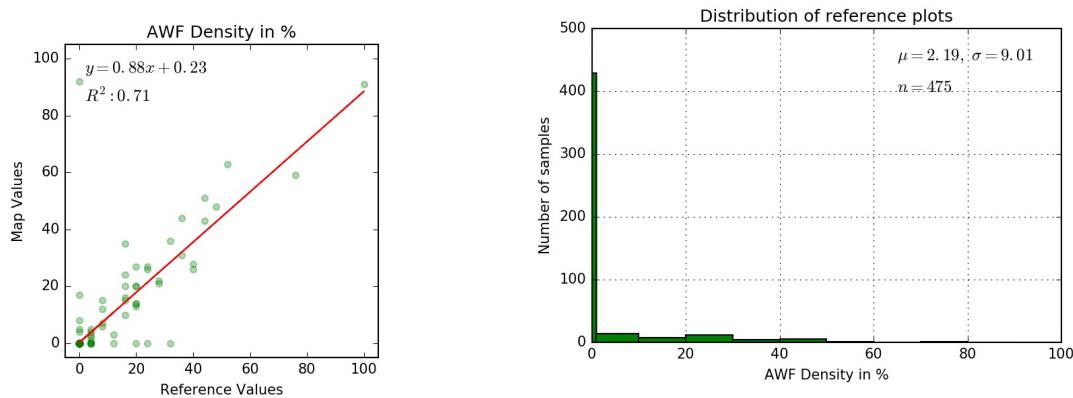


Figure 86: AWF Layer – FI country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

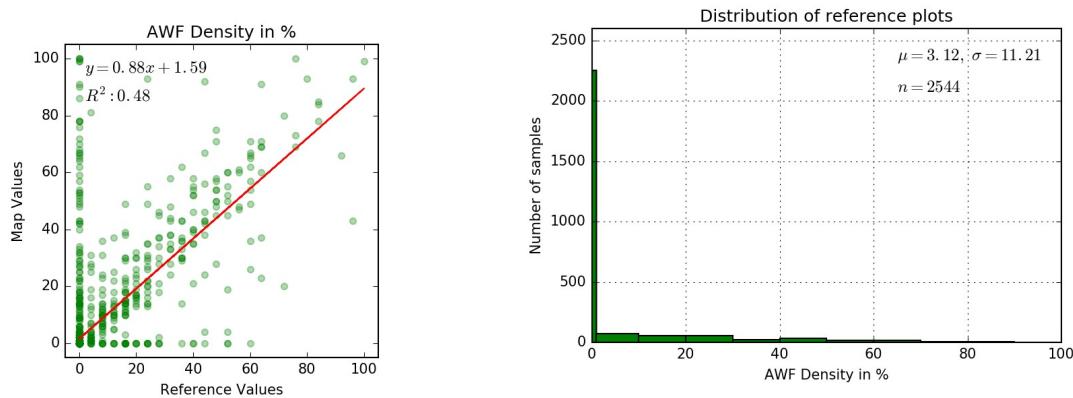


Figure 87: AWF Layer – FR_AD country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

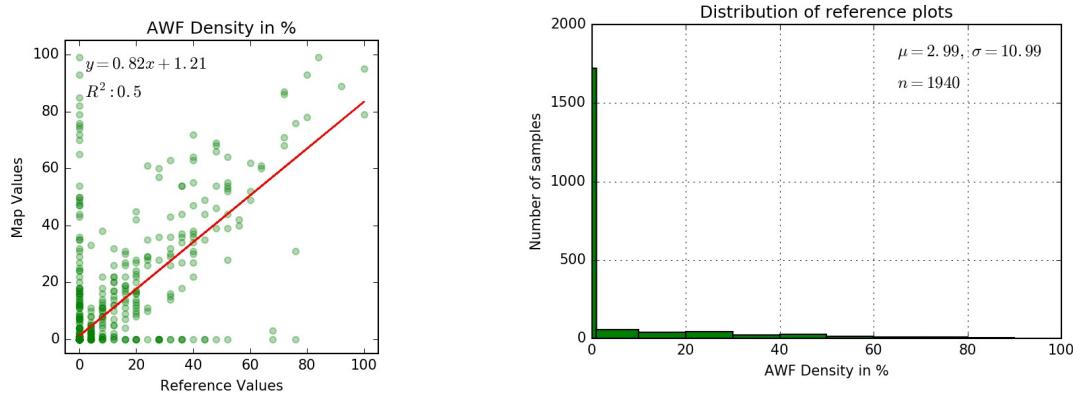


Figure 88: AWF Layer – GB_IE country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

