

# 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 IMPERVIOUSNESS DEGREE 2018 VALIDATION REPORT



## Consortium Partners

**Prime Contractor** SIRS SAS, France

**Main Partners** GAF AG, Germany

Joanneum Research, Austria

e-GEOS, Italy

**Experts** Specto Natura, UK

IGN France International, France

Terranea, Germany

NLSI, Iceland

## Document Preparation and Release

	Affiliation	Name(s)	Date	Signature
<b>Author</b>	Specto Natura	Geoff Smith	31/07/2020	
<b>Contributions</b>	SIRS	Alexandre Pennec		
<b>Review</b>	SIRS	Alexandre Pennec		
<b>Approval</b>	EEA	Hans Dufourmont		

## Document Issue Record

Issue	Date	Author(s)	Description of Change
0.1	01/07/2020	Geoff Smith	Initial draft version
0.2	15/07/2020	Geoff Smith	Inclusion of IMD blind and plausibility results
1.0	27/07/2020	Alexandre Pennec	Review and integration of CI95% results
1.1	31/07/2020	Geoff Smith	Inclusion of SBU analysis
1.2	31/07/2020	Geoff Smith / Alexandre Pennec	Inclusion of IMC analysis
1.3	24/11/2020	Geoff Smith / Alexandre Pennec	Integration of feedbacks from EEA and SPs

## Executive Summary

This report covers the validation of the High Resolution layers (HRL) on Imperviousness (IMD), Share of Built-Up (SBU) and Imperviousness change (IMC) for the 2018 update.

The IMD layer captures the spatial distribution of artificially sealed areas, including the level of sealing of the soil per area unit for a particular reference year. The level of sealed soil (imperviousness degree 1-100%) is produced using an automatic algorithm based on calibrated normalised difference vegetation index (NDVI). To date the IMD layer has been produced for five reference years; 2006, 2009, 2012, 2015 and 2018. For each of the iterations from 2009 onwards the results have contained two products: a status layer for any reference year (e.g. IMD2018), as well as an imperviousness density change layer between reference years (e.g. IMC2018 is 2015 to 2018), and based on the already existing imperviousness product for that previous reference year. For the 2018 reference year a further product representing the Share of Built-Up (SBU) was also produced.

This report represents the analysis of the fully aggregated 100 m spatial resolution IMD, SBU and IMC product deliveries for the 2018 reference year. The assessment of the IMD, SBU and IMC layers involved a review of the available datasets and the existing documentation prepared as part of the semantic checks performed during the production. The datasets and documented checks were compared against the specification prepared by the EEA and published in June 2020, as well as the original tender documents. This was a qualitative assessment reported in a pass / fail data quality list with additional comments from the validation team. A comprehensive assessment of the blind interpretation and a plausibility analysis are given to take into account input data limitations and better understand the cause of classification errors. These assessments included a quantitative analysis of the mapped imperviousness density, share of built-up and change against reference data in a stratified systematic sampling scheme.

The results in this report show that the IMD product broadly meets most of the technical specifications, but the minimum thematic accuracy requirement of 90 %, based on a 30 % threshold of IMD values to represent urban areas, are not met in all cases. The IMD layer tends to underestimate imperviousness and there is variability from country to country or based on biogeographical regions. Improved results were obtained when using a plausibility approach which allowed for some variation in the reference data to account for acceptable mapped imperviousness. The 2018 show a more balanced distribution between errors of omission and commission. The outcome of the analysis of the scatterplots show that there is a strong relationship between the reference and map data. The variation between the sub-European regionalisations is a useful tool to explore the causes of the lower than anticipated thematic accuracies. Overall, there has been an improvement in the performance of the IMD 2018 layer against 2015 results for both blind and plausibility interpretations relative to reference data. The main reason of the improvement of user's and producer's accuracies is related to the enhanced spatial resolution from 2015 to 2018. Indeed, the 2015 production was mainly based on Landsat-8 data with a spatial resolution of 30m, whereas for 2018 Sentinel-2 data with a spatial resolution of 10m was used as main input. It results in a 9-fold improvement in spatial resolution which leads to a greater detection of sealed surfaces.

The results in this report show that the SBU product broadly meets most of the technical specifications, but the minimum thematic accuracy requirement of 90 %, based on a 1 % threshold of SBU values to represent built-up areas, are not met in all cases. The use of a 30 % threshold of SBU values produced poorer thematic results. The SBU layer tends to overestimate the share of built-up and it is highly variable across regionalisations. Improved results were obtained when using a plausibility approach which allowed for some variation in the reference data to account for acceptable mapped share of built-up. As this was the first reference year for the production of this layer there were no previous results against which to assess improvements, but further work should be considered to understand the definitions and their realisation in products and reference data.

The results in this report show that the IMC product again meets most of the technical specifications, but the minimum thematic accuracy requirement of 90 % are not met for the whole dataset and not for all regionalisations. The results are broadly similar to those of the 2015 layer, but there has been an improvement in the producer's accuracy at the expense of user's accuracy.

## Table of Contents

.....	0
<b>1. Validation Framework.....</b>	<b>8</b>
1.1. Products to be validated.....	8
1.2. Validation Criteria.....	9
1.2.1. IMD2018.....	10
1.2.2. SBU2018 .....	10
1.2.3. IMC1518 .....	10
<b>2. Validation approach .....</b>	<b>11</b>
2.1. Completeness .....	11
2.2. Logical consistency .....	12
2.2.1. Conceptual consistency.....	12
2.2.2. Domain consistency .....	12
2.2.3. Format consistency .....	13
2.2.4. Additional logical consistency checks .....	14
2.3. Positional Accuracy.....	15
2.4. Thematic Accuracy.....	15
2.4.1. Level of reporting .....	15
2.4.2. Stratification and sample design.....	16
2.4.3. Response Design.....	22
2.4.4. Estimation and analyses procedures.....	24
2.5. Temporal Quality .....	25
2.6. Usability .....	25
2.7. INSPIRE compliant metadata .....	26
<b>3. Validation check list.....</b>	<b>27</b>
3.1. IMD2018 .....	27
3.2. SBU2018.....	33
3.3. IMC1518.....	39
<b>4. Thematic accuracy IMD2018.....</b>	<b>45</b>
4.1. IMD2018 scatterplots & regression analysis .....	45
4.2. IMD2018 blind interpretation correspondence analysis .....	52
4.3. IMD2018 plausibility analysis correspondence analysis .....	54
<b>5. Thematic Accuracy SBU2018.....</b>	<b>56</b>
5.1. SBU2018 scatterplots & regression analysis .....	56
5.2. SBU2018 blind and plausibility interpretation correspondence analysis .....	62

6.	Thematic Accuracy IMC1518.....	65
7.	Conclusions and recommendations .....	67

## List of Figures

<i>Figure 2-1. Level of reporting accordingly to the biogeographical regions .....</i>	15
<i>Figure 2-2. Level of reporting accordingly to the country or aggregated countries .....</i>	16
<i>Figure 2-3. Simple random (left) and random systematic (right) sampling designs.....</i>	17
<i>Figure 2-4. LUCAS points located on a regular grid .....</i>	17
<i>Figure 2-5. Number of sample points as a function of the expected error rate for two accepted standard error values (after Wack et al. 2012).....</i>	18
<i>Figure 2-6. Replicates and sub-replicates used on LUCAS grid .....</i>	19
<i>Figure 2-7: Comparison HRL 2015 (red) vs CLC 2012 (blue) and OSM (green).....</i>	20
<i>Figure 2-8: Example of SSUs organised in a 5x5 20m grid .....</i>	21
<i>Figure 2-9: An example of the PSU (yellow square) and the SSU (green and red dots) for one sample location based on a LUCAS point. ....</i>	23
<i>Figure 2-10. Representation of the behaviour of the 95% confidence interval for a 5x5 SSU grid over the whole range of imperviousness degree values .....</i>	25
<i>Figure 4-1: Scatterplots for 2018 for ALL the reference data for IMD for both the blind interpretation (TOP) and plausibility analysis (BOTTOM). ....</i>	46
<i>Figure 4-2: Scatterplots for 2018 for Black Sea (BLS) BRME zone for IMD for both the blind interpretation (TOP) and plausibility analysis (BOTTOM). ....</i>	47
<i>Figure 4-3: Scatterplots for 2018 for Atlantic (ATL) BRME zone for IMD for both the blind interpretation (TOP) and plausibility analysis (BOTTOM). ....</i>	48
<i>Figure 5-1: Scatterplots for 2018 for ALL the reference data for SBU for both the blind interpretation (TOP) and plausibility analysis (BOTTOM). ....</i>	57
<i>Figure 5-2: Scatterplots for 2018 for BLS the reference data for SBU for both the blind interpretation (TOP) and plausibility analysis (BOTTOM). ....</i>	58
<i>Figure 5-3: Scatterplots for 2018 for ATL the reference data for SBU for both the blind interpretation (TOP) and plausibility analysis (BOTTOM). ....</i>	59

## List of Tables

<i>Table 1-1: Thresholds used for the correspondence analysis of IMC1518.....</i>	10
<i>Table 2-1: Map projection details for LAEA-ERTS89.....</i>	14
<i>Table 2-2: Distribution of sample units per main strata and substrata.....</i>	21
<i>Table 2-3: Weight factor to be applied to each stratum and substratum for constructing confusion matrices .....</i>	22
<i>Table 4-1: A summary of regression line parameters for the scatterplots derived from the blind interpretation as reference data for 2015 and 2018. ....</i>	50
<i>Table 4-2: A summary of regression line parameters for the scatterplots derived from the plausibility analysis as reference data. ....</i>	51
<i>Table 4-3: A summary of the correspondence results for the blind interpretation for the 2015 and 2018 results. Green cells exceed the 90 % minimum accuracy requirement, orange cells are within 5% and red cells fall more than 5% short.....</i>	53
<i>Table 4-4: A summary of the correspondence results for the plausibility analysis for the 2015 and 2018 results. Green cells exceed the 90 % minimum accuracy requirement, orange cells are within 5% and red cells fall more than 5% short.....</i>	55
<i>Table 5-1: A summary of regression line parameters for the SBU scatterplots derived from the blind and plausibility interpretation as reference data for 2018. ....</i>	61
<i>Table 5-2: A summary of the correspondence results for the blind and plausibility interpretation for the SBU &gt; 1% results. Green cells exceed the 90 % minimum accuracy requirement, orange cells are within 5% and red cells fall more than 5% short.....</i>	63
<i>Table 5-3: A summary of the correspondence results for the blind and plausibility interpretation for the SBU &gt; 30% results. Green cells exceed the 90 % minimum accuracy requirement, orange cells are within 5% and red cells fall more than 5% short.....</i>	64
<i>Table 6-1: A summary of the correspondence results for the blind and plausibility interpretation for the IMC2018 results. Green cells exceed the 90 % minimum accuracy requirement, orange cells are within 5% and red cells fall more than 5% short.....</i>	66

## List of Abbreviations

BRME	Biogeographical Regions Map of Europe
CLC	CORINE Land Cover
CRS	Coordinate reference system
DWH	Data warehouse
EEA	European Environment Agency
EO	Earth Observation
ETC ULS	European Topic Centre on Urban, Land and Soil systems
ETRS	European Terrestrial Reference System
EUREF	Regional Reference Frame Sub-Commission for Europe
FP7	7 <sup>th</sup> Framework Programme
FTS LM	Fast Track Service Land Monitoring
GDAL	Geospatial Data Abstraction Library
GIO	GMES Initial Operations
GMES	Global Monitoring for Environment and Security
GRS	Geodetic Reference System
HRL	High Resolution Layer
IMD	Imperviousness density
IMC	Imperviousness density change
INSPIRE	Infrastructure for Spatial Information in the European Community
LAEA	Lambert Azimuthal Equal-Area
LUCAS	Land Use/Cover Area frame Survey
MMU	Minimum Mapping Unit
MMW	Minimum Mapping Width
NDVI	Normalised Difference Vegetation Index
PSU	Primary Sample Unit
SBU	Share of Built-Up
SSU	Secondary Sample Unit
TIFF	Tagged Image File Format

## 1. Validation Framework

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

### 1.1. Products to be validated

Pan-European High Resolution Layers (HRL) provide information on specific land cover characteristics, and are complementary to land cover / land use mapping such as in the CORINE land cover (CLC) datasets. The HRLs were originally produced at a 20 m spatial resolution, but for the 2018 versions this has been improved to 10 m spatial resolution as all imagery could be sourced through the Copernicus Sentinels (1 & 2). Their thematic content is generated through a combination of automatic processing and interactive rule-based classification.

5 themes have been identified so far, four corresponding with the main themes from CLC, i.e. the level of sealed soil (imperviousness), tree cover density & forest type, permanent grasslands, water & wetness, and one related to small woody features. There has been some evolution of the layers over time. For the 2015 iteration the HRL for wetland and water layers were merged into a single layer which recorded wetness and water. Also, the natural grassland layer was also redefined for 2015 to become a new grassland baseline product based on 7-year time series and including all grasslands. The 2015 iteration also saw a new product, based on VHR data, related to small patchy and linear woody features.

Although produced and available for use at a high spatial resolution of 10 m (20 m), their constituent pixels were aggregated into 100 by 100 m grid cells for final products which were validated here.

The HRLs are pan-European wall to wall products which cover the EEA39 countries (Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey and the United Kingdom plus Albania, Bosnia and Herzegovina, Kosovo, The former Yugoslav Republic of Macedonia, Montenegro and Serbia.).

The HRLs were produced in a centralized fashion involving the service industry through competitive market mechanisms on a layer by layer basis. In a decentralized approach the participating countries in the “Copernicus Local Land monitoring services: NRCs LC Copernicus supporting activities for the period 2017-2021” contracts will be performing a further verification of the HRLs if they wished, but that work will not be completed before the end of this contract.

Built-up areas are characterized by the substitution of the original (semi-) natural land cover or water surface with an artificial, often impervious cover. These artificial surfaces are usually maintained over long periods of time. The Imperviousness HRL (IMD) thus captures the spatial distribution of artificially sealed areas, including the level of sealing of the soil per area unit. The level of sealed soil (imperviousness degree 1-100%) is produced using an automatic algorithm based on calibrated normalised difference vegetation index (NDVI).

The IMD layer was the first HRL to be produced for 2006. However, at the time it was described as a “soil sealing database for Europe” and delivered as part of the GMES (Global Monitoring for Environment and Security, former name of Copernicus) Fast Track Service on Land Monitoring (Land FTS LM). It was produced during 2006-2008 from multi-sensor and bi-temporal, orthorectified satellite imagery, IMAGE2006, the same as the CORINE Land Cover 2006 update. The production of IMD2006 covering 38 European countries (32 EEA Member States and 6 West-Balkan countries) and was implemented in two phases:

- an initial soil sealing product based on the EEA specification, and
- a soil sealing enhancement product based on evaluation of the initial product by some Member States.

The production approach used an automatic algorithm based on calibrated NDVI. The main deliverable was a raster dataset of continuous degree of soil sealing ranging from 0 – 100 % in full spatial resolution (20 m x 20 m)

with the associated metadata. A derived product, a raster dataset of continuous degree of soil sealing ranging from 0 - 100% in aggregated spatial resolution (100 m x 100 m) in European projection was validated.

Since the production of IMD2006 a time series of imperviousness has been produced for reference years 2009, under the FP7 Geoland2 project, 2012, under the GMES Initial Operations (GIO) and 2015 under the Copernicus programme. For each of these iterations the results contained two products: a status layer for any reference year (e.g. IMD2012), as well as an imperviousness density change layer between reference years (e.g. IMC2009to2012) based on the already existing imperviousness product for that previous reference year. It is worth noting that there have been revisions to the previous year's IMD products during updating and the IMC time series has been reworked.

The latest imperviousness update for 2018 was also produced as part of the Copernicus Land Monitoring Services and was available with free and open access (see <http://land.copernicus.eu/>).

When considering all the IMD products a density threshold of 30% has been used to derive binary sealed mask layer from the imperviousness values. This was not intended to be a separate product, but instead was calculated for the verification / validation process only, because density products cannot be verified.

For 2015, all the previous IMD and IMC layers were reworked. The derivation of Imperviousness 2006-2009-2012 consisted of two separate procedures, the reprocessing and the re-analysis of the historical layers. The methodologies applied for both the reprocessing, and re-analysis of all existing density products, assured a properly calibrated HRL Imperviousness time-series.

For 2018, new IMD and IMC layers were produced, but there was no reprocessing of the previous data. The spatial resolution of the primary layers was reduced to 10 m thanks to the Sentinel-2 imagery, but the validation procedure was still applied to the aggregated 100 m versions. Issues around the change from 20 m – 100 m to 10 m – 100 m aggregation were considered.

An additional product related to the share of built-up was also produced at the 100 m level. An automated supervised classification was applied to level 2 Sentinel-2 10 m imagery and derived textural and spectral indicators. The Approach was trained by selection built-up and non-built-up area from Open Street Map and the European Settlement Map. The results were post-processed using the IMD2018 layer and NDVI values to apply rule-based corrections.

The HRL products to be validated as integrated Pan-European mosaics at a 100m pixel size were as follows:

- Degree of Imperviousness (IMD) 2018, 100m x 100m, European projection, LAEA
- Share of Built Up (SBU) 2018, 100m x 100m, European projection, LAEA
- Imperviousness Density Change (IMC) between 2015 & 2018, 100m x 100m, European projection, LAEA

## 1.2. Validation Criteria

For the IMD, SBU and IMC layers the main criteria selected for the validation will be:

- Completeness, the amount of omission and commission.
- Logical consistency, the adherence to formats, conventions and conceptual aspects.
- Thematic accuracy, the correspondence with reference data.
- Temporal quality, the alignment of the results with the reference year.
- Usability
- Metadata, the presence of sufficient metadata to describe the product.

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

### 1.2.1. IMD2018

The validation of the Imperviousness Degree Layer for 2018 was done at two levels:

- A scatterplot of the density values extracted from the sample units for both the reference and map data for 2018 for the blind and plausibility interpretation was 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 was applied to the density values for reference and map data to produce binary attributes of built-up for both the reference and map data layers for the blind and plausibility interpretation. For IMD, the threshold was set to 30 % with density values lower than 30 % classified as 0 (non-built-up) and density values greater than or equal to 30 % classified as 1 (built-up). The minimum acceptable thematic accuracy of 90 % should be reached for both omission and commission errors for class 1 (built-up).

