

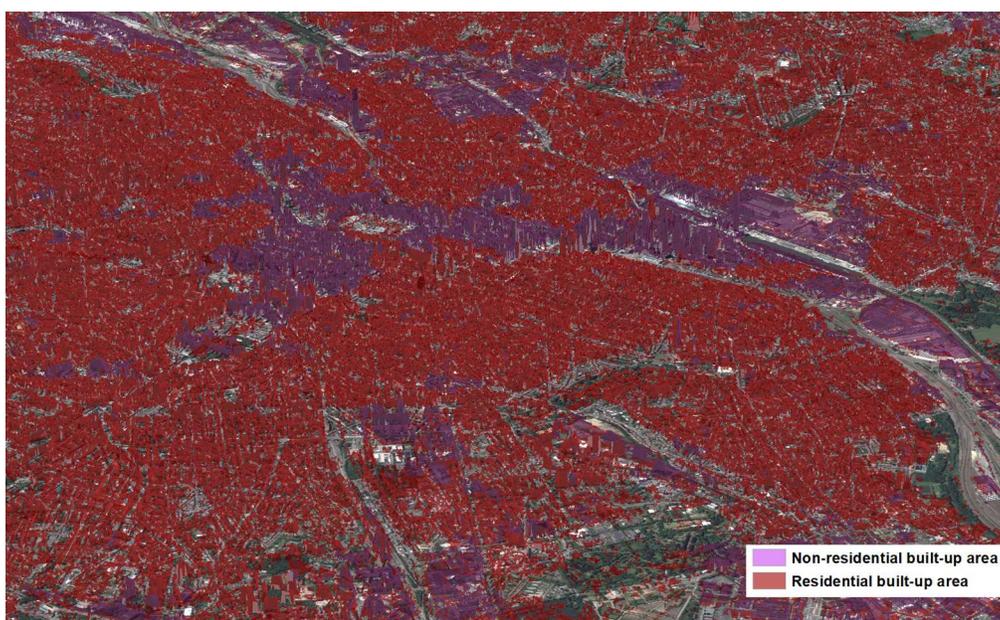


JRC TECHNICAL REPORTS

The European Settlement Map 2019 release

*Application of the Symbolic
Machine Learning to
Copernicus VHR Imagery*

Sabo, F., Corbane, C., Politis, P., Kemper, T.



This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication.

Contact information

Name: Christina Corbane

Address: European Commission, Joint Research Centre, Space, Security and Migration (Ispra), Disaster Risk Management (JRC.E.1)

Email: christina.corban@ec.europa.eu

Tel.: +39 0332 78 3545

EU Science Hub

<https://ec.europa.eu/jrc>

JRC118076

EUR 29886 EN

PDF	ISBN 978-92-76-12184-8	ISSN 1831-9424	doi:10.2760/25824
Print	ISBN 978-92-76-12183-1	ISSN 1018-5593	doi:10.2760/979189

Luxembourg: Publications Office of the European Union, 2019

© European Union, 2019

The reuse policy of the European Commission is implemented by Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Reuse is authorised, provided the source of the document is acknowledged and its original meaning or message is not distorted. The European Commission shall not be liable for any consequence stemming from the reuse. For any use or reproduction of photos or other material that is not owned by the EU, permission must be sought directly from the copyright holders.

All content © European Union, 2019

Contents

Abstract	2
1 Introduction	4
2 Input imagery	5
3 The ESM_2015 workflow	6
3.1 Cloud screening	6
3.2 Symbolic Machine Learning	6
3.3 Textural and morphological features.....	7
3.4 Level 1 classification: Delineation of built-up surfaces	9
3.5 Level 2 classification: discrimination of residential and non-residential built-up areas.....	10
3.6 Level 3 classification: Green open spaces within urban centres	11
4 Results.....	14
4.1 Level 1 classification results in terms of delineation of built-up areas at 2 meters	14
4.2 Level 2 classification results in terms of discrimination between residential and non-residential buildings at 10 meters	15
4.3 Level 3 classification results in terms of calculation of distances to green open spaces within urban centres at 10 meters	16
5 Validation against LUCAS 2015	18
6 ESM_2015 mosaic and data access	19
7 Datasets description and metadata.....	21
7.1 ESM_BUILT_VHR2015_EUROPE_R2019_3035_02_V1_0 (level 1 layer)	21
7.2 ESM_BUILT_VHR2015CLASS_EUROPE_R2019_3035_10_V1_0 (level 2 layer)	21
7.3 How to cite	22
References	23
List of figures	24
List of tables	25
Annexes	26
Annex 1. Nomenclature of ENACT refined Corine Landcover	26

Abstract

The ESM_2015 is the latest release of the European Settlement Map (ESM) produced in the frame of the Global Human Settlement Layer (GHSL) project. It is produced with the GHSL technology of the Joint Research Centre (JRC) in collaboration with the Directorate General of Regional and Urban Policy. It follows-up on the previous ESM_2012 derived from 2.5 m resolution SPOT-5/6 images acquired in the context of the pan-European GMES/Copernicus (Core_003) dataset for the reference year 2012. The ESM_2015 product exploits the Copernicus VHR_IMAGE_2015 dataset made of satellite images Pleiades, Deimos-02, WorldView-2, WorldView-3, GeoEye-01 and Spot 6/7 ranging from 2014 to 2016. Unlike the previous ESM versions, the built-up extraction is realized through supervised learning (and not only by means of image filtering and processing techniques) based on textural and morphological features. For the first time a new layer containing non-residential buildings was derived by using only remote sensing imagery and training data. The produced built-up map is delivered at 2 m pixel resolution (level 1 layer), while the residential/non-residential layer (level 2) is delivered at 10 m spatial resolution. ESM_2015 offers new opportunities in Earth Observation related research by allowing to study urbanisation and related features across Europe in urban and rural areas, from continental to country perspective, from regional to local, until single blocks.