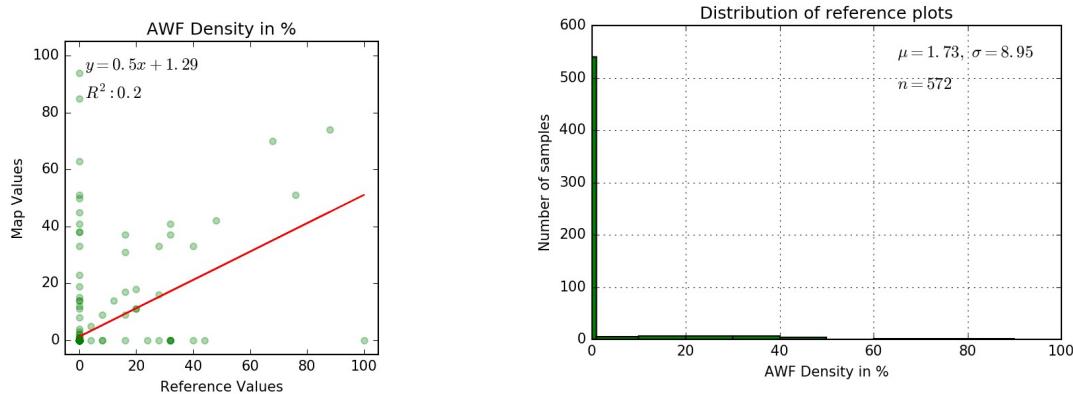


Figure 89: AWF Layer – GR_CY country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

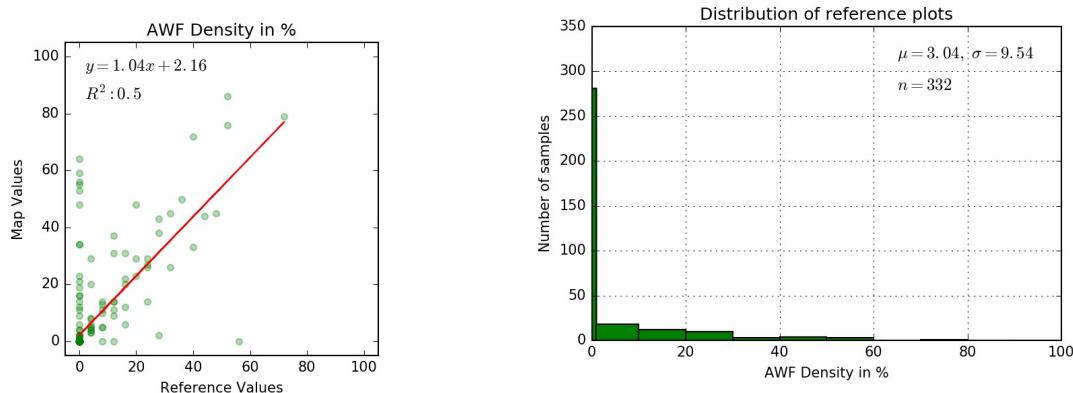


Figure 90: AWF Layer – HR_BA_SI country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