### 1.2.2. SBU2018

The validation of the Share of Built Up Layer for 2018 was done at two levels:

- A scatterplot of the percentage share values extracted from the sample units for both the reference and map data for 2018 for the blind and plausibility interpretation was 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).
- Two thresholds were applied to the percentage share values for reference and map data to produce binary attributes of built-up share for both the reference and map data layers for the blind and plausibility interpretation. For SBU, the thresholds were set to 1 % and 30 % with density values lower than the threshold classified as 0 (non-built-up) and density values greater than or equal to the threshold classified as 1 (built-up). The minimum acceptable thematic accuracy of 90 % should be reached for both omission and commission errors for class 1 (built-up).

### 1.2.3. IMC1518

The validation of the Imperviousness Degree Change Layer for 2015 involved a classification of both the reference and map data changes for the blind and plausibility interpretation into six classes (Table 1-1). For the actual thematic accuracy calculation, the changes were further generalised into 3 classes of no change (unsealed), change in imperviousness and no change (sealed).

Table 1-1: Thresholds used for the correspondence analysis of IMC1518.

Class code	Description
0	Stable unsealed areas
1	New sealed areas
2	Loss of sealed area
10	Stable sealed areas
11	Increase in imperviousness
12	Decrease in imperviousness

## 2. Validation approach

The validation approach provides 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, (iv) Thematic Accuracy, (v) Temporal quality and (vi) Usability. Each of these criteria forms a section in the Validation Check list in section 3.

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

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

The validation exercise will refer to the technical details of the IMD, SBU and IMC products at 100m spatial resolution as set out in "Copernicus Land Monitoring Service – High Resolution Layer 2018 Production: Key technical specifications Lots 1-4: VERSION 1.3", Creation date "2020-06-03" and filename "Key\_technical\_specs\_HRL2018\_lots\_1-4\_V1-3.docx". This document captures the detailed definitions and product specifications for the Copernicus Land Monitoring Service (CLMS) Imperviousness High Resolution Layer (HRL) products for the 2018 reference year.

### 2.1. 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 areas. 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:** For the completeness of the IMD / SBU / IMC the proportion of excess data beyond the required extent is used for commission errors and the proportion of missing data within the required extent is used to represent the omission errors.

However, the total extent referred to is not clearly defined as land area or whether there are territorial buffers that reach out to sea when finalising products. For delineation of the production area (EEA39) the Countries, 2020 - Administrative Units – Dataset (release date 28/02/2020) from EuroStat (<https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/countries>) was used to check all products.

This check was performed by visually checking the extent of the product against the known extents of the EEA-39 countries.

## 2.2. Logical consistency

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

### 2.2.1. 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 IMD, SBU and IMC layers the properties of conceptual consistency are defined for both the 10 m and 100 m spatial resolution products, with the inference that some properties are applied in the same way to both. The properties appropriate to raster data can be defined as:

- Type of feature used. In the case of raster dataset, the features are pixels. The raster bit depth is defined to encompass the required number of numeric values / thematic codes and unsigned 8-bit and signed 16-bit are valid for the HRLs. For the IMD, SBU and IMC the raster format is proposed but no specific depth is set. However, the range of required numeric values / thematic codes would suggest unsigned 8-bit data is appropriate for IMD and SBU and signed 16-bit is appropriate for IMC.
- 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 IMD, SBU and IMC rasters 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 IMD, SBU and IMC products should use the European LAEA.

### 2.2.2. Domain consistency

**Description:** 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 IMD, SBU and IMC, the correct encoding of data is checked.

**Indicator:** The domain consistency properties are divided into two main groups. The overall data structure of the raster file and the valid values for the pixels.

The data structure for the HRLs should be a single band raster of type unsigned 8-bit or a 16-bit signed.

The valid values for the pixels will be different between the IMD, SBU and IMC datasets as they are representing different characteristics of the surface.

IMD:

- 0: all non-impervious areas
- 1-100: imperviousness values
- 254: unclassifiable (no satellite image available, or clouds, shadows, or snow)
- 255: outside area

SBU:

- 0 to 100: share of built-up (%)
- 254: unclassifiable (no satellite image available, or clouds, shadows, or snow)
- 255: outside area

IMC:

- 0 to 99: decreased imperviousness density
- 100: sealed in both years (stable built-up)
- 101-200: increased imperviousness density
- 201: non-sealed in both years (stable non-built up)
- 254: unclassifiable (no satellite image available, or clouds, shadows, or snow)
- 255: outside area

### 2.2.3. Format consistency

**Description:** Format consistency includes the consideration of file formats, file or attribute names or attribute types. 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 IMD, SBU and IMC layers which are raster-based products plus documentation, only the following format consistency properties need to be checked:

- File format conformance
  - Raster products shall be delivered as Geotiff (\*.tif) with world file (\*.tfw), pyramids (\*.ovr), attribute table (\*.dbf) and statistics (\*.aux.xml).
  - Compression: LZW
  - Each product shall be accompanied with INSPIRE compliant metadata in XML format and an INSPIRE Mapping Table as XLS files.
  - A PDF providing CRS information, including details of parameters used to transform to ETRS89 LAEA projection as in the following example from Hungary. The pdf should be named as follows: CRS\_Information\_Sheet\_<country 2-letter ISO code>, e.g. CRS\_Information\_Sheet\_BG.pdf. See Table 6 in specification document.
  - CRS information sheets will be static and therefore will not have version numbers.

- File name conformance

A set of filename conventions were established in the product specifications, but these referred only to the country level products and associated reports. The conventions are based on the following descriptors:

- 'theme' – the particular HRL as a 3-letter abbreviation of processing stage, e.g. 'imd' for Imperviousness Degree, 'sbu' for Share of Built-Up and 'imc' for Imperviousness Change.
- 'year' – four digits e.g. 2012 for the IMD and SBU layers and 1518 for the IMC layer.
- 'resolution' – a four-character abbreviation (100m).
- 'extent' - either a 2-letter country code for country deliveries in national projection or "eu" for all deliveries in European Projection (partial and full lot mosaics).
- 'EPSG' – a 5-digit EPSG projection code (geodetic parameter dataset code by the European Petroleum Survey Group), e.g. "03035" for the European LAEA projection.
- 'version' (only for final deliveries) – 4 character qualifier of the version number, starting with "V1\_1" for a first full final version, and allowing to capture re-processing/calculation of small changes as ("V1\_2", "V1\_3" etc.). In case of major changes a second version should be used ("V2\_1").

The file name format should therefore be:

`<theme>_<year>_<resolution>_<extent>_<EPSG>_<version>`

As an example, the imperviousness layer for 2018 in LAEA projection and 100 m spatial resolution should be supplied as follows:

`imd_2018_100m_eu_03035_V1_0.tif`

## 2.2.4. 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 2-1).

*Table 2-1: 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 1980 (New International)
	Semi major axis	6 378 137 m
	Inverse flattening	298.257222101
Projection		Geographic (Ellipsoidal Coordinate System)

## 2.3. Positional Accuracy

The positional accuracy of the products is being addressed through the source EO image data and will only be considered here if unusual issues arise.

## 2.4. Thematic Accuracy

The thematic accuracy quantifies the relationship of the product to a set of reference data. The reference data can come from many sources and could be a binary mask, categorical set or continuous variables.

The thematic accuracy requirement is set relative to an urban area mask. A threshold of 30 % should be applied to transform imperviousness to built-up. A minimum thematic accuracy requirement was set at 90 % overall accuracy target, as well as 90 % for producer's accuracy (commission) and 90% for user's accuracy (omission).

### 2.4.1. Level of reporting

Regarding the level of reporting for the internal validation, results will be provided at different levels of aggregation. So, the accuracy of the products will be reported at EEA39 level and at more disaggregated scales. 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.
2. Biogeographical regions 2016 (Figure 2-1).

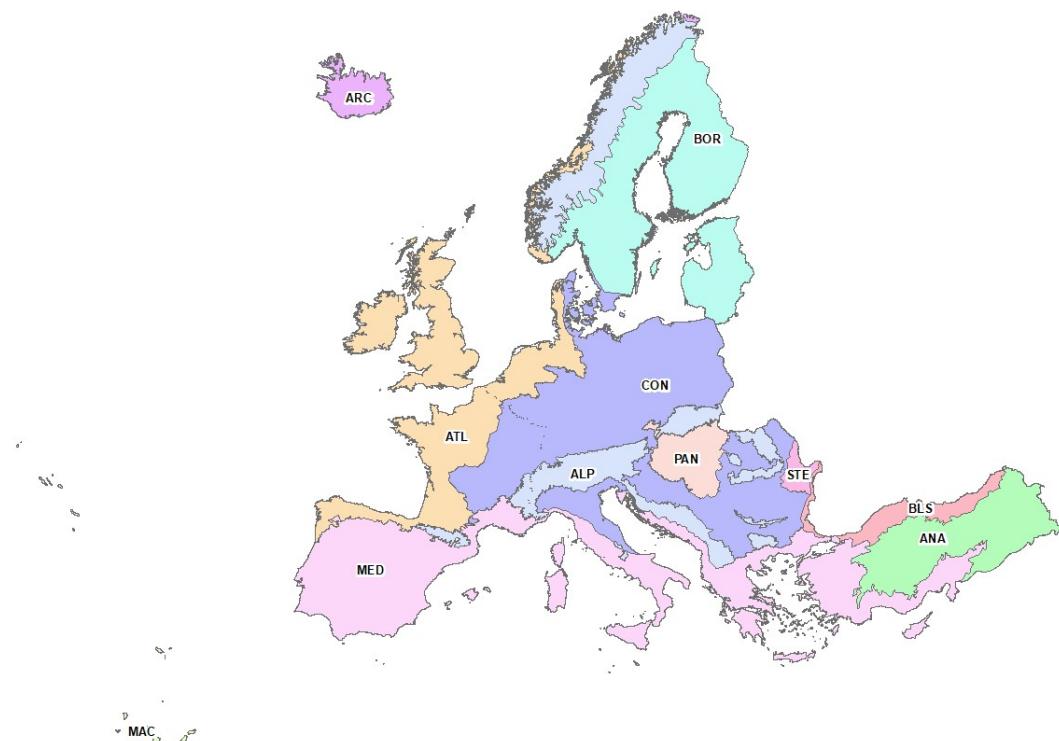


Figure 2-1. Level of reporting accordingly to the biogeographical regions

3. Country or aggregated countries as described in the previous Specific Contracts in 23 main countries or groups of countries (including French DOMs). Countries < 90,000 km<sup>2</sup> shall be grouped into contiguous groups of countries > 90,000km<sup>2</sup> as much as possible (Figure 2-2).



Figure 2-2. Level of reporting accordingly to the country or aggregated countries

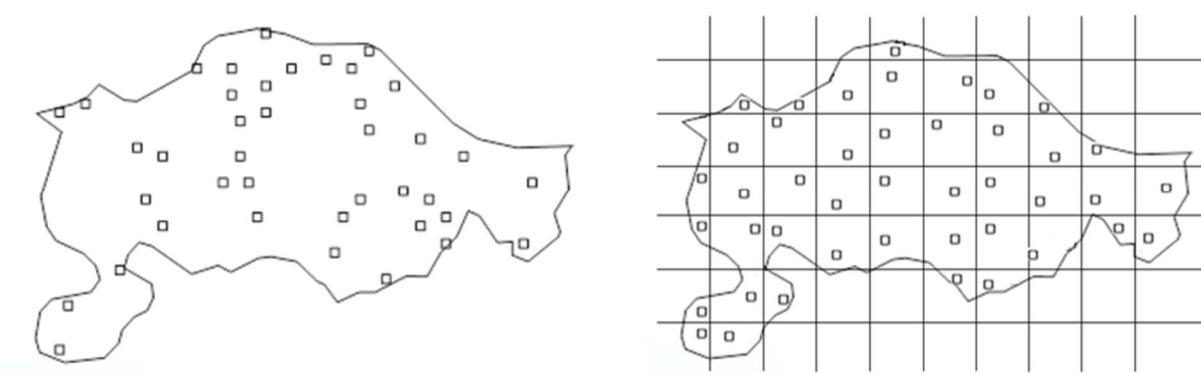
## 2.4.2. Stratification and sample design

### 2.4.2.1. Overview

The stratification and the sampling design primarily consist of 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.

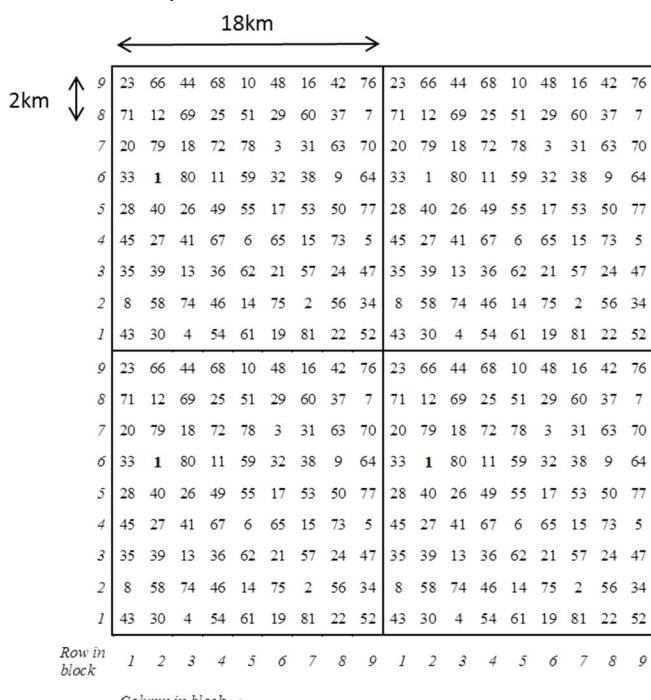


*Figure 2-3. Simple random (left) and random systematic (right) sampling designs*

The sampling and stratification design presented below is applicable to all HRL Imperviousness products including change products.

The validation approach of the previous Specific Contracts was adopted. 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. 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 2-4. 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 proper 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  $\sigma_h$ :  $n_h = \frac{p_h(1-p_h)}{\sigma_h^2}$ . (2)

From Figure 2-5 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.

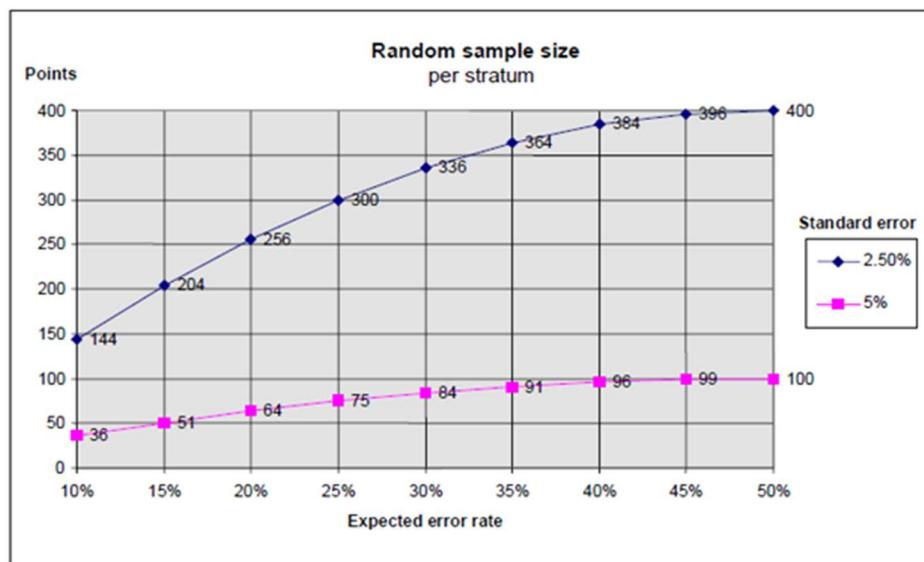


Figure 2-5. Number of sample points as a function of the expected error rate for two accepted standard error values (after Wack *et al.* 2012)

When using stratified sampling, the main issue to maximise the efficiency of the stratification (maximise 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  $\sigma_h$  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's accuracy.

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 the Figure below.

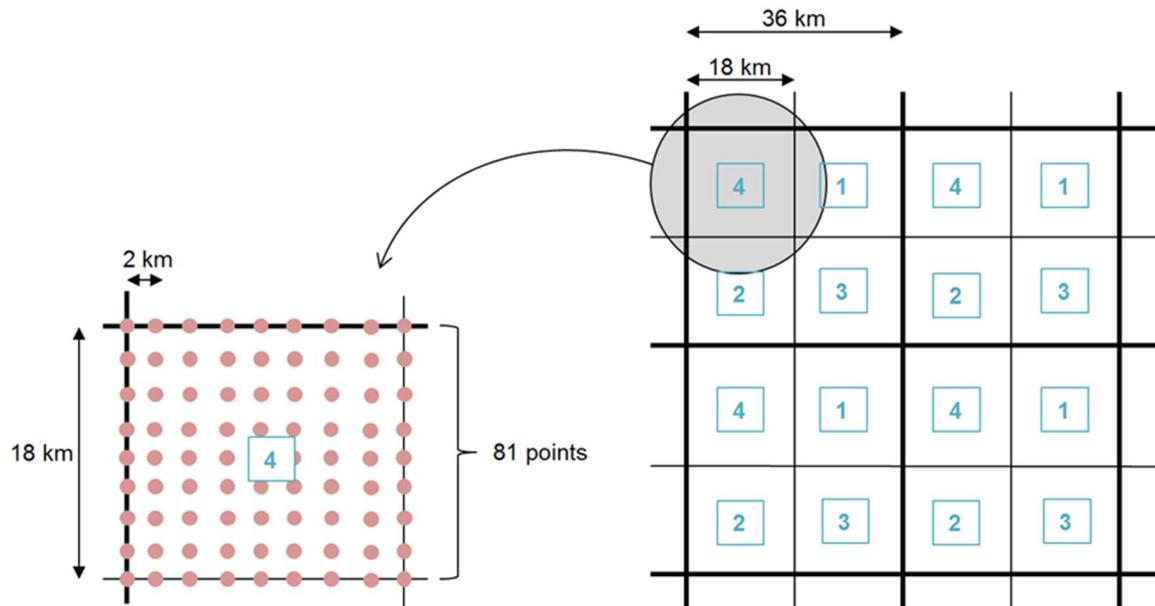


Figure 2-6. Replicates and sub-replicates used on LUCAS grid

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 200 m.