ESM_2015 was validated against the LUCAS 2015 survey database both at 2 and 10 meters resolution (including also the non-residential class). The validation has resulted in a Balanced Accuracy of 0.81 for the 2 m resolution built-up layer and of 0.71 for the 10 m non-residential built-up layer.

Authors

Filip Sabo^a, Christina Corbane^b, Panagiotis Politis^c, Thomas Kemper^b

^aArhs Developments Italia S.r.l., Milano, Italy; ^bEuropean Commission, Joint Research Centre, Disaster Risk Management Unit; ^cArhs Developments S.A., Luxembourg

1 Introduction

The Global Human Settlement Layer (GHSL) framework [1] addresses questions related to the location of human settlements, their characteristics and their vulnerability. In contrast to the GHSL global products, which use open and free data from Landsat [2] and Sentinel-1 [3], the European Settlement Map (ESM) is exploiting the Very High-Resolution (VHR) commercial image collections still relying on the GHSL methodology for automatic image information extraction [4].

According to the Data Warehouse document¹, the VHR image collection acquisitions shall be repeated every 3 years starting from 2012 as reference year. All the previous releases of the ESM (ESM_2012) were based on the optical VHR DWH_MG2b_CORE_03 image collection, also known as Core_003. This collection consisted of ~3500 Spot 5/6 false colour (red, green and near infrared bands) pansharpned data at 2.5 m for the 2011-2013 year interval (the first reference year). The VHR collection coverage is for 39 European countries, members or cooperating countries of the European Environment Agency (EEA-39).

The first ESM_2012 was released in 2016 [5] as a layer which contains the percentage of built-up area coverage per 10 m spatial grid. The methodology was based on the use of textural (Pantex) [6] and morphological features [7]. The binary built-up class at 2.5 m was aggregated to 10 m using the average operator and therefore it was made continuous. The latest release of the ESM_2012 in 2017 [8] was also based on the textural and morphological technique of unsupervised built-up area detection. However, additional layers from Open Street Map, Urban Atlas and from other sources were used in the post-processing step in order to improve the classification accuracy. The layer is a discrete classified dataset released at 2.5 m spatial resolution and it consists of several classes other than built-up. The ESM_2012 release 2017 is also the final release of the ESM_2012 based on the Core_003 collection.

Recently, as part of the Copernicus contributing missions, a Pan-European coverage (EEA39 countries) of very high resolution (VHR) images for the reference year 2015 (VHR_IMAGE_2015) was acquired for the purpose of continuation of local land component applications for hotspots, coastal areas, risk zones and protected sites. This latest 2019 release of the ESM (ESM_2015) is based on the VHR_IMAGE_2015 image collection and it is described in this report.

¹ https://spacedata.copernicus.eu/documents/12833/14545/DAP_Release_Phase_2

2 Input imagery

The VHR_IMAGE_2015 data has been acquired during the 2014, 2015 and 2016 vegetation seasons. The data providers were: Airbus Defence and Space, European Space Imaging (EUSI) and Deimos Imaging. The satellite missions are: Pleiades (Airbus), WorldView-2 (WV-2), WorldView-3 (WV-3), GeoEye-1 (GE-1) (EUSI) and Deimos-2 (Deimos Imaging). SPOT 6/7 images were additionally delivered to improve the coverage over the remaining gaps. All the delivered data is composed of 4 multispectral bands (red, green, blue and near infrared) and a panchromatic image in the European ETRS89-Lambert azimuthal equal-area (LAEA) projection (EPSG:3035). Compared to the previous image collection (Core_003), the main additions are: availability of 4 multispectral (MS) bands and a panchromatic image, higher spatial resolution and improved radiometry. The previous image collection, Core_03, was delivered as a 3 band (red, green and near infrared) pan sharpened image at 2.5 m. The VHR_IMAGE_2015 collection was downloaded and stored on the Joint Research Centre Earth Observation Data and Processing Platform (JEODPP) [9]. The total size of the downloaded data is 127TB. All the data is delivered as a 4 band composite (blue, green, red and infrared) and a panchromatic image.

The VHR_IMAGE_2015 delivered collection per contributing satellite mission is shown in figure 1 (a) and the image collection per acquisition year is shown in figure 1 (b).

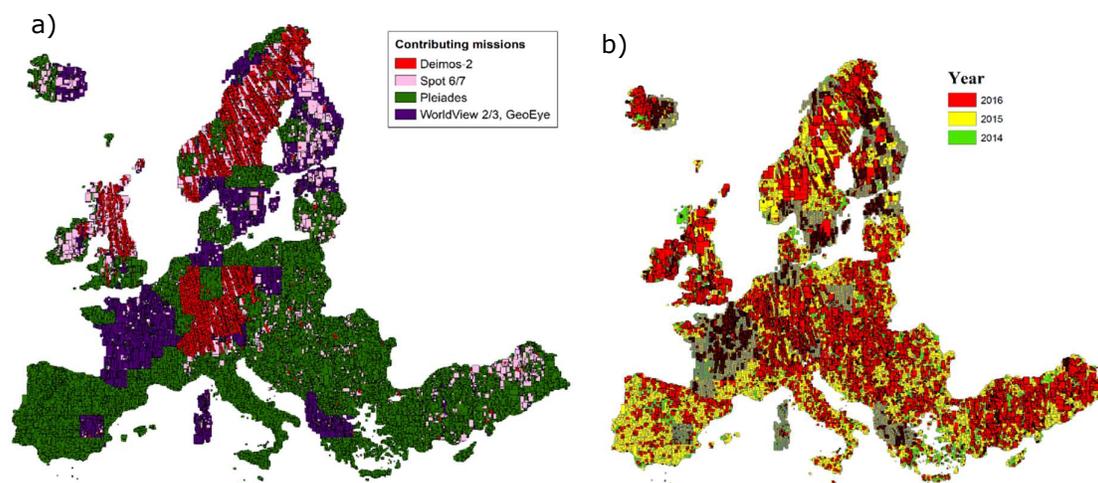


Figure 1 (a) Delivered imagery per contributing mission and (b) the imagery year distribution.