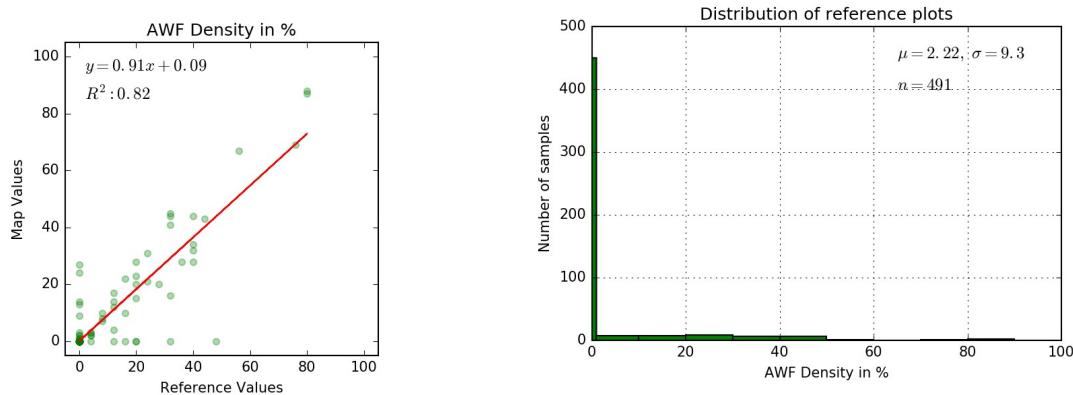


Figure 91: AWF Layer – HU country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

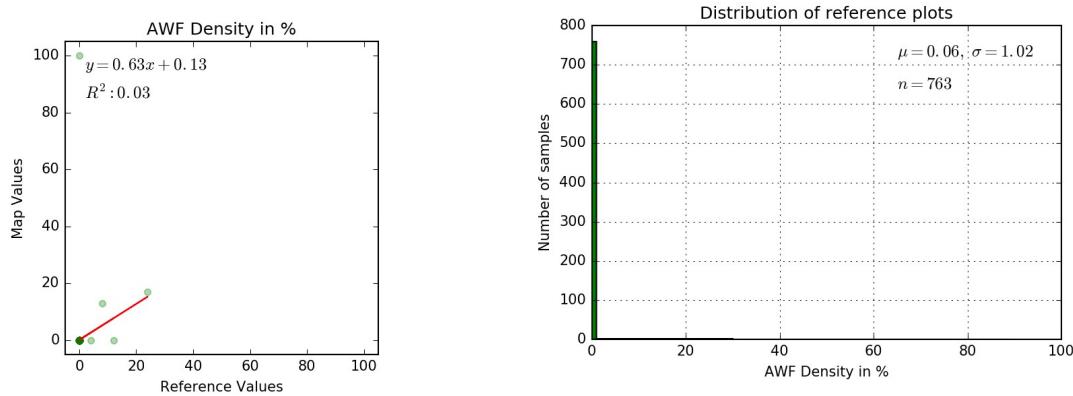


Figure 92: AWF Layer – IS country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

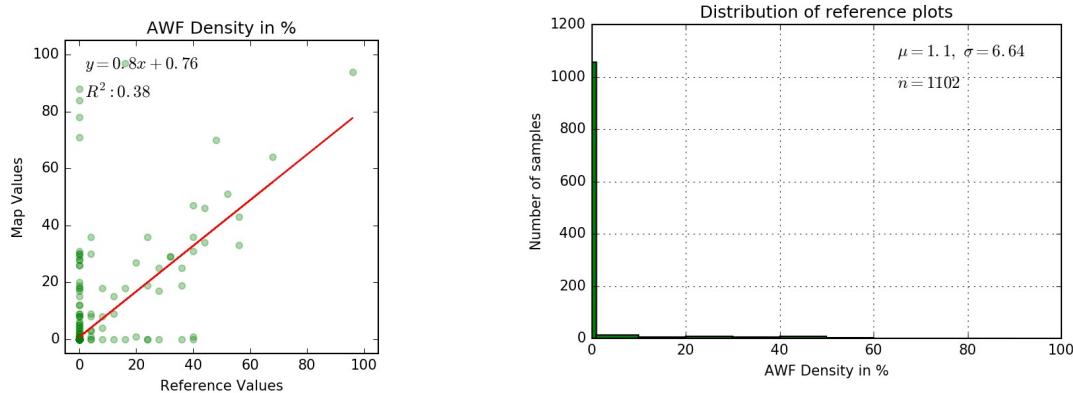


Figure 93: AWF Layer – IT_MT country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