#### 2.4.2.2. Stratification approach

For the IMD/IMC products, a stratification is applied based on a series of omission/commission strata at pan-European level.

The number of sample units called **Primary Sampling Units (PSUs)** per stratum based on LUCAS and densified LUCAS grid should be such to ensure a sufficient level of precision at reporting level. Priority is given to strata which are known to be difficult to map: i.e. changes and difficult classes. That is why different sampling intensities can be applied to focus on strata for which there is a higher probability that errors will be found.

The level stratification based on omission/commission strata for the Imperviousness Status and Change layers is defined as follows:

- Commission: Imperviousness Degree 1-100% in 2015
- Omission High Probability: Imperviousness Degree 0% in 2015 and Open Street Map & CLC “impervious classes”
- Omission Low Probability: Rest of the area
- Commission Change: all changes 2006-2009, 2009-2012, 2012-2015 and 2015-2018 [gain, loss, increased and decreased]

It should be noticed that the stratification is based on the HRL IMD 2015 product, performed during the Third Specific Contract. Since the IMD product for the reference year 2018 was not available at the beginning of the validation exercise, this approach allowed us to take advantage of the sample units already drawn and interpreted for the reference year 2015. Nevertheless, the sample units have been reinterpreted/updated for the reference year 2018 following the specifications of the product. This approach was considered as valid due

to the relatively low urban dynamics over the 3-year time period. It also explains the use of CLC 2012 instead of CLC 2018.

The selection of the PSUs in these strata will also take into account the sample units drawn for the omission and commission strata to avoid duplicates.

For both status and change layers, CLC artificial classes and Open Street Map road network are used and converted to a pseudo artificial layer. Relevant OSM road types are selected and rasterized to 100m (for example, abandoned, construction, cycleway, path, planned, trail, track... are removed) to obtain the artificial areas. Using a relevant selection of OSM road types tend to lead to a better spatialization of artificial and impervious areas.

CLC impervious classes are defined as follows based on CLC2012:

- 1.1.1 = continuous urban fabric
- 1.1.2 = discontinuous urban fabric
- 1.2.1 = industrial, commercial areas
- 1.2.3 = ports
- 1.2.4 = airports

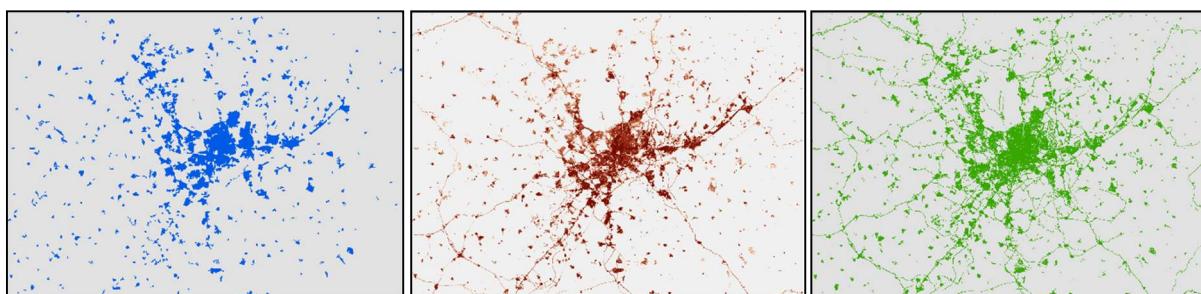


Figure 2-7: Comparison HRL 2015 (red) vs CLC 2012 (blue) and OSM (green)

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

Each PSU corresponds to one HRL aggregated pixel (100m). Each PSU is then associated to secondary sampling units (SSUs) corresponding to a 5x5 grid with 20m between each SSU (Figure 2-8). The original idea is that each SSU can then be associated with the corresponding HRL 20m layer pixel. It should be noticed that the HRL 2018 IMD product is now based on a 10-meter spatial resolution using the Sentinel-2 data and so, the SSUs cannot be associated to one HRL 10m pixel anymore. This approach was still considered suitable both for assessing the 2018 imperviousness density status values as well as the change values. Each PSU will be interpreted based on VHR image data for each reference date (2015 and 2018) and change values will then be derived from the actual change layers allowing the assessment of the accuracy.



*Figure 2-8: Example of SSUs organised in a 5x5 20m grid*

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

There was a total of 25,777 sample units selected for this specific contract (Table 2-2):

*Table 2-2: Distribution of sample units per main strata and substrata*

	Strata	Number of sample units
Commission	Status 2015	3,959
	Change 0609	1,590
	Change 0912	1,579
	Change 1215	1,575
	Change 0609 & Change 0912	3
	Change 0609 & Change 1215	7
	Change 0912 & Change 1215	18
	Change 1518	3,000
Omission	High Probability	1,019
	Low Probability	13,027
<b>Total</b>		<b>25,777</b>

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 2-3) following the formula presented at the beginning of section 2.4.4 .

Table 2-3: Weight factor to be applied to each stratum and substratum for constructing confusion matrices

	Strata	Number of sample units
Commission	<b>Status 2015</b>	0.4149424
	<b>Change 0609</b>	0.0069280
	<b>Change 0912</b>	0.0086511
	<b>Change 1215</b>	0.0053778
	<b>Change 0609 &amp; Change 0912</b>	0.0102973
	<b>Change 0609 &amp; Change 1215</b>	0.0066540
	<b>Change 0912 &amp; Change 1215</b>	0.0063842
	<b>Change 1518</b>	0.0614105
Omission	<b>High Probability</b>	0.4147891
	<b>Low Probability</b>	1.8034853

The sample units were provided to the bulk interpretation team as separate shapefiles in which all the information on strata was removed to ensure the independence of the interpretation.

### 2.4.3. Response Design

#### 2.4.3.1. Overview

The LUCAS points were re-interpreted based on available in situ and reference data, but the LUCAS thematic information is not used directly during the process.

The response design for most data sets being validated during this project was based on the interpretation of thematic classes or surface characteristics (e.g. soil sealing) at the SSU point level taking into account the details of the product specifications (MMU, MMW, class definitions, etc. ...). The interpretation was based on a combination of available in situ data, for example, the HRL and CLC2012 validation could use the VHR mosaic alongside the imagery used in production and virtual globes datasets.

A double blind approach was adopted as the initial process to guarantee the complete independence of the validation data from the map products. This may underestimate their accuracy where SSU points are uncertain. This was resolved by the plausibility approach for which the interpreter checks the map value to assess whether it can be considered correct or not, within the frame of accepted product specifications. Also, density values may be adjusted based on experience of known uncertainties to allow a more realistic comparison.

The interpretation of each PSU at a particular point in time is based on the assessment (sealed or unsealed) of 5 x 5 grid of SSU points to derive a density value of soil sealing (imperviousness) at the PSU level (*Figure 2-8*). The values are thus represented by percentage values in 4% steps from 0 – 100%. The process is then repeated for each time point.

The datasets against which the interpretation is performed are divided in two main groups, guiding data and reference data.

The guiding data were used in the production of the IMD and IMC layers and are hosted by the EEA as web services. The available guiding data were:

- 2018 VHR IMAGE dataset (not used for original production)

- 2018 HR IMAGE 10 m dataset (used for original production)
- 2015 VHR IMAGE dataset (not used for original production)
- 2015 HR IMAGE dataset (used for original production)
- 2012 HRIMC1 25 m dataset (used for original production)
- 2009 HRIMC1 20 m dataset (used during production)
- 2006 HRIMC1 20 m dataset (known issue: Shift in relation to 2012 & 2009 datasets)

The reference data were to provide more spatial detail and strong landscape context to the assessment. The available reference data were:

- Bing maps image / cartography layer
- Open Street map data
- GoogleEarth image / cartography data
- Further in-situ data

The interpretation process was controlled, and the results recorded by the selected sample locations. For each location there were two vector layers. The sample points, or PSU, were represented by a 100 m by 100 m rectangle with a centre point at the selected location and the grid points, or SSU, were represented by a set of points on a 5 x 5 grid with 20 m spacing within each PSU (Figure 2-9). Each dataset contained a set of attributes to both define the characteristics of the sample point, PSU or SSU, and the results of the assessment.



*Figure 2-9: An example of the PSU (yellow square) and the SSU (green and red dots) for one sample location based on a LUCAS point.*

For the IMD / IMC the assessment process was applied to approximately 23,000 PSUs each of which contained 25 SSU making in excess of 550,000 total interpretations per time point. To perform an internal quality check 10% of the assessments undertaken by the bulk interpretation team were repeated by an independent team.

For each PSU and its constituent SSUs the following process was followed to complete the assessment.

At the PSU level the guiding data was screened for clouds and cloud shadows and a note was made in the comment attributes when no useable data was available for a reference year. The presence of any spatial shifts either between the guiding datasets or against reference datasets were also noted.

At the SSU level within each PSU the first step was to assess the guiding data in the context of the available reference data to understand the relationship between the guiding data and the reference data. Then, using the guiding data for 2018 in the context of the reference data, an assessment was made whether the surface at each SSU was sealed or unsealed. If the surface was sealed a value of 1 was assigned to the CODE18 attribute on the SSU point feature. If the surface was unsealed a check was made to ensure that the CODE18 attribute for SSUs point feature was set to 0. If necessary, the UNCERTAIN and COMMENT attributes were used to record SSUs

where the coding of sealing was unclear or certain issues occurred. Using the sample units for the third Specific Contract, initially, the 2015 sealing status was copied to the 2018 sealing status. If a change had occurred the appropriate attributes (CODE15 & CODE18) were updated as necessary

The assessment was made for the actual location where each SSU point feature lay in the guiding and reference data and visible in the imagery. For instance, roads under tree canopies are unsealed and structures over lawns are sealed.

## 2.4.4. Estimation and analyses procedures

### 2.4.4.1. Analysis of density values

As described above, density values from the reference data are not directly assessed, but generated from sampled data (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. The approach described below would be applicable to both status and change density layers.

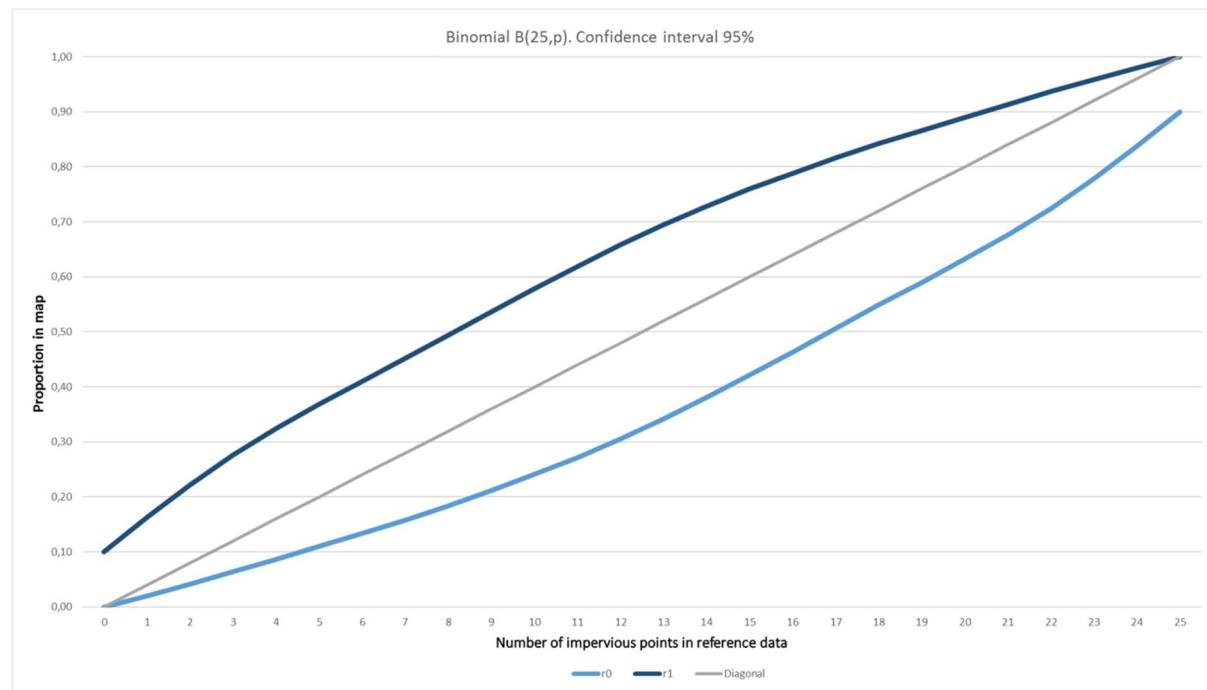
If we had a complete information on the cell for our reference data, a reasonable measure of the commission  $\varphi$  and omission  $\psi$  errors would be:

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

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 100 m 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 impervious 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 2-10. Representation of the behaviour of the 95% confidence interval for a 5x5 SSU grid over the whole range of imperviousness degree values*

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 2-10 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  $\varphi$  and omission errors  $\psi$  would be:

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

#### 2.4.4.2. Thematic accuracy

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 ration 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  $\pi$  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  $+/- 1.96 \sigma$ .

## 2.5. 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. 2018  $+/- 1$  year(s).

## 2.6. 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.

Assess the appropriateness of the metadata description and accompanying documentation. The accompanying documentation should be provided as a pdf and named as follows: CRS\_Information\_Sheet\_<country 2-letter ISO code> e.g. CRS\_Information\_Sheet\_EU.pdf. These will be static and therefore will not have version numbers.

## 2.7. INSPIRE compliant metadata

Presence of INSPIRE compliant metadata should be verified. Each product shall be accompanied with INSPIRE-compliant metadata in XML format and an INSPIRE Mapping Table as \*.xls file.

To test the INSPIRE-compliant metadata in XML format the "Conformance Class 'XML encoding of ISO 19115/19119 metadata'" test suite at <https://inspire.ec.europa.eu/validator/about/> was applied.

### 3. Validation check list

#### 3.1. IMD2018

<b>PRODUCT:</b> IMD2018							
<b>VALIDATION LEVEL:</b> Pan-European							
<b>SERVICE PROVIDER:</b> SPECTO NATURA			<b>SERVICE USER:</b>	<b>ISSUE/REVISION:</b> 1.0			
<b>VALIDATION DATE:</b> 15/07/2020			<b>REVIEW DATE:</b>				
<b>CONDUCTED BY:</b> GEOFF SMITH			<b>REVIEWED BY:</b>	<b>APPROVED BY:</b>			
No.	Data Quality Sub-	Data Quality Measure	Data Quality Result	Comments by Audit Team		Draft Audit Conclusion	Final Audit Conclusion
1	<b>Completeness</b>						
1.1	Commission	Rate of excess items	OK	<ul style="list-style-type: none"> <li>Areas of intertidal / marine water are mapped as unsealed (0), thus can't use as extent of terrestrial areas.</li> <li>Many small islets included which may not be in the countries data.</li> <li>Small islets exaggerated beyond what would be expected for 100 m generalisation.</li> </ul>		Accepted	
1.2	Omission	Rate of missing items	OK			Accepted	
2	<b>Logical consistency</b>						
2.1	Format consistency	File format	OK	<ul style="list-style-type: none"> <li>Additional supplementary files.</li> </ul>		Accepted	
2.2		File name	OK			Accepted	
2.3		Attributes names	n/a				

<b>PRODUCT:</b> IMD2018						
<b>VALIDATION LEVEL:</b> Pan-European						
<b>SERVICE PROVIDER:</b> SPECTO NATURA			<b>SERVICE USER:</b>	<b>ISSUE/REVISION:</b> 1.0		
<b>VALIDATION DATE:</b> 15/07/2020			<b>REVIEW DATE:</b>			
<b>CONDUCTED BY:</b> GEOFF SMITH			<b>REVIEWED BY:</b>	<b>APPROVED BY:</b>		
No.	Data Quality Sub-	Data Quality Measure	Data Quality Result	Comments by Audit Team		Draft Audit Conclusion
2.4		Attributes types	OK			Accepted
2.5	Conceptual consistency	Feature type	n/a	not applicable for raster data		
2.6		MMU	OK			Accepted
2.7		Coordinate reference system	OK			Accepted
2.8		Unique identifier	n/a	not applicable for raster data		
2.9		Nomenclature	OK			Accepted
2.10	Domain consistency	Value domain non-conformance	OK			Accepted
2.11	Topological consistency	Overlaps	n/a	not applicable for raster data		

<b>PRODUCT:</b> IMD2018							
<b>VALIDATION LEVEL:</b> Pan-European							
<b>SERVICE PROVIDER:</b> SPECTO NATURA			<b>SERVICE USER:</b>	<b>ISSUE/REVISION:</b> 1.0			
<b>VALIDATION DATE:</b> 15/07/2020			<b>REVIEW DATE:</b>				
<b>CONDUCTED BY:</b> GEOFF SMITH			<b>REVIEWED BY:</b>	<b>APPROVED BY:</b>			
No.	Data Quality Sub-	Data Quality Measure	Data Quality Result	Comments by Audit Team		Draft Audit Conclusion	Final Audit Conclusion
2.12		Gaps	n/a	not applicable for raster data			
2.13		Multipart features	n/a	not applicable for raster data			
2.14		Neighbouring features	n/a	not applicable for raster data			
2.15		Self-intersections	n/a	not applicable for raster data			
2.16		Null geometry	n/a	not applicable for raster data			
2.17		Unclosed rings	n/a	not applicable for raster data			
2.18		Duplicate vertex	n/a	not applicable for raster data			
2.19		Pseudo nodes	n/a	not applicable for raster data			