The image size ranges from 4,096x4,096 pixels to more than 15,000x15,000 pixels. The panchromatic spatial resolution is in the range 0.4-1 m, while the multispectral (MS) spatial resolution ranges from 1.4 to 6 m. Deimos sensor captures the data as 10-bit values, while the Airbus and EUSI missions collect the data in a 12-bit range.

3 The ESM_2015 workflow

The ESM_2015 workflow is a fully automatic image information extraction technique split into three levels of processing. Unlike the workflows for the ESM_2012 production, the main building block of the ESM_2015 workflow is the use of the supervised scene based classification algorithm - the Symbolic Machine Learning (SML) [10], [11] in combination with the textural (Pantex) and morphological features.

The ESM_2015 workflow is divided into three levels of processing corresponding to three different levels of outputs. The learning sets used in the workflow are: the latest ESM_2012 release, the ENhancing ACTivity and population mapping (ENACT) map [12] and the Water and Wetness (WAW) product from the EEA. The learning sets are resampled to the MS pixel size with the nearest operator and the training is done at the input data resolution at the scene level.

3.1 Cloud screening

Prior to the workflow execution, a cloud screening technique was applied to the input data since the VHR collection was not delivered as cloud-free. Automatic cloud screening was based on the brightness values of the input bands with additional morphological filtering followed with dilation and erosion operations. A cloud screening interface (based on html) was developed in order to visually check all the scenes which have more than 30% of clouds as identified. Visual check resulted in the exclusion of 6221 scenes. More precisely, 8% of the total Deimos scenes and 3% of the total EUSI tiles were excluded. The html interface is shown in figure 2.



Figure 2. Html interface developed for the visual cloud check

In addition to this, data gaps over built-up areas were noticed in large parts of England and Ireland, smaller parts of Germany and in Austria, close to the Alps.

3.2 Symbolic Machine Learning

The two main steps of the SML method are the data reduction and association analysis (AAx). In the data reduction step, the distribution of every SML input vector, i.e. of each feature extracted from the multispectral image, is partitioned in few groups using a uniform quantization approach. The quantization parameter is sensor and landscape specific. Moreover, different quantization levels are necessary to be applied per scene since the sensors collect the data in a bit depth prescribed by their design (10 or 12 bits as regards to the data set used in this study). The input image features were quantized as follows:

$$X_{qi} = [x/q : x \in x_i] \quad (1)$$

where x are the input data instances, $i = 1,2,3,4$ and q is the quantization parameter determined so that the average support is in the range 10^3 - 10^4 . The average support controls the SML generalization ability and computational performances [11]. Depending on the average support, for EUSI and Airbus scenes $q \in$

[16,32,64] while for the Deimos scenes $q \in [128,256,512]$. After data reduction, the unique data sequences are identified together with their frequencies. The sequences are then used in the AAx step where they are scored according to the frequencies of their joint occurrences with positive and negative data instances from the learning set. The output of the AAx is the Evidence-based Normalized Differential Index (ENDI) which is a continuous layer ranging from -1 to +1. Values close to 1 indicate that the data sequence is in a strong association with the class of interest (built-up area). The ENDI is shown in figure 3.

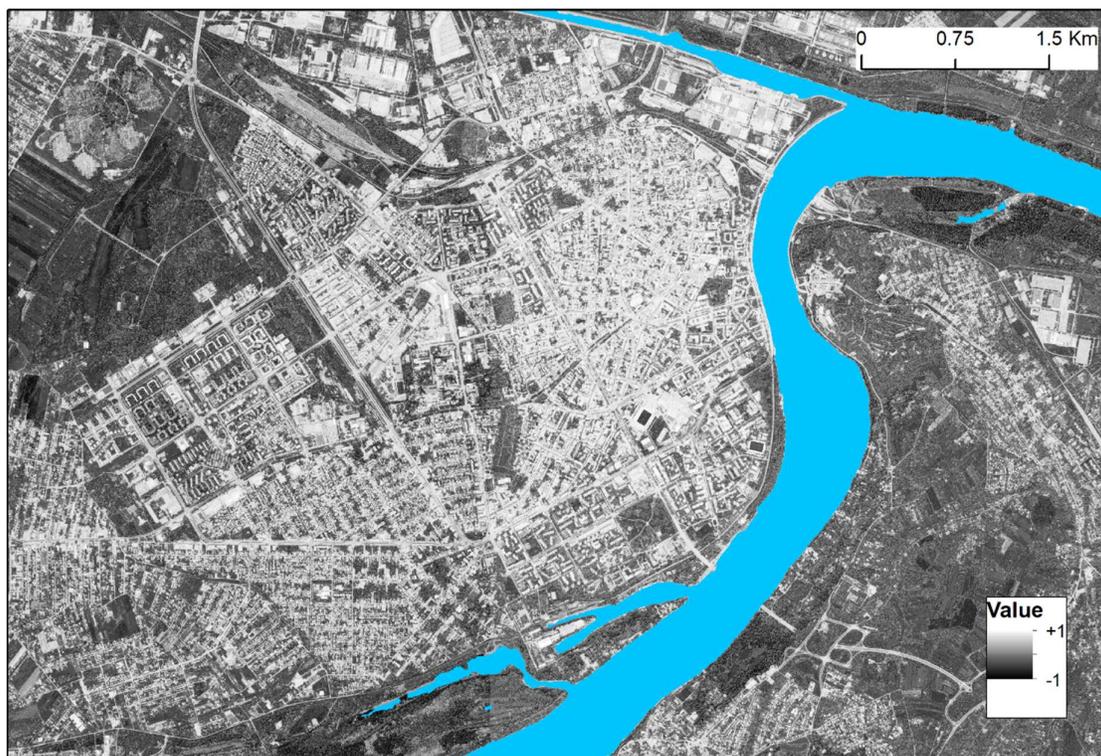


Figure 3. The output from the Association analysis - Evidence-based Normalized Differential Index in Novi Sad (Serbia). Water is shown with blue colour.