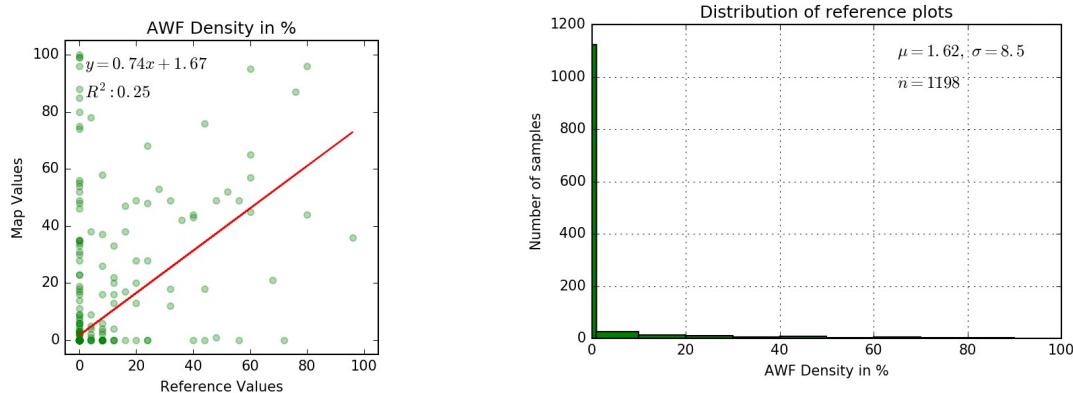


Figure 94: AWF Layer – NO country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

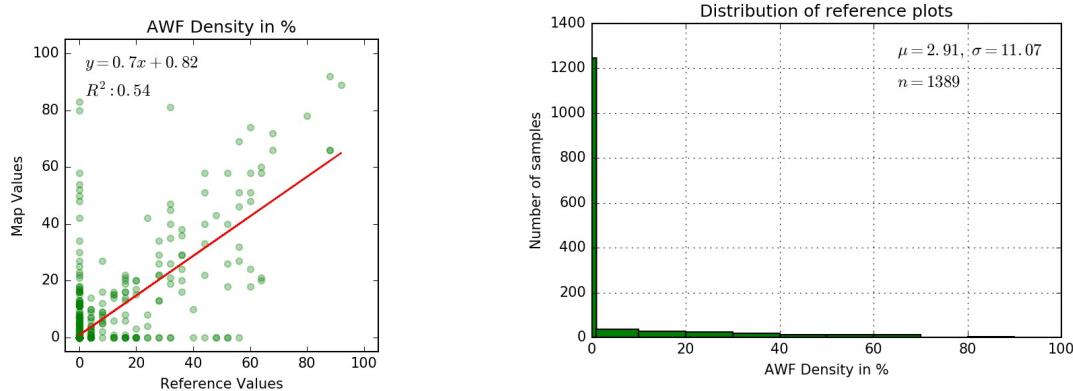


Figure 95: AWF Layer – PL country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