<b>PRODUCT:</b> IMD2018						
<b>VALIDATION LEVEL:</b> Pan-European						
<b>SERVICE PROVIDER:</b> SPECTO NATURA			<b>SERVICE USER:</b>	<b>ISSUE/REVISION:</b> 1.0		
<b>VALIDATION DATE:</b> 15/07/2020			<b>REVIEW DATE:</b>			
<b>CONDUCTED BY:</b> GEOFF SMITH			<b>REVIEWED BY:</b>	<b>APPROVED BY:</b>		
No.	<b>DATA QUALITY SUB-</b>	<b>DATA QUALITY MEASURE</b>	<b>DATA QUALITY RESULT</b>	<b>COMMENTS BY AUDIT TEAM</b>	<b>DRAFT AUDIT CONCLUSION</b>	<b>FINAL AUDIT CONCLUSION</b>
2.20		Non matching nodes	n/a	not applicable for raster data		
3	<b>POSITIONAL ACCURACY</b>					
3.1	Absolute or external accuracy	RMSEP	n/a	This is dependent on the assessment of the input image data		
3.2	Relative or internal accuracy	RMSEP	n/a	This is dependent on the assessment of the input image data		
4	<b>THEMATIC ACCURACY</b>					
4.1	Classification correctness	Overall accuracy	OK	See section 4 for final thematic accuracy values of the products	Accepted	
4.2		Min. producer's accuracy	OK NOK	See section 4 for final thematic accuracy values of the products		
4.3		Min. user's accuracy	OK NOK	See section 4 for final thematic accuracy values of the products		
4.4		Kappa	n/a	Not specified		

<b>PRODUCT:</b> IMD2018							
<b>VALIDATION LEVEL:</b> Pan-European							
<b>SERVICE PROVIDER:</b> SPECTO NATURA			<b>SERVICE USER:</b>	<b>ISSUE/REVISION:</b> 1.0			
<b>VALIDATION DATE:</b> 15/07/2020			<b>REVIEW DATE:</b>				
<b>CONDUCTED BY:</b> GEOFF SMITH			<b>REVIEWED BY:</b>	<b>APPROVED BY:</b>			
No.	DATA QUALITY SUB-	DATA QUALITY MEASURE	DATA QUALITY RESULT	<b>COMMENTS BY AUDIT TEAM</b>		DRAFT AUDIT CONCLUSION	FINAL AUDIT CONCLUSION
5	<b>TEMPORAL QUALITY</b>						
5.1	Temporal quality	Closeness of the acquired image data to the reference year		No information available			
6	<b>USABILITY</b>						
6.1	Usability	Usability description	OK	<ul style="list-style-type: none"> <li>Different supplementary files</li> <li>Behaviour is different in different image processing systems, likely to be related to associated files.</li> </ul>		Accepted	
7	<b>METADATA</b>						
7.1	INSPIRE compliant metadata	Presence	OK			Accepted	
7.2		File format	OK			Accepted	
7.3		File name	OK			Accepted	

<b>PRODUCT:</b> IMD2018						
<b>VALIDATION LEVEL:</b> Pan-European						
<b>SERVICE PROVIDER:</b> SPECTO NATURA			<b>SERVICE USER:</b>	<b>ISSUE/REVISION:</b> 1.0		
<b>VALIDATION DATE:</b> 15/07/2020			<b>REVIEW DATE:</b>			
<b>CONDUCTED BY:</b> GEOFF SMITH			<b>REVIEWED BY:</b>	<b>APPROVED BY:</b>		
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		Accepted	

### 3.2. SBU2018

<b>PRODUCT:</b> SBU2018							
<b>VALIDATION LEVEL:</b> Pan-European							
<b>SERVICE PROVIDER:</b> SPECTO NATURA		<b>SERVICE USER:</b>		<b>ISSUE/REVISION:</b> 1.0			
<b>VALIDATION DATE:</b> 18/08/2020		<b>REVIEW DATE:</b>					
<b>CONDUCTED BY:</b> GEOFF SMITH		<b>REVIEWED BY:</b>		<b>APPROVED BY:</b>			
No.	Data Quality Sub-	Data Quality Measure	Data Quality Result	<b>COMMENTS BY AUDIT TEAM</b>			Draft Audit Conclusion
1	<b>Completeness</b>						
1.1	Commission	Rate of excess items	OK	<ul style="list-style-type: none"> <li>Areas of intertidal / marine water are mapped as 0 % built-up, thus can't use as extent of terrestrial areas.</li> <li>Issues carried over from the IMD layer</li> <li>Many small islets included which may not be in the countries data.</li> <li>Small islets exaggerated beyond what would be expected for 100 m generalisation.</li> </ul>			
1.2	Omission	Rate of missing items	OK				
2	<b>Logical consistency</b>						
2.1	Format consistency	File format	OK	<ul style="list-style-type: none"> <li>Additional supplementary files.</li> </ul>			
2.2		File name	OK				
2.3		Attributes names	n/a	not applicable for raster data			

<b>PRODUCT:</b> SBU2018							
<b>VALIDATION LEVEL:</b> Pan-European							
<b>SERVICE PROVIDER:</b> SPECTO NATURA			<b>SERVICE USER:</b>	<b>ISSUE/REVISION:</b> 1.0			
<b>VALIDATION DATE:</b> 18/08/2020			<b>REVIEW DATE:</b>				
<b>CONDUCTED BY:</b> GEOFF SMITH			<b>REVIEWED BY:</b>	<b>APPROVED BY:</b>			
No.	Data Quality Sub-	Data Quality Measure	Data Quality Result	Comments by Audit Team		Draft Audit Conclusion	Final Audit Conclusion
2.4		Attributes types	OK				
2.5	Conceptual consistency	Feature type	n/a	not applicable for raster data			
2.6		MMU	OK				
2.7		Coordinate reference system	OK				
2.8		Unique identifier	n/a	not applicable for raster data			
2.9		Nomenclature	OK				
2.10	Domain consistency	Value domain non-conformance	OK				
2.11	Topological consistency	Overlaps	n/a	not applicable for raster data			

<b>PRODUCT:</b> SBU2018							
<b>VALIDATION LEVEL:</b> Pan-European							
<b>SERVICE PROVIDER:</b> SPECTO NATURA			<b>SERVICE USER:</b>	<b>ISSUE/REVISION:</b> 1.0			
<b>VALIDATION DATE:</b> 18/08/2020			<b>REVIEW DATE:</b>				
<b>CONDUCTED BY:</b> GEOFF SMITH			<b>REVIEWED BY:</b>	<b>APPROVED BY:</b>			
No.	Data Quality Sub-	Data Quality Measure	Data Quality Result	Comments by Audit Team		Draft Audit Conclusion	Final Audit Conclusion
2.12		Gaps	n/a	not applicable for raster data			
2.13		Multipart features	n/a	not applicable for raster data			
2.14		Neighbouring features	n/a	not applicable for raster data			
2.15		Self-intersections	n/a	not applicable for raster data			
2.16		Null geometry	n/a	not applicable for raster data			
2.17		Unclosed rings	n/a	not applicable for raster data			
2.18		Duplicate vertex	n/a	not applicable for raster data			
2.19		Pseudo nodes	n/a	not applicable for raster data			

<b>PRODUCT:</b> SBU2018							
<b>VALIDATION LEVEL:</b> Pan-European							
<b>SERVICE PROVIDER:</b> SPECTO NATURA			<b>SERVICE USER:</b>	<b>ISSUE/REVISION:</b> 1.0			
<b>VALIDATION DATE:</b> 18/08/2020			<b>REVIEW DATE:</b>				
<b>CONDUCTED BY:</b> GEOFF SMITH			<b>REVIEWED BY:</b>	<b>APPROVED BY:</b>			
No.	Data Quality Sub-	Data Quality Measure	Data Quality Result	Comments by Audit Team		Draft Audit Conclusion	Final Audit Conclusion
2.20		Non matching nodes	n/a	not applicable for raster data			
3	<b>POSITIONAL ACCURACY</b>						
3.1	Absolute or external accuracy	RMSEP	n/a	This is dependent on the assessment of the input image data			
3.2	Relative or internal accuracy	RMSEP	n/a	This is dependent on the assessment of the input image data			
4	<b>THEMATIC ACCURACY</b>						
4.1	Classification correctness	Overall accuracy	OK	<ul style="list-style-type: none"> <li>See section 5 for final thematic accuracy values of the products</li> <li>Visual check: Rocky islet with no buildings given a share of built-up value &gt; 0%</li> </ul>			
4.2		Min. producer's accuracy	OK NOK	See section 5 for final thematic accuracy values of the products			
4.3		Min. user's accuracy	OK NOK	See section 5 for final thematic accuracy values of the products			

<b>PRODUCT:</b> SBU2018							
<b>VALIDATION LEVEL:</b> Pan-European							
<b>SERVICE PROVIDER:</b> SPECTO NATURA			<b>SERVICE USER:</b>	<b>ISSUE/REVISION:</b> 1.0			
<b>VALIDATION DATE:</b> 18/08/2020			<b>REVIEW DATE:</b>				
<b>CONDUCTED BY:</b> GEOFF SMITH			<b>REVIEWED BY:</b>	<b>APPROVED BY:</b>			
No.	DATA QUALITY SUB-	DATA QUALITY MEASURE	DATA QUALITY RESULT	<b>COMMENTS BY AUDIT TEAM</b>		DRAFT AUDIT CONCLUSION	FINAL AUDIT CONCLUSION
4.4		Kappa	n/a	Not specified			
5	<b>TEMPORAL QUALITY</b>						
5.1	Temporal quality	Closeness of the acquired image data to the reference year		No information available			
6	<b>USABILITY</b>						
6.1	Usability	Usability description	OK	<ul style="list-style-type: none"> <li>Different supplementary files</li> <li>Behaviour is different in different image processing systems, likely to be related to associated files.</li> </ul>			
7	<b>METADATA</b>						
7.1	INSPIRE compliant metadata	Presence	OK				
7.2		File format	OK				

<b>PRODUCT:</b> SBU2018							
<b>VALIDATION LEVEL:</b> Pan-European							
<b>SERVICE PROVIDER:</b> SPECTO NATURA			<b>SERVICE USER:</b>	<b>ISSUE/REVISION:</b> 1.0			
<b>VALIDATION DATE:</b> 18/08/2020			<b>REVIEW DATE:</b>				
<b>CONDUCTED BY:</b> GEOFF SMITH			<b>REVIEWED BY:</b>	<b>APPROVED BY:</b>			
No.	DATA QUALITY SUB-	DATA QUALITY MEASURE	DATA QUALITY RESULT	<b>COMMENTS BY AUDIT TEAM</b>		<b>DRAFT AUDIT CONCLUSION</b>	<b>FINAL AUDIT CONCLUSION</b>
7.3		File name	OK				
7.4		INSPIRE compliance	OK				

### 3.3. IMC1518

<b>PRODUCT:</b> IMC1518							
<b>VALIDATION LEVEL:</b> Pan-European							
<b>SERVICE PROVIDER:</b> SPECTO NATURA		<b>SERVICE USER:</b>		<b>ISSUE/REVISION:</b> 1.0			
<b>VALIDATION DATE:</b> 08/09/2020		<b>REVIEW DATE:</b>					
<b>CONDUCTED BY:</b> GEOFF SMITH		<b>REVIEWED BY:</b>		<b>APPROVED BY:</b>			
No.	Data Quality Sub-	Data Quality Measure	Data Quality Result	<b>COMMENTS BY AUDIT TEAM</b>		Draft Audit Conclusion	Final Audit Conclusion
1	<b>COMPLETENESS</b>						
1.1	Commission	Rate of excess items	OK	<ul style="list-style-type: none"> <li>Areas of intertidal / marine water are mapped as unclassifiable and unsealed/unchanged (201), thus can't use as extent of terrestrial areas.</li> <li>Issues carried over from the IMD layer.</li> <li>Many small islets included which may not be in the countries data.</li> <li>Small islets exaggerated beyond what would be expected for 100 m generalisation.</li> </ul>			
1.2	Omission	Rate of missing items	OK				
2	<b>LOGICAL CONSISTENCY</b>						
2.1	Format consistency	File format	OK	<ul style="list-style-type: none"> <li>Additional supplementary files.</li> </ul>			
2.2		File name	OK				
2.3		Attributes names	n/a	not applicable for raster data			

<b>PRODUCT:</b> IMC1518							
<b>VALIDATION LEVEL:</b> Pan-European							
<b>SERVICE PROVIDER:</b> SPECTO NATURA			<b>SERVICE USER:</b>	<b>ISSUE/REVISION:</b> 1.0			
<b>VALIDATION DATE:</b> 08/09/2020			<b>REVIEW DATE:</b>				
<b>CONDUCTED BY:</b> GEOFF SMITH			<b>REVIEWED BY:</b>	<b>APPROVED BY:</b>			
No.	Data Quality Sub-	Data Quality Measure	Data Quality Result	Comments by Audit Team		Draft Audit Conclusion	Final Audit Conclusion
2.4		Attributes types	OK				
2.5	Conceptual consistency	Feature type	n/a	not applicable for raster data			
2.6		MMU	OK				
2.7		Coordinate reference system	OK				
2.8		Unique identifier	n/a	not applicable for raster data			
2.9		Nomenclature	OK				
2.10	Domain consistency	Value domain non-conformance	OK				
2.11	Topological consistency	Overlaps	n/a	not applicable for raster data			

<b>PRODUCT:</b> IMC1518							
<b>VALIDATION LEVEL:</b> Pan-European							
<b>SERVICE PROVIDER:</b> SPECTO NATURA			<b>SERVICE USER:</b>	<b>ISSUE/REVISION:</b> 1.0			
<b>VALIDATION DATE:</b> 08/09/2020			<b>REVIEW DATE:</b>				
<b>CONDUCTED BY:</b> GEOFF SMITH			<b>REVIEWED BY:</b>	<b>APPROVED BY:</b>			
No.	Data Quality Sub-	Data Quality Measure	Data Quality Result	Comments by Audit Team		Draft Audit Conclusion	Final Audit Conclusion
2.12		Gaps	n/a	not applicable for raster data			
2.13		Multipart features	n/a	not applicable for raster data			
2.14		Neighbouring features	n/a	not applicable for raster data			
2.15		Self-intersections	n/a	not applicable for raster data			
2.16		Null geometry	n/a	not applicable for raster data			
2.17		Unclosed rings	n/a	not applicable for raster data			
2.18		Duplicate vertex	n/a	not applicable for raster data			
2.19		Pseudo nodes	n/a	not applicable for raster data			

<b>PRODUCT:</b> IMC1518						
<b>VALIDATION LEVEL:</b> Pan-European						
<b>SERVICE PROVIDER:</b> SPECTO NATURA			<b>SERVICE USER:</b>	<b>ISSUE/REVISION:</b> 1.0		
<b>VALIDATION DATE:</b> 08/09/2020			<b>REVIEW DATE:</b>			
<b>CONDUCTED BY:</b> GEOFF SMITH			<b>REVIEWED BY:</b>	<b>APPROVED BY:</b>		
No.	Data Quality Sub-	Data Quality Measure	Data Quality Result	<b>COMMENTS BY AUDIT TEAM</b>		Draft Audit Conclusion
2.20		Non matching nodes	n/a	not applicable for raster data		
3	<b>POSITIONAL ACCURACY</b>					
3.1	Absolute or external accuracy	RMSEP	n/a	This is dependent on the assessment of the input image data		
3.2	Relative or internal accuracy	RMSEP	n/a	This is dependent on the assessment of the input image data		
4	<b>THEMATIC ACCURACY</b>					
4.1	Classification correctness	Overall accuracy	OK	See section 6 for final thematic accuracy values of the products		
4.2		Min. producer's accuracy	OK NOK	See section 6 for final thematic accuracy values of the products		
4.3		Min. user's accuracy	OK NOK	See section 6 for final thematic accuracy values of the products		

<b>PRODUCT:</b> IMC1518							
<b>VALIDATION LEVEL:</b> Pan-European							
<b>SERVICE PROVIDER:</b> SPECTO NATURA			<b>SERVICE USER:</b>	<b>ISSUE/REVISION:</b> 1.0			
<b>VALIDATION DATE:</b> 08/09/2020			<b>REVIEW DATE:</b>				
<b>CONDUCTED BY:</b> GEOFF SMITH			<b>REVIEWED BY:</b>	<b>APPROVED BY:</b>			
No.	DATA QUALITY SUB-	DATA QUALITY MEASURE	DATA QUALITY RESULT	<b>COMMENTS BY AUDIT TEAM</b>		DRAFT AUDIT CONCLUSION	FINAL AUDIT CONCLUSION
4.4		Kappa	n/a	Not specified			
5	<b>TEMPORAL QUALITY</b>						
5.1	Temporal quality	Closeness of the acquired image data to the reference year		No information available			
6	<b>USABILITY</b>						
6.1	Usability	Usability description	OK	<ul style="list-style-type: none"> <li>Different supplementary files Behaviour is different in different image processing systems, likely to be related to associated files.</li> </ul>			
7	<b>METADATA</b>						
7.1	INSPIRE compliant metadata	Presence	OK				
7.2		File format	OK				

<b>PRODUCT:</b> IMC1518							
<b>VALIDATION LEVEL:</b> Pan-European							
<b>SERVICE PROVIDER:</b> SPECTO NATURA			<b>SERVICE USER:</b>	<b>ISSUE/REVISION:</b> 1.0			
<b>VALIDATION DATE:</b> 08/09/2020			<b>REVIEW DATE:</b>				
<b>CONDUCTED BY:</b> GEOFF SMITH			<b>REVIEWED BY:</b>	<b>APPROVED BY:</b>			
No.	DATA QUALITY SUB-	DATA QUALITY MEASURE	DATA QUALITY RESULT	<b>COMMENTS BY AUDIT TEAM</b>		DRAFT AUDIT CONCLUSION	FINAL AUDIT CONCLUSION
7.3		File name	OK				
7.4		INSPIRE compliance	OK				

## 4. Thematic accuracy IMD2018

The thematic accuracy is assessed using a number of approaches and within a number of regionalisations to better understand the thematic characteristics across biogeographic regions and countries.

The assessment was based on a sample set of 18,005 reference points.