3.3 Textural and morphological features

In addition to the 4 spectral bands (radiometric features), the textural (Pantex) and morphological (Characteristic-Saliency-Level, CSL) spatial features were extracted from the panchromatic image in order to improve as much as possible the BU detection in complex urban environments.

The Pantex is an edge density detector and can be considered as an index indicating the presence of BU. The higher the edge density is, the higher the probability of a pixel is to contain information about the presence of a building in its neighbourhood [6]. The window size for the calculation of the Pantex corresponds to the local area of the textural measure (contrast). The window size value is estimated empirically. It was demonstrated that increasing gradually the window size leads to a better detection of large buildings and industrial facilities. Extracted textural features are shown in figure 5.

CSL is a compact representation of the multi-scale information that is conveyed by the differential morphological profiles (DMP) computed over the raster layers. The CSL reduces the dimensionality of the DMPs by mapping three layers, the scale, the saliency and the level features that highlight the dominant image components and lead actually to the segmentation of the image [13]. An example of the CSL model output is displayed in figure 6.



Figure 4. Example of a panchromatic image used for the calculation of the Pantex and the CSL in the area of New Belgrade (Serbia)

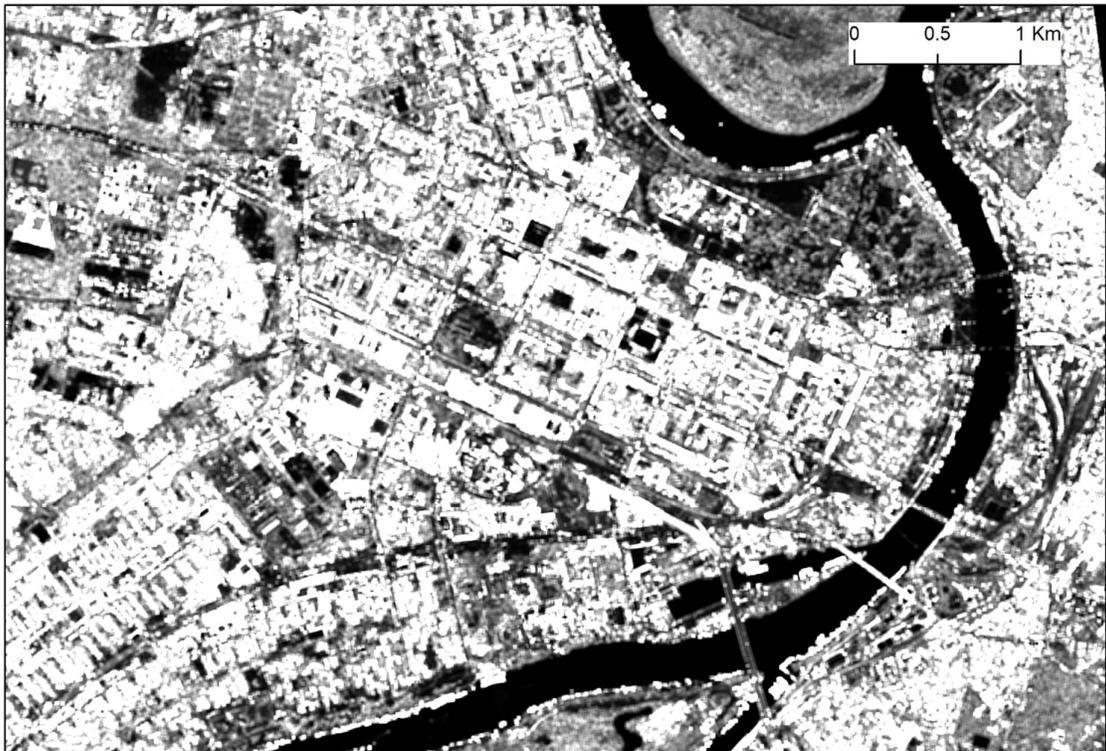


Figure 5. Textural features (pantex) derived from the panchromatic image (figure 4)