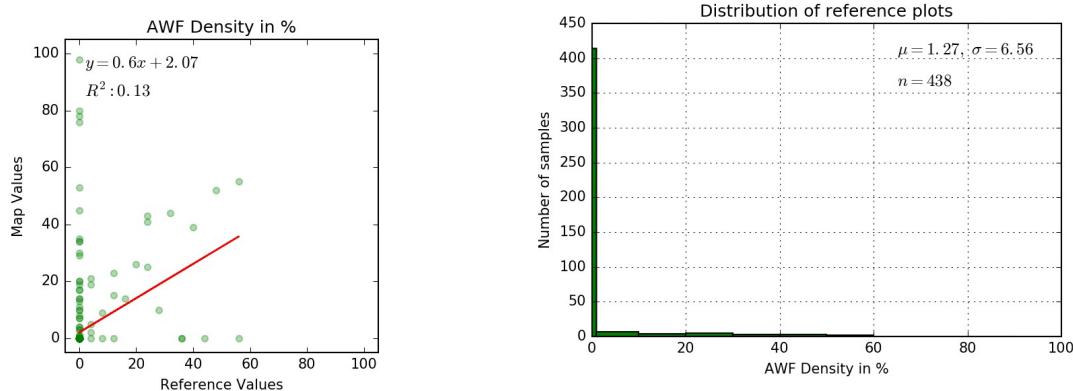


Figure 96: AWF Layer – PT country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

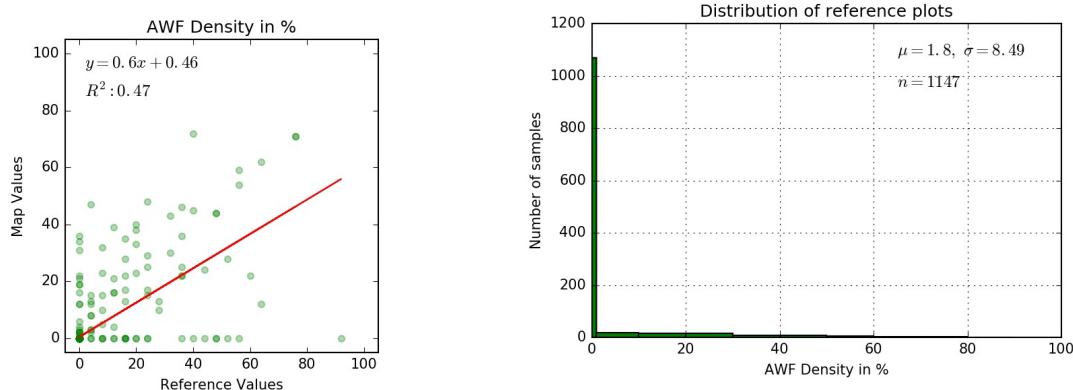


Figure 97: AWF Layer – RO country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

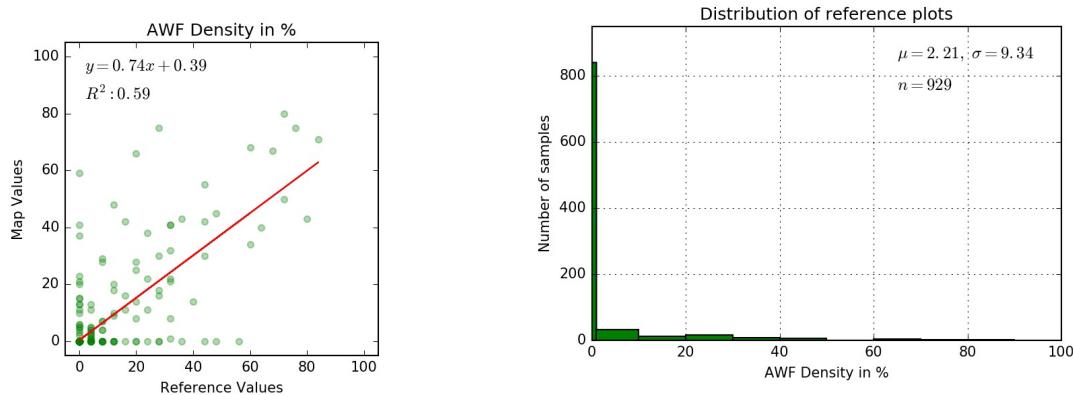


Figure 98: AWF Layer – SE country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