### 4.1. IMD2018 scatterplots & regression analysis

A scatterplot is a way of displaying data against Cartesian coordinates to show and compare values for two variables within a dataset. The data is displayed as a series of points, where the x and y locations relate two variables assigned to a particular recording instance, in this case a PSU. The three available measurements for each PSU are the original reference data (CODE), the reference data after plausibility analysis (QC) and the mapped value from the product (MAP). For this validation exercise the position / value on the horizontal axis represented the reference (CODE or QC) information and the position / value on the vertical axis represents product (MAP) information. In this way the relation of the reference and product information for a point can be compared to a 1:1 line which runs diagonally across the scatter plot. The closeness of a point to the point to the 1:1 line is an indication of the similarity between the reference and mapped results. The points that lie exactly on the x and y axes are related to omission and commission rather than the calibration of the IMD values themselves, but continue to be included in the regression calculations for consistency with previous reference years.

The challenge of accurately recording imperviousness in 10 m EO data and then aggregating this to a 1 ha grid is obvious from the distribution of points displayed on Figure 4-1. The plots for both blind interpretation and plausibility analysis shows considerable scatter of up to 20 % each side of the best fit line in the majority of cases. There are also considerable numbers of point on the x and y axes showing commission, where sealing is mapped that is not present in reality, and omission, where sealed areas are missed. These cases seemed to be limited in most cases to examples which represent sealing degrees of less than 35%. It can also be seen that the distribution is not centred on the 1:1 line but falls below the line indicating that the mapped products are underestimating the actual imperviousness as reported by the reference data. In the case of the plausibility analysis the spread of points around the best fit line is slightly reduced as would be expected. These conclusions are repeated consistently across the different regions and between the blind interpretation and the plausibility analysis.

The scatterplots and the resulting best fit lines are controlled by the actual geography of the regions being considered so sub-regionalisations can differ significantly from the results presented in Figure 4-1 when considering Europe as a whole. In Figure 4-2 the results for the Black Sea BRME zone show that there are limited number of sealed areas, the areas present tend to have low imperviousness values and the relationship between the reference and mapped data is controlled by a few point with only medium levels of sealing. For this regionalisation this has resulted in imperviousness being more underestimated relative to the reference data than for the whole dataset and other regionalisations. There are also a number of omissions and commissions which is highly likely for a region where the population is relatively small and there are fewer urban.

The Atlantic BRME zone (Figure 4-3) is more representative of Europe as a whole and contains significant sealed areas with a broad range of imperviousness density. The distribution of points appears relatively tight around the best fit line given the known +20 % uncertainties visible in Figure 4-1. Given the variation in landscape types across Europe the Atlantic region is likely to be optimal in terms of the methods applied and representative of what would be expected. There may be two sub-groups of points; a large group with imperviousness density below approximately 40 % related to suburban areas and a second group with higher imperviousness density related to city centres and industrial sites. There is also a significant group of points which are marked as completely sealed or unsealed in the reference data but exhibit a range of values in the mapped data, which may indicate that certain sealed surfaces are more difficult to detect or calibrated for in the EO data. In these examples it can be seen that the plausibility analysis removes some of the commission and omission errors.

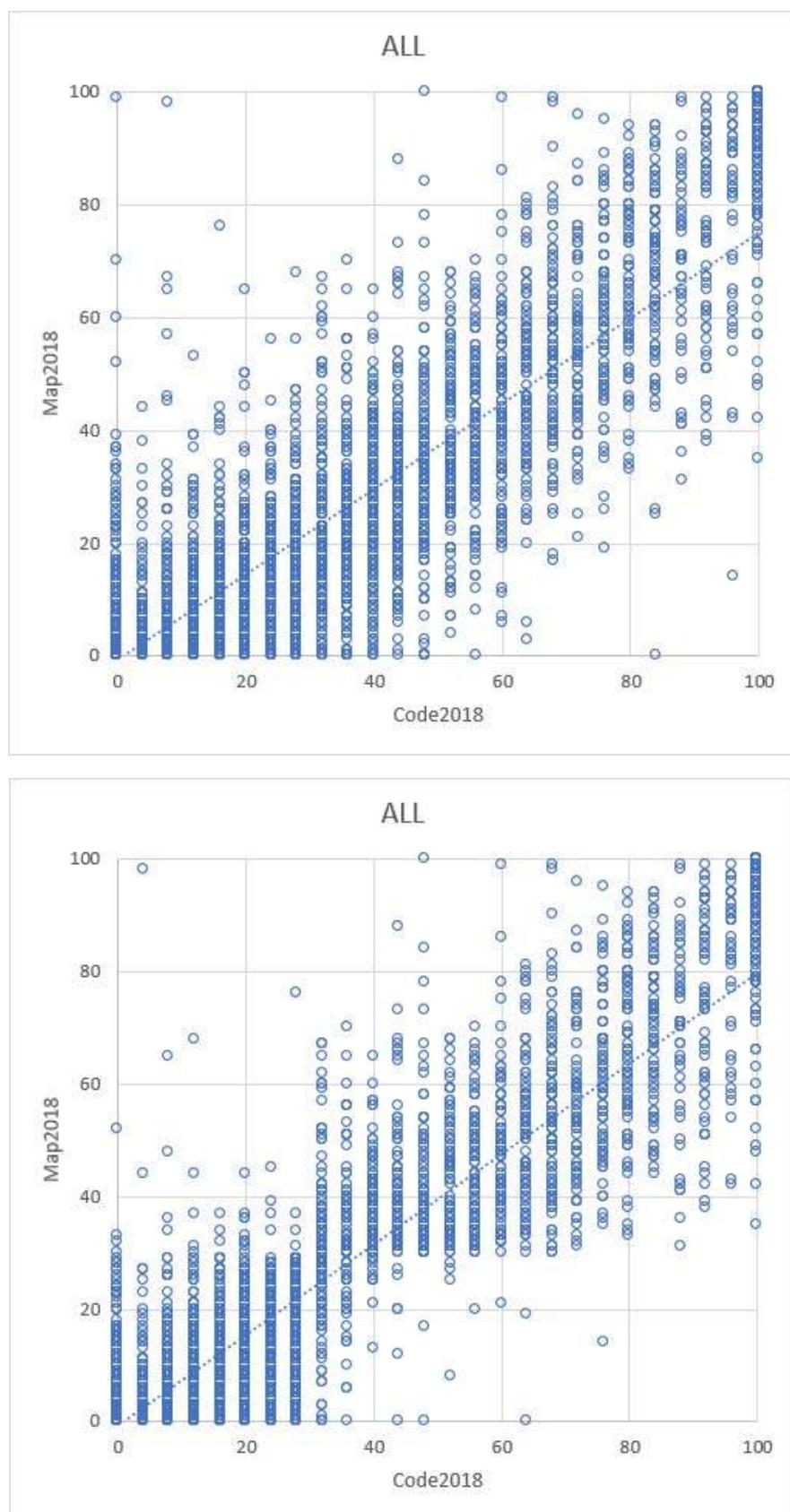


Figure 4-1: Scatterplots for 2018 for ALL the reference data for IMD for both the blind interpretation (TOP) and plausibility analysis (BOTTOM).

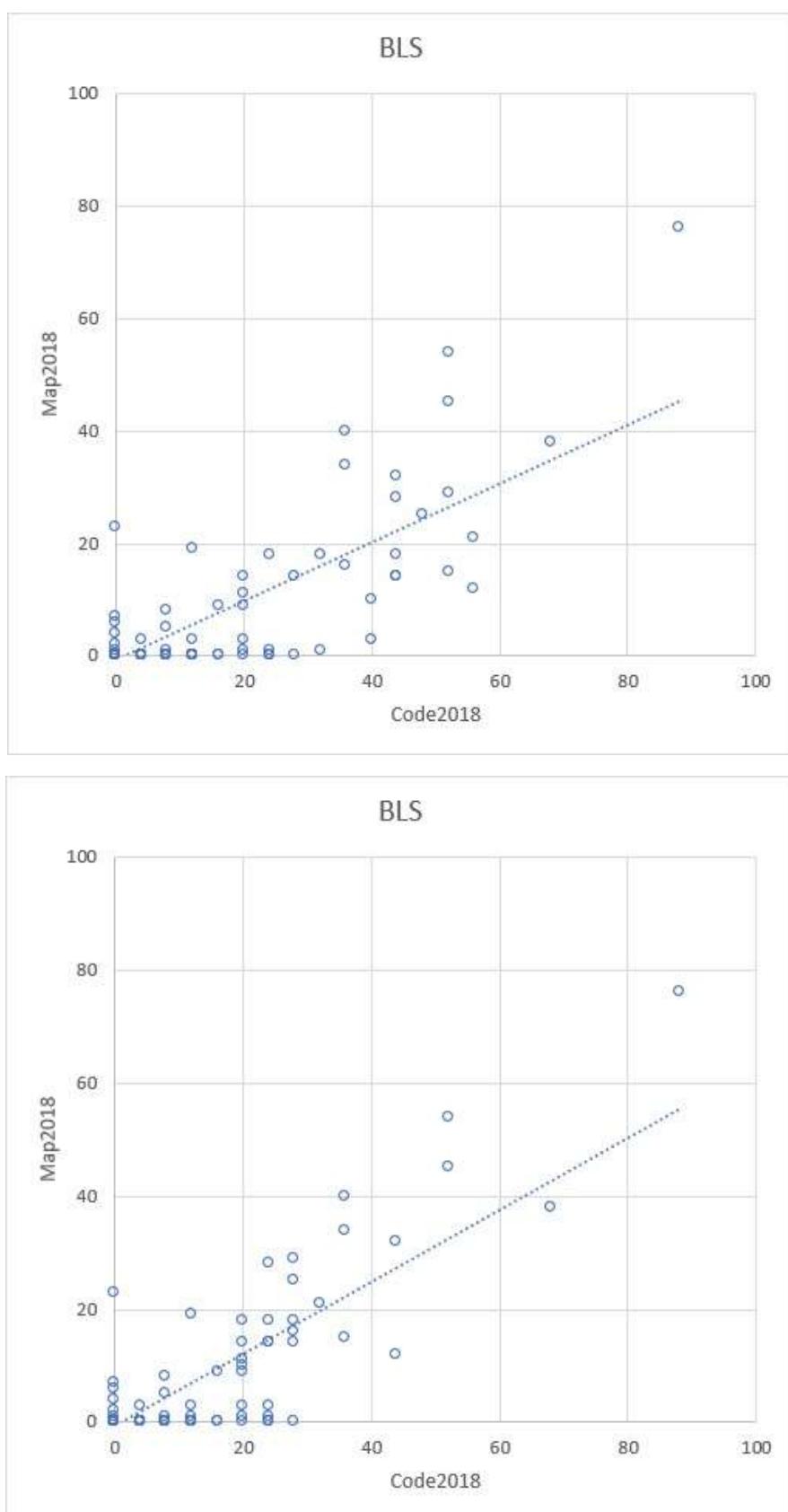


Figure 4-2: Scatterplots for 2018 for Black Sea (BLS) BRME zone for IMD for both the blind interpretation (TOP) and plausibility analysis (BOTTOM).

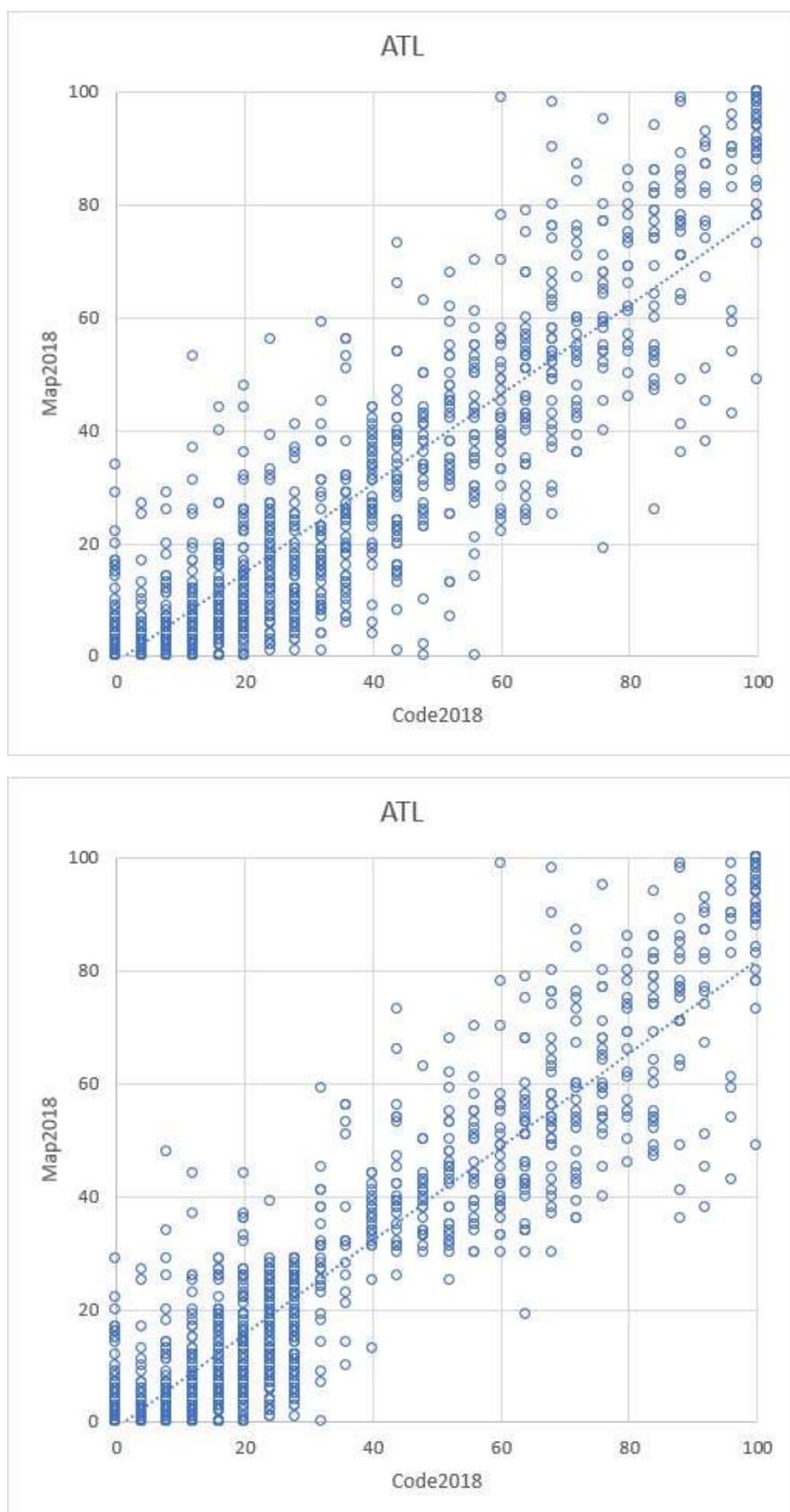


Figure 4-3: Scatterplots for 2018 for Atlantic (ATL) BRME zone for IMD for both the blind interpretation (TOP) and plausibility analysis (BOTTOM).

To quantitatively summarise the results displayed in the scatterplots above a linear regression analysis is performed to estimate the relationships between the reference and mapped product information. The analysis produces a coefficient of determination ( $R^2$ ) which gives information about the goodness of fit of the estimated regression model. Coefficients of determination closer to 1 represent a better fit. In this case as the reference and map information are meant to represent the same information then it is useful to also consider the slope and intercept of the estimated regression model. The slope should therefore approach 1 and the intercept should be close to 0 for the required relationships. Deviations from the expected values give an indication of the correspondence of the reference and mapped imperviousness data.

The results of the regression analysis of the map imperviousness densities to the reference data are given in Table 4-1 for the blind interpretation and Table 4-2 for the plausibility analysis. There are some missing items in the table due to data for the French DOMs not being available at the time of analysis.

As expected, the plausibility analysis results give higher  $R^2$  values and slopes closer to 1 as more of the outliers have been adjusted to account for uncertainties in the validation approach. Only in the 2015 results were there two minor reduction in  $R^2$  for particular regionalisations. Subsequent discussions here will focus on the results of the plausibility analysis (Table 4-2).

Compared to the 2015 results the  $R^2$  for the pan-European (ALL) dataset rose from around 0.84 to 0.88 continuing a trend from 2012 to 2015, but with a slightly lower increase of 0.04 compared to 0.06 previously. A more dramatic increase was found in the  $R^2$  values at the lower end of the distribution for the regions which rose from around 0.48 to 0.72. Overall, there continues to be a slight overall increasing trend in the  $R^2$  values across time within the majority of regionalisations, however this is unlikely to be significant in most cases. There is still some variation between the regions with  $R^2$  ranging from 0.72, a moderate relationship, to 0.92, suggesting a strong correspondence between the reference and mapped data. As expected, the lowest values are for the regions where there are few points or limited amounts of sealed surfaces such as the Black Sea and Macaronesian BRME regions and the Scandinavian countries. The highest values of  $R^2$  had consistently been for Iceland where the reference points are dominated by low or zero imperviousness values and there is a single point with an imperviousness value. However, for 2018 Bulgaria came out with the highest  $R^2$  value of 0.92 which represented a wider range of imperviousness density values. It should always be noted that these values are not only affected by the changing character and make-up of the landscape, but also the numbers of samples used in each region.

The slope values have all remained less than 1.0 suggesting that there is still a consistent underestimation of the imperviousness across the extent of the product and across the epochs of production which is indicative of a relatively stable calibration. The slope values for all regions and dates increased towards 1.0 suggesting an improvement to that calibration. The pan-European (ALL) slopes increased by 0.07 and the regions had a maximum increase in of 0.19. The slope closest to the line of equality was 0.92 for Portugal, but there was still a considerable range of slopes with the minimum at 0.59 for Finland.

Overall, the regression results show a continued underestimation of the of imperviousness by the IMD layer compared to the reference data, but in comparison to 2015 the calibration has been improved and there is great consistency across the various regionalisations.