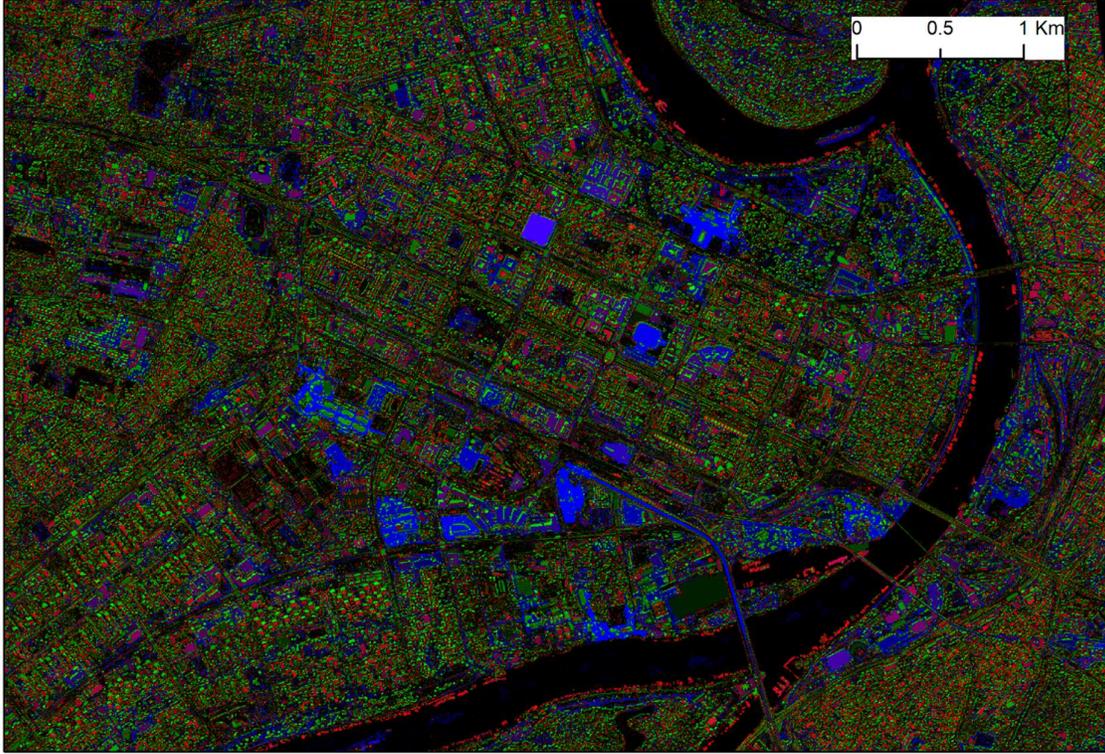


Figure 6. CSL2RGB model representation derived from the input panchromatic image (figure 4)

3.4 Level 1 classification: Delineation of built-up surfaces

The delineation of built-up, water and land surfaces is achieved with the level 1 classification. ESM_2012 buildings are used as a learning set for built-up area (BU) detection. AAx for the level 1 works at scene-level; it considers both the ESM_2012 training set for a specific scene and the unique sequences of radiometric and morphological characteristics extracted from that input scene. Both training and classification were performed per scene at the spatial resolution of the MS bands. ENDI from level 1 processing was weighted by the Pantex layer in order to reduce the over detections in non-built-up areas. Combined with the SML, it improves the BU detection by enhancing the BU confidence values and at the same time removing over detections such as roads, bare land, sand dunes and other open spaces. However, a drawback with the Pantex is that it can remove relevant information when the edge density is relatively low. This can be the case with large industrial facilities or large commercial and public buildings. Weighting the ENDI with the Pantex in this case can lead to a complete or partial removal of the buildings in the binary layer. In order to overcome this issue, the Pantex features were calculated for two resampled panchromatic images as follows:

- Panchromatic image at 1 m, window size 100 x 100 m (ptx1)
- Panchromatic image at 2.5 m, window size 25 x 25 m (ptx2)

The ptx1 setup was providing good results in dense urban areas, however in the low urban density areas some information was removed, scattered buildings were not detected as BU. Therefore the ptx2 was used to complement these omissions. Increasing the window size to 100 m contributed to better detection of large residential and industrial BU, but also to the over detections of roads and railroads. This commission error was allowed in order to capture as many BU pixels as possible. Furthermore, in order to increase the probability for the large non-residential facilities to be captured, ENDI was not weighted by the Pantex layer in the areas where non-residential buildings can be found as defined by the ENACT training set. These rules are explained with the following equations using a logical OR operator:

$$BU_{ptx} = BU_{ptx1} \vee BU_{ptx2} \quad (2)$$

$$BU_{no_ptx} = BU_{ptx} \vee BU_{org} \quad (3)$$

where BU is the binarized ENDI using the Otsu method [14], BU_{ptx1} is the BU with ptx1 setup, BU_{ptx2} is the BU with ptx2 setup, BU_{org} is the BU not weighted by the Pantex. BU_{no_ptx} is then derived by applying Eq. (2) in areas where large buildings can be found according to the ENACT learning set.

The WAW 2015 product with 20m pixel size is used as a learning set for water classification. In addition to the WAW, the ENACT map is also used in combination with the water from ESM_2012 for coastal water detection (sea and ocean) since the WAW product is dedicated only to inland water. The ESM_2012 water could not be used for water classification, since there are false detections of water in urban areas, which can seriously disrupt the SML. The classes 1, 2 and 3 from the WAW were used in the workflow, code 4 (temporary wet areas) was excluded. Similarly to the ENDI for BU, the AAX for the water is also measured and the ENDI water is extracted.

3.5 Level 2 classification: discrimination of residential and non-residential built-up areas

At the level 2 processing the discrimination of residential and non-residential BU areas is achieved by combining the ENACT map with the ESM_2012. The ENACT map is a thematic and spatial refinement of the Corine Land Cover. The ENACT methodology is combining external layers such as Urban Atlas, TomTom, Corine change layers and other higher resolution layers at European scale (Annex 1). The learning set for the ESM_2015 BU is defined as an intersection of ESM_2012 BU with commercial, production facilities and port areas from the ENACT map. Therefore, the learning set used per scene are buildings from the ESM_2012 that correspond to non-residential areas from the ENACT map. The logical operation applied to discriminate the residential from non-residential BU is the following:

$$LRN_{non_res} = ESM_{2012_{50}} \wedge ENACT_{non_res} \quad (4)$$

where LRN_{non_res} is the learning set used for defining non-residential BU, $ESM_{2012_{50}}$ are the buildings from the ESM_2012 (class 50) and $ENACT_{non_res}$ are the non-residential pixels corresponding to the ENACT map.