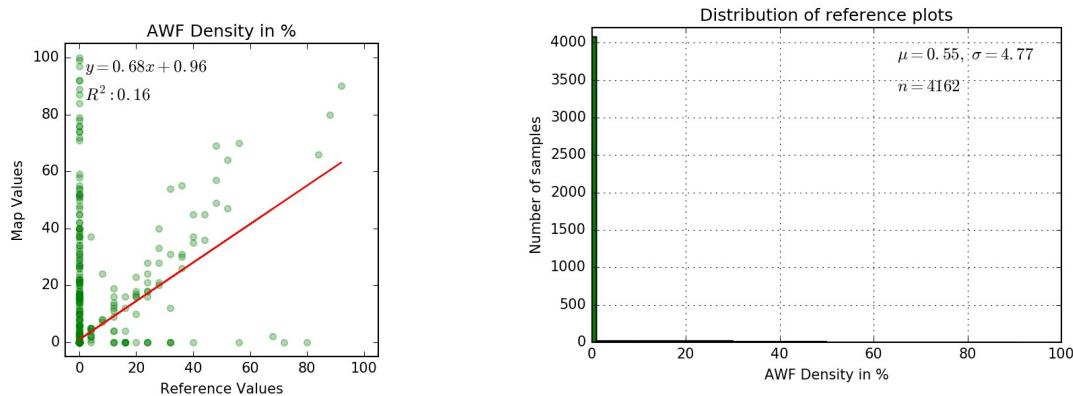


Figure 99: AWF Layer – TR country zone: (left) Scatterplot of all sample units. (Right) Distribution of density values.

Annex 5. SWF2015 - confusion matrices

Table 7.1: Estimated area proportion matrix for AWF mapping on pan-european level. Columns correspond to reference and rows to mapping.

| blind | AWF < 1 | AWF >= 1 | UA % | EC % | OA % |
|----------|----------------|-----------------|-----------------|-------|-----------------|
| AWF < 1 | 88.79 | 1.44 | 98.4 (+/-0.17) | 1.6 | |
| AWF >= 1 | 3.93 | 5.83 | 59.73 (+/-1.96) | 40.27 | |
| PA % | 95.76 (+/-0.2) | 80.16 (+/-1.75) | | | |
| EO % | 4.24 | 19.84 | | | |
| OA % | | | | | 94.62 (+/-0.24) |

Table 7.2: Raw error matrix (counts) for AWF mapping on pan-european level. Columns correspond to reference and rows to mapping.

| blind | AWF < 1 | AWF >= 1 | UA % | EC % | OA % |
|----------|-----------------|-----------------|-----------------|-------|-----------------|
| AWF < 1 | 21822 | 355 | 98.4 (+/-0.17) | 1.6 | |
| AWF >= 1 | 967 | 1434 | 59.73 (+/-1.96) | 40.27 | |
| PA % | 95.76 (+/-0.56) | 80.16 (+/-2.36) | | | |
| EO % | 4.24 | 19.84 | | | |
| OA % | | | | | 94.62 (+/-0.98) |

Table 7.3: Estimated area proportion matrix for SWF mapping on pan-european level. Columns correspond to reference and rows to mapping.

| blind | SWF < 1 | SWF >= 1 | UA % | EC % | OA % |
|----------|-----------------|-----------------|-----------------|-------|----------------|
| SWF < 1 | 71.31 | 5.09 | 93.34 (+/-0.36) | 6.66 | |
| SWF >= 1 | 5.61 | 18 | 76.24 (+/-1.1) | 23.76 | |
| PA % | 92.71 (+/-0.31) | 77.97 (+/-0.95) | | | |
| EO % | 7.29 | 22.03 | | | |
| OA % | | | | | 89.3 (+/-0.38) |

Table 7.4: Raw error matrix (counts) for SWF mapping on pan-european level. Columns correspond to reference and rows to mapping.

| blind | SWF < 1 | SWF >= 1 | UA % | EC % | OA % |
|----------|-----------------|-----------------|-----------------|-------|----------------|
| SWF < 1 | 17525 | 1250 | 93.34 (+/-0.36) | 6.66 | |
| SWF >= 1 | 1379 | 4424 | 76.24 (+/-1.1) | 23.76 | |
| PA % | 92.71 (+/-0.49) | 77.97 (+/-0.94) | | | |
| EO % | 7.29 | 22.03 | | | |
| OA % | | | | | 89.3 (+/-0.58) |