*Table 4-1: A summary of regression line parameters for the scatterplots derived from the blind interpretation as reference data for 2015 and 2018.*

	2015			2018		
	R <sup>2</sup>	Slp	Int	R <sup>2</sup>	Slp	Int
European All	0.78	0.67	-0.63	0.84	0.76	-0.59
ALP	0.78	0.59	-0.31	0.83	0.65	-0.31
ANA	0.62	0.48	0.05	0.73	0.65	-0.05
ARC	0.84	0.63	-0.03	0.83	0.77	-0.01
ATL	0.81	0.73	-1.25	0.87	0.79	-0.95
BLS	0.50	0.43	0.00	0.68	0.52	-0.48
BOR	0.68	0.46	-0.31	0.70	0.62	-0.41
CON	0.78	0.70	-0.71	0.83	0.77	-0.61
MAC	0.43	0.48	0.33	0.60	0.58	-0.26
MED	0.79	0.67	-0.62	0.84	0.78	-0.62
PAN	0.83	0.52	-0.29	0.82	0.61	-0.16
STE	0.85	0.53	-0.47	0.83	0.72	-0.53
AL+ME+MK+RS+XK	0.76	0.51	-0.29	0.84	0.59	-0.38
AT + CH + LI	0.82	0.72	-0.78	0.89	0.75	-0.67
BA + HR + SI	0.76	0.65	-0.64	0.78	0.69	-0.57
BE + LU+ NL + DK	0.82	0.79	-0.82	0.86	0.85	-0.34
BG	0.81	0.66	0.19	0.91	0.70	-0.18
CZ + SK	0.84	0.79	0.11	0.83	0.86	0.22
DE	0.82	0.75	-1.11	0.88	0.82	-0.40
EE + LT + LV	0.78	0.63	-0.44	0.76	0.73	-0.43
EL	0.82	0.65	-0.75	0.81	0.74	-0.93
ES	0.77	0.63	-0.54	0.84	0.78	-0.48
FI	0.71	0.41	-0.21	0.74	0.56	-0.32
FR	0.78	0.69	-1.38	0.84	0.77	-1.19
FR DOMs						
HU	0.82	0.54	-0.60	0.84	0.65	-0.65
IE + UK	0.82	0.71	-0.58	0.88	0.75	-0.55
IS	0.90	0.68	-0.01	0.88	0.83	0.02
IT	0.79	0.66	-1.25	0.83	0.80	-1.19
NO	0.63	0.45	-0.22	0.69	0.56	-0.23
PL	0.72	0.73	-0.65	0.73	0.71	-0.47
PT	0.81	0.75	-1.00	0.89	0.89	0.85
RO	0.81	0.49	-0.38	0.83	0.57	-0.36
SE	0.63	0.45	-0.39	0.70	0.64	-0.51
TR	0.70	0.60	-0.16	0.77	0.68	-0.32

*Table 4-2: A summary of regression line parameters for the scatterplots derived from the plausibility analysis as reference data.*

	2015			2018		
	R <sup>2</sup>	Slp	Int	R <sup>2</sup>	Slp	Int
European All	0.84	0.73	-0.68	0.88	0.80	-0.60
ALP	0.83	0.66	-0.32	0.87	0.72	-0.30
ANA	0.74	0.57	-0.06	0.83	0.73	-0.21
ARC	0.84	0.63	-0.03	0.83	0.77	-0.01
ATL	0.86	0.78	-1.29	0.90	0.83	-0.79
BLS	0.53	0.58	-0.19	0.72	0.64	-0.60
BOR	0.73	0.56	-0.42	0.74	0.67	-0.48
CON	0.85	0.77	-0.76	0.88	0.83	-0.65
MAC	0.48	0.53	0.10	0.76	0.72	-1.09
MED	0.85	0.72	-0.67	0.88	0.82	-0.68
PAN	0.84	0.59	0.02	0.86	0.67	-0.10
STE	0.82	0.61	-0.24	0.84	0.74	-0.54
AL+ME+MK+RS+XK	0.82	0.60	-0.28	0.89	0.67	-0.35
AT + CH + LI	0.87	0.77	-0.78	0.91	0.79	-0.51
BA + HR + SI	0.80	0.70	-0.66	0.82	0.75	-0.69
BE + LU+ NL + DK	0.89	0.83	-1.09	0.88	0.88	-0.01
BG	0.89	0.74	0.01	0.92	0.74	-0.10
CZ + SK	0.86	0.83	0.20	0.87	0.88	0.09
DE	0.87	0.81	-1.01	0.91	0.85	-0.38
EE + LT + LV	0.78	0.67	-0.38	0.81	0.75	-0.52
EL	0.80	0.65	-0.58	0.84	0.76	-0.91
ES	0.82	0.67	-0.59	0.88	0.81	-0.57
FI	0.72	0.48	-0.27	0.74	0.59	-0.35
FR	0.84	0.76	-1.38	0.87	0.82	-0.97
FR DOMs						
HU	0.84	0.64	-0.30	0.87	0.71	-0.59
IE + UK	0.86	0.76	-0.64	0.90	0.78	-0.44
IS	0.90	0.68	-0.01	0.88	0.83	0.02
IT	0.87	0.73	-1.42	0.90	0.84	-1.46
NO	0.72	0.55	-0.30	0.76	0.66	-0.28
PL	0.80	0.81	-0.85	0.84	0.83	-0.81
PT	0.84	0.80	-0.93	0.91	0.92	-0.79
RO	0.82	0.57	-0.21	0.86	0.63	-0.32
SE	0.73	0.59	-0.56	0.74	0.68	-0.56
TR	0.80	0.67	-0.26	0.83	0.74	-0.37

## 4.2. IMD2018 blind interpretation correspondence analysis

For each of the regionalisations, the correspondence analysis (user's and producer's accuracies) and the 95 % confidence intervals were calculated. The results for the blind interpretation are given in *Table 4-3* which shows the correspondence analysis for the blind interpretation and mapped values of imperviousness converted to urban area using a 30% threshold. The overall accuracies, which were not part of the product specification, exceeded the target thematic accuracy 90% (not shown for clarity) but these were not representative as the urban area makes up a relatively small fraction of the European landscape even in the most densely built-up areas.

It is more appropriate therefore to consider the user's and producer's accuracies, where just under half of the regionalisations achieved the required thematic accuracy of 90 % with a small number falling just short of the requirement, but within 5 %. The user's accuracy, or error of commission, tends to reach the requirement in two thirds of cases, whereas the producer's accuracy, errors of omission, tend to fall considerably short of the requirement. The correspondence results are more variable across the regionalisations due to the sizes of the regions and the distribution of urban areas across Europe. Also, the high thematic accuracies tend to be associated with the user's rather than producer's accuracy supporting the conclusion that the IMD layers are underestimating imperviousness and therefore creating omission errors in the urban layer.

Table 4-3: A summary of the correspondence results for the blind interpretation for the 2015 and 2018 results. Green cells exceed the 90 % minimum accuracy requirement, orange cells are within 5% and red cells fall more than 5% short.

	2015				2018			
	User	CI95%	Prod	CI95%	User	CI95%	Prod	CI95%
European All	91.78%	0.05%	55.10%	0.34%	92.79%	0,13%	66.40%	0,32%
ALP	94.23%	0.02%	52.13%	0.04%	95,24%	0,02%	56,08%	0,04%
ANA	67.86%	0.03%	43.18%	0.04%	81,58%	0,03%	58,49%	0,04%
ARC	100.00%	0.00%	100.00%	0.00%	100,00%	0,00%	100,00%	0,00%
ATL	95.53%	0.07%	59.24%	0.26%	95,14%	0,07%	69,52%	0,32%
BLS	80.00%	0.04%	22.22%	0.04%	100,00%	0,00%	33,33%	0,05%
BOR	100.00%	0.00%	34.56%	0.49%	85,71%	0,04%	47,56%	0,52%
CON	90.32%	0.09%	57.28%	0.35%	92,41%	0,16%	65,01%	0,35%
MAC	50.00%	0.15%	25.00%	0.25%	66,67%	0,14%	40,01%	0,25%
MED	90.71%	0.05%	56.78%	0.44%	92,24%	0,18%	73,81%	0,28%
PAN	96.00%	0.05%	43.64%	0.12%	94,59%	0,05%	64,82%	0,11%
STE	100.00%	0.00%	30.00%	0.07%	75,00%	0,06%	50,00%	0,07%
AL+ME+MK+RS+XK	100.00%	0.00%	43.24%	0.08%	100,00%	0,00%	55,81%	0,08%
AT + CH + LI	94.59%	0.07%	66.04%	0.14%	97,62%	0,04%	66,13%	0,14%
BA + HR + SI	85.71%	0.07%	58.06%	0.10%	84,00%	0,07%	63,64%	0,10%
BE + LU+ NL + DK	94.59%	0.13%	66.67%	0.28%	95,65%	0,12%	71,93%	0,35%
BG	75.00%	0.07%	50.00%	0.08%	100,00%	0,00%	60,00%	0,08%
CZ + SK	86.49%	0.08%	66.67%	0.12%	85,71%	0,09%	80,00%	0,10%
DE	93.66%	0.11%	63.78%	0.35%	94,45%	0,10%	79,17%	0,29%
EE + LT + LV	100.00%	0.00%	37.50%	0.08%	66,67%	0,08%	46,15%	0,09%
EL	87.50%	0.05%	58.33%	0.08%	93,75%	0,04%	75,00%	0,07%
ES	88.89%	0.04%	50.46%	0.39%	89,47%	0,04%	80,32%	0,05%
FI	100.00%	0.00%	36.00%	0.06%	83,33%	0,05%	55,56%	0,06%
FR	93.24%	0.07%	52.33%	0.31%	92,38%	0,24%	60,41%	0,34%
HU	93.33%	0.06%	38.89%	0.11%	91,67%	0,06%	62,86%	0,11%
IE + UK	93.52%	0.08%	63.25%	0.37%	95,20%	0,07%	73,23%	0,34%
IS	100.00%	0.00%	100.00%	0.00%	100,00%	0,00%	100,00%	0,00%
IT	96.90%	0.06%	56.03%	0.51%	93,38%	0,25%	66,55%	0,43%
NO	100.00%	0.00%	45.00%	0.03%	100,00%	0,00%	54,17%	0,03%
PL	78.46%	0.10%	55.44%	0.12%	84,72%	0,09%	46,32%	0,40%
PT	96.55%	0.06%	57.15%	0.15%	100,00%	0,00%	79,59%	0,11%
RO	100.00%	0.00%	35.82%	0.08%	96,77%	0,03%	45,91%	0,40%
SE	100.00%	0.00%	29.36%	0.64%	95,32%	0,02%	49,58%	0,68%
TR	79.17%	0.04%	47.62%	0.43%	90,00%	0,03%	57,31%	0,29%

### 4.3. IMD2018 plausibility analysis correspondence analysis

One reason for the lower correspondences results from the blind interpretations is the arbitrary decision between urban and non-urban being imposed by the 30% threshold. Therefore, a point where the reference data is 30 % and the map density is 29 % will not correspond and vice versa. To cope with this situation and incorporate a level of plausibility onto the results the plausibility reference data were used where the interpreter was allowed to assess whether the mapped data was plausible.

The results in Table 4-4 using the plausibility analysis therefore represent a more realistic assessment of the thematic accuracy of the IMD product for 2018. They show considerable improvement in the correspondence results compared to the target accuracies of 90 % and the results of the blind interpretation (Table 4-3) as expected.

The 2018 plausibility-based correspondence results, when compared to 2015, show an increase in thematic accuracy for over 85 % of the regionalisations. For the individual user's and producer's accuracies, 57 out of 68 regionalisation exceeded the 90% target with a further 5 falling only 5 % short. This is a considerable increase in performance over the 2015 IMD product where only 41 regionalisations exceed the target. The majority of the improvement was focused on the producer's accuracy which indicates far less errors of omission in the 2018 IMD product. The regionalisations falling short all had producer's accuracies that were greater than 70% compared to some falling as low as 50 % in 2015.

The range of changes in thematic accuracy between 2015 and 2018 are from -31 % to 50 % with a large amount of variability and with producer's accuracy dominating as was identified above. Many of the negative changes in thematic accuracy were associated with regionalisations which reported at or near 100% in 2015, suggesting that this may be more of a correction representativeness of the accuracy assessment.

There has been a considerable improvement of the overall performance of the IMD layer in 2018 with the both European level user's and producer's accuracies exceeding the target of 90%. More importantly the improvement of producer's accuracy across all regionalisation shows a step-up inconsistency and reduction in the level of omissions. The main reason of the improvement of user's and producer's accuracies is related to the enhanced spatial resolution from 2015 to 2018. Indeed, the 2015 production was mainly based on Landsat-8 data with a spatial resolution of 30m, whereas for 2018 Sentinel-2 data with a spatial resolution of 10m was used as main input. It results in a 9-fold improvement in spatial resolution which leads to a greater detection of sealed surfaces.

Table 4-4: A summary of the correspondence results for the plausibility analysis for the 2015 and 2018 results. Green cells exceed the 90 % minimum accuracy requirement, orange cells are within 5% and red cells fall more than 5% short.

	2015				2018			
	User	CI95%	Prod	CI95%	User	CI95%	Prod	CI95%
European All	95.41%	0.04%	84.50%	0.17%	97.50%	0,03%	91.37%	0,17%
ALP	100.00%	0.00%	92.86%	0.02%	95,24%	0,02%	83,33%	0,03%
ANA	78.57%	0.03%	88.00%	0.02%	97,37%	0,01%	92,50%	0,01%
ARC	100.00%	0.00%	100.00%	0.00%	100,00%	0,00%	100,00%	0,00%
ATL	97.76%	0.05%	82.40%	0.20%	97,33%	0,05%	92,24%	0,22%
BLS	80.00%	0.04%	80.00%	0.04%	100,00%	0,00%	70,00%	0,05%
BOR	100.00%	0.00%	71.52%	0.49%	100,00%	0,00%	87,50%	0,04%
CON	94.79%	0.07%	87.95%	0.18%	97,76%	0,04%	92,09%	0,16%
MAC	100.00%	0.00%	50.00%	0.23%	100,00%	0,00%	100,00%	0,00%
MED	93.68%	0.04%	82.09%	0.07%	97,05%	0,03%	91,56%	0,18%
PAN	100.00%	0.00%	89.29%	0.07%	100,00%	0,00%	92,50%	0,06%
STE	100.00%	0.00%	100.00%	0.00%	100,00%	0,00%	100,00%	0,00%
AL+ME+MK+RS+XK	100.00%	0.00%	80.00%	0.06%	100,00%	0,00%	80,00%	0,06%
AT + CH + LI	100.00%	0.00%	92.50%	0.08%	97,62%	0,04%	87,23%	0,10%
BA + HR + SI	85.71%	0.07%	90.00%	0.06%	92,00%	0,05%	88,46%	0,07%
BE + LU+ NL + DK	98.65%	0.07%	89.02%	0.19%	95,65%	0,12%	94,62%	0,13%
BG	100.00%	0.00%	80.00%	0.06%	100,00%	0,00%	92,31%	0,04%
CZ + SK	91.89%	0.07%	91.89%	0.07%	97,62%	0,04%	100,00%	0,00%
DE	95.12%	0.09%	83.57%	0.25%	97,62%	0,07%	94,86%	0,20%
EE + LT + LV	100.00%	0.00%	100.00%	0.00%	100,00%	0,00%	69,23%	0,08%
EL	87.50%	0.05%	72.41%	0.07%	96,88%	0,03%	96,88%	0,03%
ES	93.83%	0.03%	71.70%	0.06%	95,61%	0,03%	91,60%	0,04%
FI	100.00%	0.00%	81.82%	0.05%	100,00%	0,00%	92,31%	0,03%
FR	95.95%	0.06%	86.59%	0.10%	95,22%	0,06%	92,76%	0,07%
HU	100.00%	0.00%	88.24%	0.07%	100,00%	0,00%	96,00%	0,05%
IE + UK	97.22%	0.05%	79.94%	0.29%	98,63%	0,04%	89,59%	0,35%
IS	100.00%	0.00%	100.00%	0.00%	100,00%	0,00%	100,00%	0,00%
IT	99.22%	0.03%	90.78%	0.10%	98,72%	0,04%	89,56%	0,10%
NO	100.00%	0.00%	64.29%	0.02%	100,00%	0,00%	92,86%	0,01%
PL	89.23%	0.07%	93.55%	0.06%	98,61%	0,03%	95,95%	0,05%
PT	100.00%	0.00%	85.30%	0.10%	100,00%	0,00%	97,50%	0,05%
RO	100.00%	0.00%	96.00%	0.03%	100,00%	0,00%	79,49%	0,06%
SE	100.00%	0.00%	62.65%	0.66%	100,00%	0,00%	91,43%	0,03%
TR	86.11%	0.03%	86.11%	0.03%	97,78%	0,01%	83,54%	0,33%

## 5. Thematic Accuracy SBU2018

The thematic accuracy of the SBU2018 layer is assessed using a number of approaches and within a number of regionalisations to better understand the thematic characteristics across biogeographic regions and countries.

The assessment was based on a sample set of 18,005 reference points.

### 5.1. SBU2018 scatterplots & regression analysis

Scatterplots are again used to display data against Cartesian coordinates to show any trends or relationships in the results at the PSU level. The three available measurements for SBU at each PSU are the original reference data from blind analysis (CODE), the reference data after plausibility analysis (QC) and the mapped value from the product (MAP). The reference (CODE or QC) information was presented on horizontal axis and product (MAP) information on the vertical axis. The closeness of a point to the 1:1 line is an indication of the similarity between the reference and mapped results and deviations show issues with the calibration or offsets.

The scatterplots for SBU2018 indicate the increased uncertainties when recording the share of built-up in 10 m EO data and then aggregating this to a 1 ha grid compared to the IMD scatterplots in section 4.1. The distribution of points displayed on Figure 5-1, Figure 5-2 and Figure 5-3 have considerable scatter about the best fit line. The plots for both blind interpretation and plausibility analysis shows a large number of points exceeding a difference of 20 % each side of the best fit line. The level of scatter seems to be uniform over the range of share of built-up values. There are some points on the x and y axes showing commission, where built-up is mapped that is not present in reality, and omission, where built-up areas are missed.