For the level 2 classification, the AAX was applied by considering only the non-residential training set and unique sequences of radiometric and morphological features from the input imagery. Non-residential buildings are larger than residential buildings and generally with different shapes and radiometry represented by white, grey and black colour. Morphological features can contribute in discriminating residential from non-residential buildings based on their size and shape by measuring the AAX between the non-residential training set and the actual input data. On the other hand, the radiometric features can also improve the SML outputs by giving high confidence to the buildings with specific reflection in the visible and NIR domain. The output of the AAX is the confidence for level 2 layer (ENDI BU non-residential). Generally, the confidence is high for large industrial facilities, port areas, large public buildings, stadiums and other commercial facilities. Generally, the confidence will be high for large industrial facilities, port areas, large public buildings, stadiums and other commercial facilities. It must be noted that at this stage the Pantex was not used to weight the ENDI BU non-residential because it can lower the confidence for very large buildings. Furthermore, since the learning set is constrained to the intersection between two layers, the false positives were already significantly reduced. This resulted in less commission errors in bare land, agricultural areas, roads, and other non-BU areas compared to the level 1 classification. The main issue for the level 2 classification was the over detection of non-residential BU when using only the ENDI BU non-residential. To bypass this issue the ENDI from the level 1 was combined with the ENDI from level 2. In this way, the threshold value is automatically increased for each scene but still keeping the ENDI for non-residential BU areas high and allowing the detection of new non-residential BU areas outside the learning zones. Finally, the level 2 ENDI is aggregated to 10 m output resolution and binarized following the Otsu's approach.

The core workflow with the input data and input learning sets for level 1 and 2 is illustrated in figure 7. Level 3 processing follows after the core workflow and the results are outputted per city. Level 3 workflow is presented in figure 9.

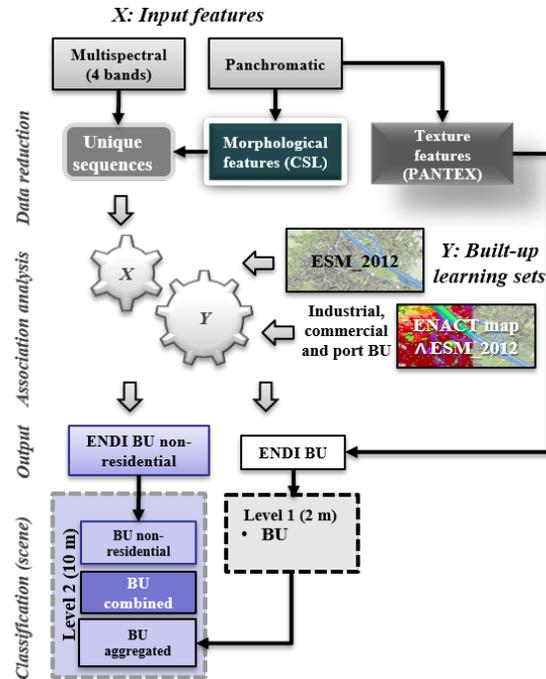


Figure 7. The ESMA_2015 simplified workflow for level 1 and level 2 processing levels

3.6 Level 3 classification: Green open spaces within urban centres

Level 3 classification aims at characterizing the urban centres in terms access to green open spaces. It builds on the combination of: 1) the urban/rural settlement model (GHS-SMOD) for delineating the urban centres, 2) the built-up areas derived from the ESMA_2015, 3) the open spaces derived from the Urban Atlas (2012) and 4) the greenness derived from the Normalized Difference Vegetation Index (NDVI). The integration of these layers allows deriving meaningful indicators describing the European cities in a consistent way.

The location and delineation of urban centres builds on the urban/rural settlement model (GHS-SMOD) which applies the Degree of Urbanization classification [15] to the GHS baseline data (i.e. to the built-up and to the population grids of the GHS). The GHS-SMOD allows deriving a globally consistent delineation of city boundaries at a spatial resolution of 1 km. The Degree of Urbanisation definition relies on population thresholds and spatial contiguity rules. In the urban/rural settlement model (GHS-SMOD) formulation, the 'urban centres' class is implemented as a spatial generalization of contiguous population grid cells of 1 km² with a density of at least 1,500 inhabitants per km² or a density of built-up surface >50%, and a minimum total resident population of 50,000. More details on the classification schema can be found in [15].

In the context of the GHS, the application of the Degree of Urbanisation classification to the 1 km grids produced a total of 10,323 urban centres in 2015 at global scale and 888 at the European Scale (figure 8).