The distribution of points for the whole dataset on Figure 5-1 is not centred on the 1:1 line but falls above it indicating that the mapped products are overestimating the actual share of built-up as reported by the reference data. In the case of the plausibility analysis the spread of points around the best fit line is very similar with only the extreme values removed, particularly commission errors. In fact, the best fit line for the plausibility plot is further from the 1:1 line suggesting uncertainties in the blind analysis compensated for the overestimation. There also appears to be a tendency for the map product to saturate at 100% built-up where the reference data shows less which also supports the conclusion that the product overestimates the share of built-up.

The conclusions for the whole dataset (ALL) are not repeated consistently across the different regions and between the blind interpretation and the plausibility analysis. The Black Sea and Atlantic BRME results are included for comparative purposes with the IMD results and as expected the scatterplots and the resulting best fit lines are controlled by the actual geography of the regions being considered. The Figure 5-2 results for the Black Sea BRME zone show that there are limited numbers of built-up areas in this region and the points present tend to have a lower share of built-up. The relationship between the reference and mapped data is visually quite reasonable, but with a small number of outliers. However, the  $R^2$  is low and the best fit line shows a stronger overestimation than the whole dataset. The outliers tend to be mapped with greater than 50 % share of built-up, but all the reference data is below 50 % share of built-up. The Atlantic BRME zone (Figure 5-3) is more representative of Europe as a whole and should contain a broad range of building types / patterns and thus share of built-up. The distribution of points appears tighter around the best fit line than the results for the whole dataset with a greater spread towards to higher values of share of built-up. The best fit line for the plausibility analysis is very close to the 1:1 line and in fact slightly below, showing a possible underestimation by the product, but not significantly so. There appears to be limited number of points with the share of built-greater than 70 % which may be an artefact of the generalisation to 1 ha as few building, with the exceptions of industrial complexes, have footprints great than 100 m on both sides.

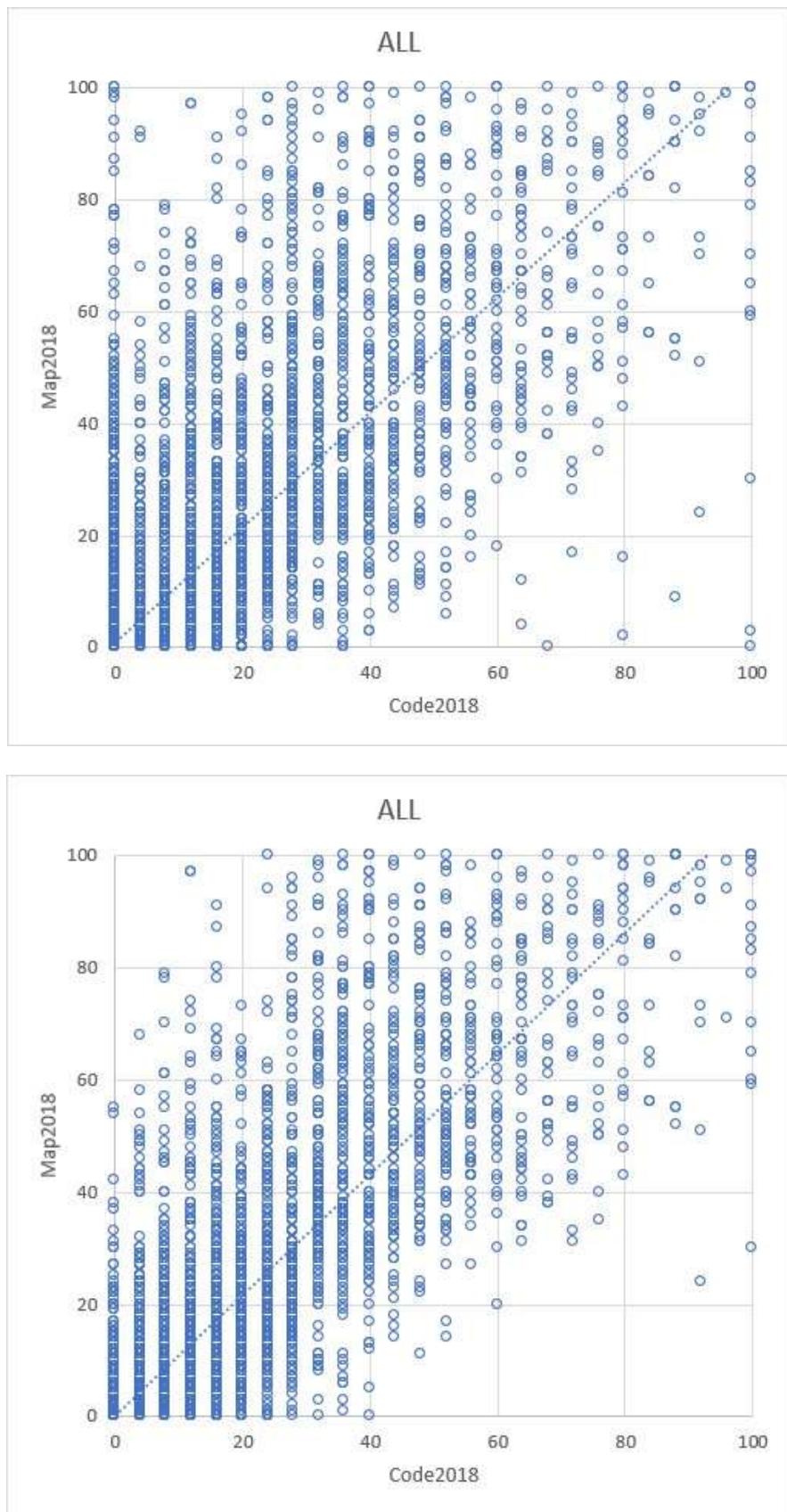


Figure 5-1: Scatterplots for 2018 for ALL the reference data for SBU for both the blind interpretation (TOP) and plausibility analysis (BOTTOM).

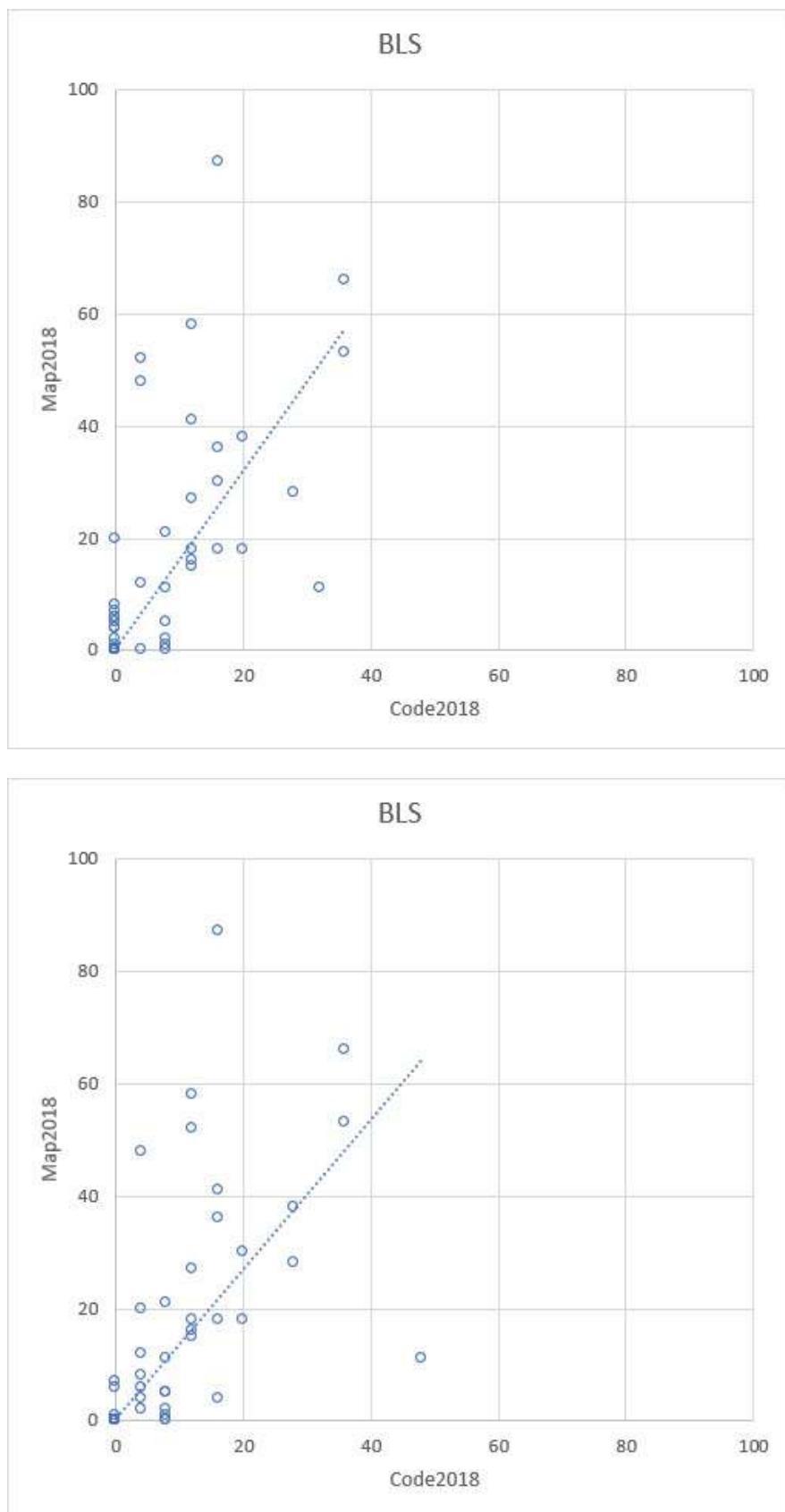


Figure 5-2: Scatterplots for 2018 for BLS the reference data for SBU for both the blind interpretation (TOP) and plausibility analysis (BOTTOM).

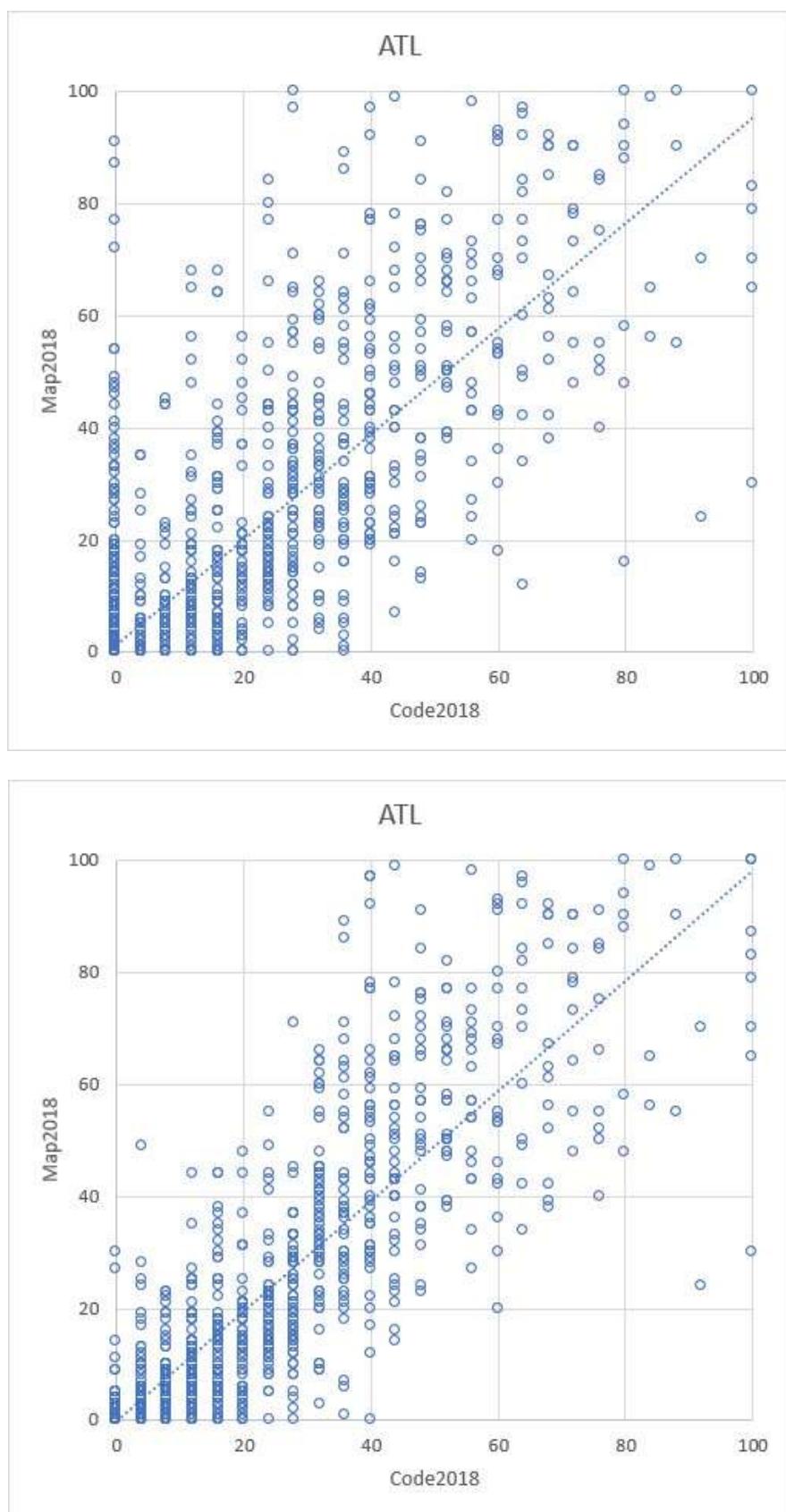


Figure 5-3: Scatterplots for 2018 for ATL the reference data for SBU for both the blind interpretation (TOP) and plausibility analysis (BOTTOM).

The regression analysis results for the blind and plausibility interpretations for all regionalisations are given in Table 5-1. As would be expected the plausibility results in general show improvements in  $R^2$  and slope over the blind results, but this is not as dramatic as that seen in the IMD layers analysis suggesting there are uncertainties remaining in the interpretation of the share of build-up against reference imagery. Two of the plausibility results, Boreal and Black Sea BRMEs, actually show a decrease in  $R^2$  and only two thirds of the best fit line slopes move closer to 1.0. The plausibility interpretations obviously remove some of the more extreme discrepancies between the reference data and the product, particularly commission errors, but underlying uncertainties remain.

For the plausibility results the  $R^2$  range from 0.54 to 1.00, but the later refers to the Iceland regionalisation where only one point is non-zero and the slope of 0.65 shows a dramatic underestimation for the point in question. The second highest  $R^2$  is 0.99 for the Arctic BMRE which is representing almost the same data. The third highest  $R^2$  is more reasonable at 0.90 is for the Steppic BRME and has a corresponding slope of 1.23 matching the general overestimation of the share of built-up by the product.

The slopes from the plausibility interpretation analysis range from 0.64, a strong underestimation, to 2.33, a strong overestimation. This range and the relatively even spread of the differing slope values between regionalisations suggests that the calibration of the product is significantly variable spatially. The two regionalisations with slopes below 0.95 are associated with arctic with very few built-up points in these areas. The most significant overestimation is related to Bulgaria where the results fit a consistent pattern, but a small group of points seem to have almost 3 times the mapped share of built-up compared to the reference data. This would suggest that there are one or more confounding factors which the production methodology is not coping with and that they may be location specific.

*Table 5-1: A summary of regression line parameters for the SBU scatterplots derived from the blind and plausibility interpretation as reference data for 2018.*

	Blind			Plausibility		
	R <sup>2</sup>	Slp	Int	R <sup>2</sup>	Slp	Int
European All	0.65	1.03	1.13	0.77	1.07	0.39
ALP	0.58	0.92	0.50	0.72	0.98	0.12
ANA	0.70	2.03	0.41	0.72	1.73	0.22
ARC	1.00	0.65	0.01	0.99	0.64	-0.01
ATL	0.70	0.94	1.38	0.82	0.98	-0.03
BLS	0.57	1.58	0.56	0.54	1.33	0.49
BOR	0.96	1.26	0.33	0.76	1.24	0.03
CON	0.70	0.97	1.25	0.80	1.04	0.43
MAC	0.27	0.59	1.29	0.69	1.40	-0.86
MED	0.58	1.20	1.85	0.72	1.26	0.97
PAN	0.76	1.19	1.33	0.82	1.22	0.55
STE	0.85	2.93	0.31	0.90	1.23	0.08
AL+ME+MK+RS+XK	0.80	1.30	0.90	0.86	1.33	0.30
AT + CH + LI	0.63	1.30	1.42	0.79	1.11	0.37
BA + HR + SI	0.85	1.26	0.44	0.86	1.25	-0.22
BE + LU+ NL + DK	0.62	0.94	2.96	0.77	1.02	0.51
BG	0.74	2.49	0.92	0.82	2.33	0.14
CZ + SK	0.78	1.26	1.21	0.82	1.30	0.57
DE	0.74	0.90	1.08	0.85	0.98	0.29
EE + LT + LV	0.71	1.62	0.22	0.73	1.51	-0.13
EL	0.58	1.71	1.89	0.67	1.70	1.18
ES	0.52	1.08	1.40	0.74	1.14	0.54
FI	0.79	1.36	0.35	0.87	1.29	0.02
FR	0.76	0.96	1.00	0.84	1.01	0.11
FR DOMs						
HU	0.76	1.26	0.76	0.80	1.28	0.03
IE + UK	0.63	0.94	1.66	0.82	0.97	-0.17
IS	1.00	0.65	0.00	1.00	0.65	0.00
IT	0.60	0.83	1.47	0.72	0.95	0.65
NO	0.65	0.99	0.15	0.83	0.97	-0.01
PL	0.73	1.15	0.84	0.81	1.18	0.08
PT	0.57	0.94	2.27	0.64	0.98	1.32
RO	0.63	1.36	1.83	0.80	1.19	0.58
SE	0.73	1.10	0.22	0.78	1.11	0.04
TR	0.65	1.89	0.69	0.68	1.60	0.53

## 5.2. SBU2018 blind and plausibility interpretation correspondence analysis

For each of the regionalisations, the correspondence analysis (user's and producer's accuracies) and the 95 % confidence intervals were calculated. As no threshold was supplied for the SBU2018 layer two separate correspondence analyses with thresholds share of built-up of 1 % and 30 % were performed.

The results for the 1 % threshold for the blind and plausibility interpretations are given in Table 5-2. As with the IMD results, the overall accuracies, which were not part of the product specification and exceeded the target thematic accuracy 90% (not shown for clarity). Also, the discussion of the results will focus on the plausibility analysis as there is an obvious improvement relative to the blind analysis and the issues of built-up uncertainty have already been raised.

For the threshold of 1 %, the user's and producer's accuracies for the whole dataset were slightly higher than the target of 90 %. More than half of the results for the regionalisations achieved the required thematic accuracy, with almost 90 % passing when those within 5 % of the threshold were considered. There is no obvious pattern to the regionalisations that fail to reach the target accuracies. The thematic accuracies tend to be evenly balanced between user's and producer's accuracy suggesting there are similar levels of omission and commission errors, but as the IMD layer served as a mask the omission errors may have therefore been overestimated. A more detailed investigation of the impact of the IMD layer mask should be undertaken.

The results for the 30 % threshold of share of built-up for the blind and plausibility interpretations are given in Table 5-3. The user's and producer's accuracies for the whole dataset did not reach the required target accuracy, but the producer's accuracy fell within 5% of it. Just over a third of the regionalisations achieved the required thematic accuracy, with only half achieving it when those within 5 % of the target were considered. The higher thematic accuracies tend to be focused on the producer's accuracy suggesting there are more commission errors.

*Table 5-2: A summary of the correspondence results for the blind and plausibility interpretation for the SBU > 1% results. Green cells exceed the 90 % minimum accuracy requirement, orange cells are within 5% and red cells fall more than 5% short.*

	Blind				Plausibility			
	User	CI95%	Prod	CI95%	User	CI95%	Prod	CI95%
European All	63.51%	0.51%	86.83%	0.43%	90.55%	0.26%	91.01%	0.36%
ALP	61.46%	0.47%	82.95%	0.39%	89.50%	0.24%	87.63%	0.33%
ANA	57.04%	0.47%	83.64%	0.40%	85.25%	0.02%	88.43%	0.34%
ARC	50.00%	0.01%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
ATL	62.23%	0.55%	89.42%	0.36%	93.22%	0.26%	93.55%	0.27%
BLS	70.59%	0.05%	92.31%	0.03%	91.18%	0.03%	93.94%	0.02%
BOR	49.92%	0.60%	74.69%	0.64%	86.01%	0.29%	83.57%	0.54%
CON	68.31%	0.49%	89.64%	0.38%	92.81%	0.22%	93.08%	0.33%
MAC	66.66%	0.14%	57.15%	0.23%	100.00%	0.00%	75.01%	0.19%
MED	61.60%	0.47%	82.56%	0.52%	85.22%	0.31%	87.01%	0.46%
PAN	69.44%	0.44%	91.90%	0.36%	88.09%	0.32%	93.50%	0.32%
STE	73.34%	0.07%	100.00%	0.00%	93.33%	0.04%	100.00%	0.00%
AL+ME+MK+RS+XK	62.50%	0.07%	77.38%	0.60%	84.37%	0.06%	82.20%	0.54%
AT + CH + LI	60.16%	0.50%	94.45%	0.05%	90.85%	0.30%	96.25%	0.04%
BA + HR + SI	53.00%	0.59%	96.53%	0.04%	90.15%	0.32%	97.93%	0.03%
BE + LU + NL + DK	63.38%	0.48%	97.18%	0.09%	94.49%	0.13%	98.09%	0.07%
BG	51.24%	0.60%	100.00%	0.00%	80.36%	0.45%	100.00%	0.00%
CZ + SK	63.41%	0.57%	92.90%	0.36%	95.47%	0.05%	95.17%	0.30%
DE	77.77%	0.44%	96.22%	0.16%	94.61%	0.16%	97.73%	0.15%
EE + LT + LV	61.39%	0.08%	81.56%	0.45%	86.64%	0.06%	86.19%	0.39%
EL	59.69%	0.62%	91.49%	0.04%	85.64%	0.38%	93.91%	0.03%
ES	49.07%	0.53%	80.57%	0.48%	77.87%	0.32%	87.70%	0.40%
FI	41.22%	0.56%	95.24%	0.02%	84.54%	0.04%	97.62%	0.01%
FR	62.22%	0.58%	86.73%	0.46%	93.31%	0.26%	92.48%	0.36%
HU	73.54%	0.39%	90.51%	0.42%	89.41%	0.38%	92.06%	0.38%
IE + UK	58.36%	0.59%	90.82%	0.23%	91.65%	0.34%	93.95%	0.19%
IS	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
IT	75.97%	0.44%	81.86%	0.51%	91.64%	0.29%	84.48%	0.48%
NO	63.89%	0.02%	48.19%	0.83%	86.11%	0.02%	55.62%	0.82%
PL	67.55%	0.42%	80.33%	0.54%	94.46%	0.05%	87.37%	0.45%
PT	71.05%	0.14%	91.09%	0.07%	93.05%	0.07%	93.05%	0.06%
RO	50.37%	0.48%	95.65%	0.03%	85.74%	0.40%	97.40%	0.02%
SE	50.26%	0.73%	65.04%	0.71%	85.79%	0.43%	76.05%	0.61%
TR	64.17%	0.39%	82.04%	0.49%	87.47%	0.03%	86.16%	0.44%

*Table 5-3: A summary of the correspondence results for the blind and plausibility interpretation for the SBU > 30% results. Green cells exceed the 90 % minimum accuracy requirement, orange cells are within 5% and red cells fall more than 5% short.*

	Blind				Plausibility			
	User	CI95%	Prod	CI95%	User	CI95%	Prod	CI95%
European All	55.36%	0.10%	78.91%	0.18%	69.22%	0.13%	88.88%	0.16%
ALP	49.02%	0.04%	56.82%	0.04%	68.63%	0.03%	77.78%	0.03%
ANA	17.14%	0.03%	100.00%	0.00%	28.57%	0.03%	100.00%	0.00%
ARC	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
ATL	67.43%	0.24%	73.96%	0.23%	88.03%	0.11%	82.64%	0.27%
BLS	20.00%	0.04%	66.67%	0.05%	20.00%	0.04%	66.67%	0.05%
BOR	40.59%	0.49%	90.00%	0.04%	57.16%	0.06%	92.69%	0.03%
CON	61.87%	0.21%	81.24%	0.21%	71.51%	0.21%	93.38%	0.07%
MAC	100.00%	0.00%	50.00%	0.15%	100.00%	0.00%	100.00%	0.00%
MED	42.34%	0.09%	87.22%	0.06%	53.65%	0.09%	94.84%	0.04%
PAN	56.10%	0.12%	92.00%	0.06%	63.41%	0.11%	92.86%	0.06%
STE	0.00%	0.00%	N/A	N/A	80.00%	0.06%	100.00%	0.00%
AL+ME+MK+RS+XK	56.67%	0.08%	94.44%	0.03%	60.00%	0.07%	94.74%	0.03%
AT + CH + LI	44.12%	0.15%	88.24%	0.09%	79.41%	0.12%	96.43%	0.05%
BA + HR + SI	57.69%	0.10%	100.00%	0.00%	65.38%	0.10%	100.00%	0.00%
BE + LU + NL + DK	68.12%	0.28%	73.44%	0.26%	84.06%	0.22%	81.69%	0.23%
BG	20.00%	0.06%	100.00%	0.00%	20.00%	0.06%	100.00%	0.00%
CZ + SK	44.74%	0.12%	80.95%	0.10%	52.63%	0.12%	90.91%	0.07%
DE	82.04%	0.17%	81.07%	0.17%	91.02%	0.12%	89.94%	0.13%
EE + LT + LV	23.08%	0.07%	100.00%	0.00%	38.46%	0.08%	100.00%	0.00%
EL	34.48%	0.07%	90.91%	0.04%	41.38%	0.07%	92.31%	0.04%
ES	48.10%	0.07%	84.44%	0.05%	60.76%	0.07%	94.12%	0.03%
FI	44.96%	0.77%	85.71%	0.04%	85.01%	0.04%	91.90%	0.03%
FR	67.53%	0.13%	77.04%	0.12%	78.57%	0.12%	90.98%	0.08%
HU	58.33%	0.11%	87.50%	0.08%	62.50%	0.11%	88.24%	0.07%
IE + UK	60.33%	0.32%	72.47%	0.34%	91.62%	0.09%	79.21%	0.38%
IS	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%	100.00%	0.00%
IT	61.11%	0.17%	63.86%	0.34%	71.30%	0.15%	87.50%	0.11%
NO	33.33%	0.02%	75.00%	0.02%	66.67%	0.02%	100.00%	0.00%
PL	50.77%	0.12%	94.29%	0.06%	60.00%	0.12%	97.50%	0.04%
PT	45.16%	0.16%	77.78%	0.13%	54.84%	0.16%	89.47%	0.10%
RO	17.50%	0.06%	87.50%	0.05%	50.00%	0.07%	86.96%	0.05%
SE	42.16%	0.67%	81.82%	0.03%	42.16%	0.67%	81.82%	0.03%
TR	21.59%	0.04%	95.00%	0.02%	30.68%	0.04%	96.43%	0.02%

## 6. Thematic Accuracy IMC1518

The thematic accuracy of the IMC1518 layer is assessed using a correspondence analysis within a number of regionalisations to better understand the thematic characteristics across biogeographic regions and countries. For the thematic accuracy assessment the change values were simplified to:

- 0 – No change with zero imperviousness
- 1 – Change in imperviousness
- 10 – No change with imperviousness greater than zero

The user's and producer's accuracies were then calculated for class 1, change in imperviousness, and the overall accuracy was not reported as it would be unrepresentative, being dominated by no change. The assessment was based on a sample set of 10,000 reference points.

From Table 6-1, it can be seen that for the blind analysis only a single regionalisation achieved a thematic accuracy above the target of 90 % and it only contained a single changed point. The remainder of the discussion will therefore focus on the plausibility results. For the whole dataset the thematic accuracies do not reach the target of 90 % and only the producer's accuracy fell within 5 % of the target. Only a third of the plausibility results reached the required target accuracy of 90%, mainly producer's accuracies, and this was still less than half when those within 5% of the target were included. Many of the regionalisations which exceed the threshold are at 100% correspondence suggesting there are only a small number of change points present relative to the sample size. Also, these regionalisations tended to more unique compared to the overall biogeography of Europe. The accuracies below the target were still quite variable with no obvious pattern to those with relatively higher or lower values.

In comparison to IMC1215, the results for IMC1518 show a slight improvement in the numbers of regionalisations reaching the target accuracy, but there is a clear shift from user's to producer's accuracy suggesting there are less omission errors this time.

*Table 6-1: A summary of the correspondence results for the blind and plausibility interpretation for the IMC2018 results. Green cells exceed the 90 % minimum accuracy requirement, orange cells are within 5% and red cells fall more than 5% short.*

	Blind				Plausibility			
	User	CI95%	Prod	CI95%	User	CI95%	Prod	CI95%
European All	37.48%	0.06%	10.80%	0.68%	51.01%	0.04%	85.08%	0.07%
ALP	49.15%	0.01%	4.25%	0.76%	61.14%	0.01%	70.14%	0.03%
ANA	36.54%	0.01%	7.08%	0.73%	40.23%	0.01%	98.05%	0.01%
ARC	0.00%	0.01%	0.00%	0.00%	100.00%	0.00%	100.00%	0.00%
ATL	42.97%	0.08%	15.75%	0.59%	56.57%	0.03%	85.89%	0.12%
BLS	31.81%	0.01%	2.78%	0.81%	34.59%	0.01%	100.00%	0.00%
BOR	40.29%	0.05%	15.01%	0.57%	77.88%	0.01%	89.79%	0.04%
CON	38.36%	0.05%	8.63%	0.65%	50.35%	0.02%	83.44%	0.10%
MAC	0.00%	0.00%	0.00%	0.07%	0.00%	0.00%	N / A	N / A
MED	29.98%	0.06%	12.79%	0.76%	38.24%	0.06%	83.37%	0.06%
PAN	30.93%	0.11%	6.94%	0.07%	64.64%	0.01%	100.00%	0.00%
STE	100.00%	0.00%	12.90%	0.04%	100.00%	0.00%	100.00%	0.00%
AL+ME+MK+RS+XK	37.26%	0.07%	11.41%	0.05%	78.05%	0.01%	100.00%	0.00%
AT + CH + LI	66.28%	0.01%	16.48%	0.14%	71.09%	0.01%	52.20%	0.15%
BA + HR + SI	28.57%	0.01%	2.74%	0.77%	28.57%	0.01%	66.67%	0.01%
BE + LU + NL + DK	19.21%	0.03%	4.03%	0.16%	39.97%	0.04%	70.71%	0.26%
BG	15.83%	0.01%	2.06%	0.02%	35.28%	0.01%	100.00%	0.00%
CZ + SK	35.88%	0.01%	3.41%	0.75%	37.43%	0.01%	44.66%	0.06%
DE	20.82%	0.11%	6.53%	0.56%	39.74%	0.02%	84.57%	0.15%
EE + LT + LV	35.43%	0.09%	20.82%	0.07%	83.24%	0.01%	99.45%	0.01%
EL	45.86%	0.02%	11.19%	0.79%	46.82%	0.02%	99.29%	0.01%
ES	7.57%	0.06%	6.30%	0.06%	13.14%	0.06%	100.00%	0.00%
FI	82.18%	0.00%	24.68%	0.06%	98.83%	0.01%	100.00%	0.00%
FR	32.59%	0.02%	5.14%	0.76%	42.08%	0.02%	69.70%	0.13%
HU	57.14%	0.01%	7.65%	0.07%	57.14%	0.01%	100.00%	0.00%
IE + UK	63.69%	0.07%	41.80%	0.55%	74.74%	0.02%	100.00%	0.00%
IS	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	100.00%	0.00%
IT	37.79%	0.02%	18.97%	0.54%	40.73%	0.02%	86.55%	0.11%
NO	19.74%	0.01%	9.99%	0.01%	40.79%	0.01%	100.00%	0.00%
PL	46.64%	0.02%	6.07%	0.46%	66.19%	0.02%	100.00%	0.00%
PT	26.61%	0.15%	22.62%	0.15%	41.61%	0.12%	100.00%	0.00%
RO	67.29%	0.01%	10.10%	0.10%	68.72%	0.01%	100.00%	0.00%
SE	29.45%	0.04%	10.15%	0.74%	67.13%	0.01%	78.39%	0.04%
TR	50.39%	0.02%	12.60%	0.82%	63.01%	0.01%	77.55%	0.03%

## 7. Conclusions and recommendations

This report represents a thorough examination of the IMD, SBU and IMC layers for the 2018 reference year. It provides an independent validation assessment of the performance of the products against their design specifications with blind interpretations and plausibility analyses putting clear boundaries on the thematic quality of the products being made available to the wider community.

The review of the specifications of the products in chapter 3 shows a very close alignment to the design requirements. There are still a few short comings with the products including the documentation and metadata which have been noted. The map data currently does not include details on the image data used for the production (e.g. time stamps, sensor, ancillary information etc.) which would be extremely useful when trying to interpret the results of studies such as this. This information is supplied by the service provider, but not made available as part of the standard product.

The scatterplots for the IMD product presented in section 4.1 show the overall situation for the layers across Europe. The majority of the points analysed are within 20 % of the best fit line, but there is some variability in the distribution depending on the actual sealing level. From the regression results in section 4.1 it can be seen that for the larger regionalisations the coefficient of determination tends to be around 0.88 and 0.90, but there is variability between biogeographical regions and countries. Even though the regression results show a continued underestimation of the of imperviousness by the IMD layer compared to the reference data, in comparison to 2015 the calibration has been improved and there is great consistency across the various regionalisations. At the European level the R<sup>2</sup> is 0.88 with a slope of 0.8 and reduced ranges for the individual regionalisation values.

The thematic accuracy results used an arbitrary 30 % threshold for the conversion of IMD and reference data to urban areas. The blind interpretation (section 4.2) produced results that did not fully meet the requirements of 90 % minimum thematic accuracy in the specification, but the plausibility results (section 4.3) with only 11 of the 64 test not exceeding the target and 5 of those within 5%. The results are therefore fully in line with specifications for commission errors, but are slightly outside the specifications for omission errors. There has been a considerable improvement of the overall performance of the IMD layer in 2018 with the both European level user's and producer's accuracies exceeding the target of 90%. More importantly the improvement of producer's accuracy across all regionalisation shows a step-up inconsistency and reduction in the level of omissions.

Overall, there has been an improvement in the performance of the IMD2018 layer against 2015 for both blind and plausibility interpretations relative to reference data related to the improved spatial resolution from 2015 to 2018 which leads to a better detection of sealed surfaces (4-fold improvement in spatial resolution). It is recommended that in a future Specific Contract, an investigation is undertaken into the effect of using a 5 x 5 sampling grid for the SSUs (1 ha) for the HRL IMD 2018 product, which is now based on a 10 m spatial resolution.

The scatterplots and correspondence analyses for the SBU2018 layer in section 5 represent the first time these layers have been produced and therefore validated. The scatterplots and regression analyses show considerable variability within the product overall and between the different regionalisations. There is a tendency to overestimate the share of built-up, but this is not consistent across all regionalisations and some sparsely built-up areas appear to actually underestimate the share of built-up. The whole dataset reaches the required accuracy target when a built-up threshold of 1 % is selected, but that may not be realistic as it will increase commissions. These differences in results between the regionalisations could be related to the differences in building / roof types and / or urban patterns, but would require further examinations. The thematic accuracies tend to be evenly balanced between user's and producer's accuracy suggesting there are similar levels of omission and commission errors, but as the IMD layer served as a mask the omission errors may have therefore been overestimated. A more detailed investigation of the impact of the IMD layer mask should be undertaken.

Overall, the IMC1518 results are similar to those for IMC1215, however there has generally been an increase in producer's accuracy, seemingly at the expense of user's accuracy. The producer's accuracy for the whole data set fell only 5% short of the target, but the user's accuracy was only just over 50%. The presence of some

correspondence results of 100% are likely to be a consequence of small numbers of change points within specific regionalisations with only small numbers of samples which are all mapped correctly.