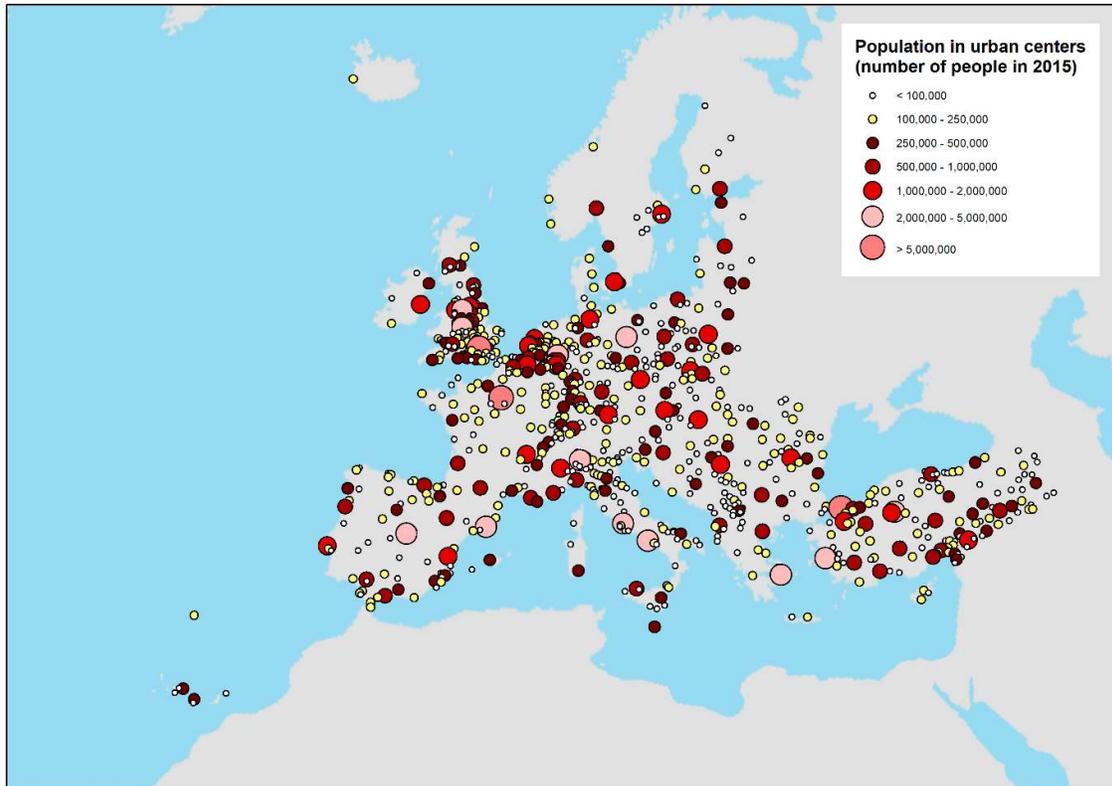


Figure 8. Location and size (in terms of population) of urban centres in Europe

These spatially delineated urban centres represent the spatial reference upon which the accessibility to open spaces from built-up areas is calculated.

The open spaces are derived from the Urban Atlas 2012 and correspond to the following classes:

Green urban areas (code 14100), Sports and leisure facilities (code 14200), Forests (code 31000) and Open spaces with little or no vegetation (code 33000). Open spaces were rasterized at 10 m spatial resolution and clipped with the urban centres defined by the GHS-SMOD [16].

Accessibility to open spaces, is typically measured using rules for spatial proximity between objects, such as the distance between peoples permanent place of residence and open spaces. The methodology developed in the ESM_2015, consists in calculating the geodesic distance from built-up areas, with resident population, to any open space (without discriminating between classes of open spaces). For that purpose we use the gray-weighted distance function based on the geodesic time algorithm developed by Soille in 1994 [17]. The level 3 classification workflow is shown in figure 9.

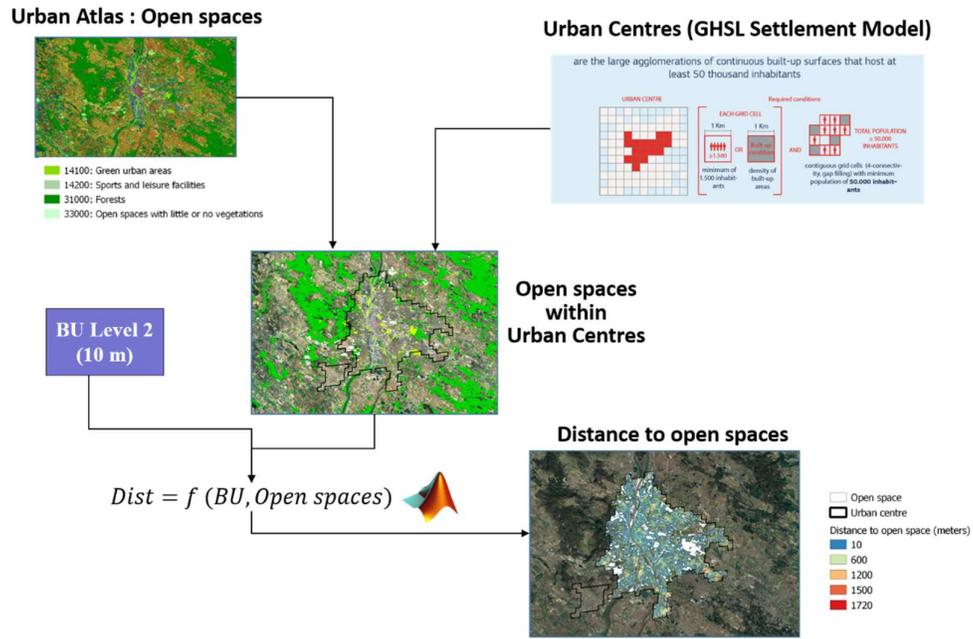


Figure 9. Level 3 processing workflow for extracting the distances from built-up areas to open spaces as defined in the Urban Atlas.

The open spaces were defined as the seed locations and the level 2 BU layer is the binary image from which the gray-weighted distance transform is calculated.

Level 3 classification results are not included in the ESM_2015 release package. They will be available soon upon request.

4 Results

The BU density map for the ESM_2015 level 1 dataset is shown in figure 10. The original dataset at 2 m pixel size was aggregated to 5 km pixel size for pan-European visualization purposes.

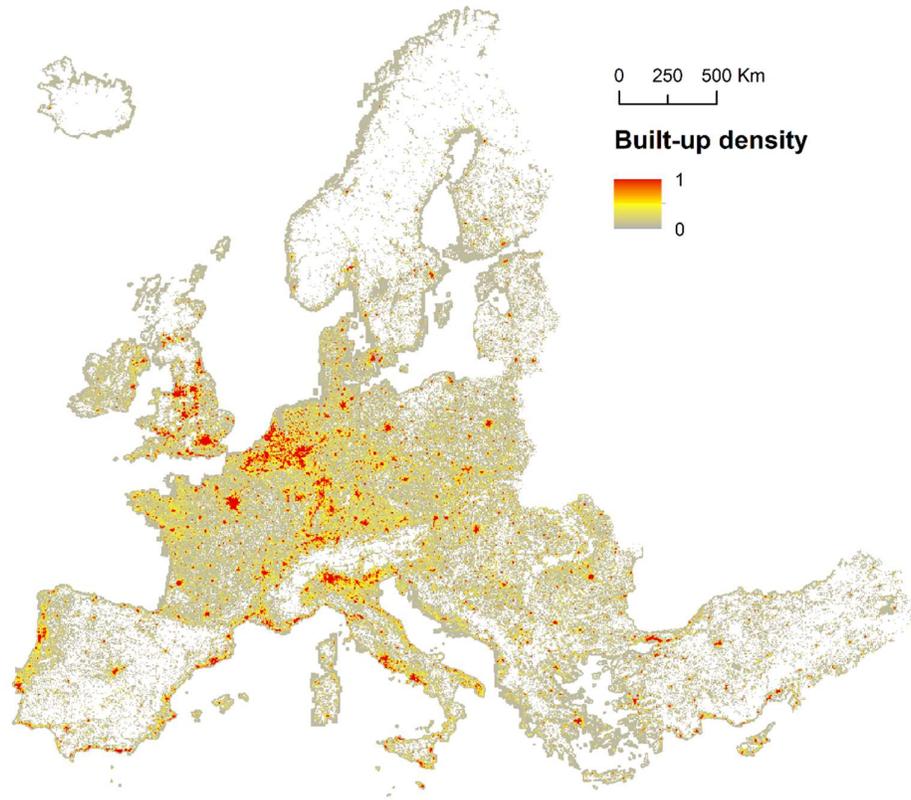


Figure 10. The full ESM_2015 level 1 resampled product at 5 km pixel size. Pixel values correspond to the BU densities since an aggregation operator was applied to the 2 m product

4.1 Level 1 classification results in terms of delineation of built-up areas at 2 meters

The level 1 layer is provided as a discrete classified dataset at 2 m spatial resolution. An example of the BU detection for level 1 layer is shown in figure 11 below.

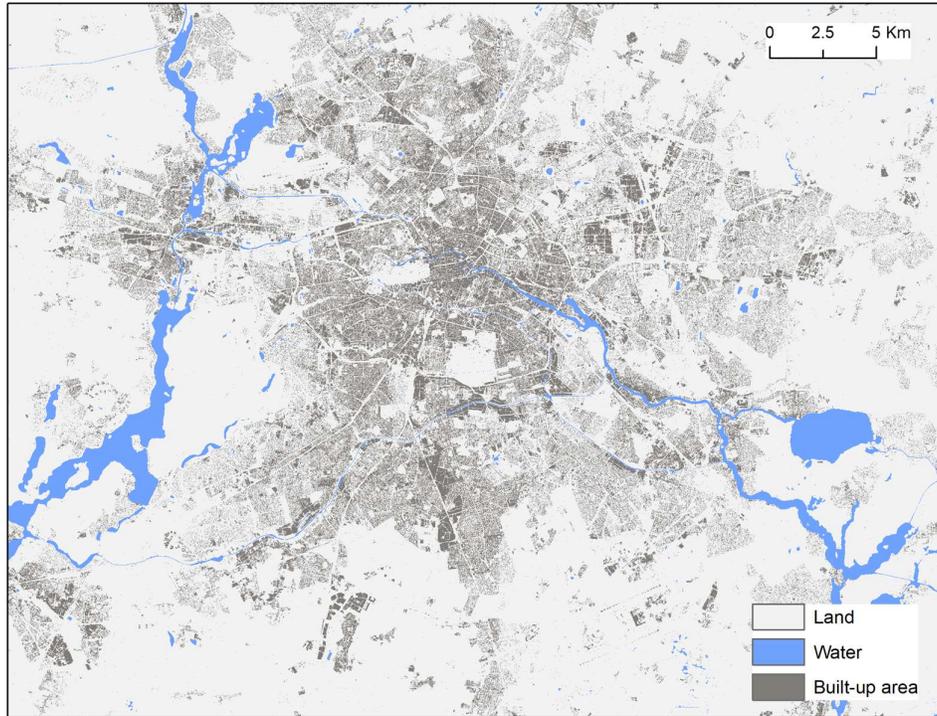


Figure 11. Built-up area detection for level 1 layer (2 m) in Berlin (Germany)

Increasing the window size of the Pantex contributed to better detection of large residential buildings while the exclusion of the Pantex in the non-residential areas improved the detection of industrial and commercial facilities with some commission errors on parking lots or roads between them. In addition, the benefit is the additional detection of new BU areas, which were not provided by the learning set.

4.2 Level 2 classification results in terms of discrimination between residential and non-residential buildings at 10 meters

Level 2 layer is also a discrete classified dataset providing residential and non-residential BU areas at 10m pixel size. The results of the BU detection for the non-residential class are shown in figure 12 in purple colour. Even with the exclusion of the public facilities from the learning set (ENACT), the majority of them were detected as BU because many connected components with specific reflection and size were already provided to the SML. In addition, new buildings outside training zones were detected, providing in that way not only the spatial refinement but also new non-residential BU detections. In some cases, entire new industrial zones were detected. However, there are examples where large residential buildings were also captured as non-residential due to their size and radiometry.

Commission errors of roads and railways for both levels 1 and 2 were decreased by using the following classes from Urban Atlas: 12210 (fast transit roads), 12220 (other roads) and 12230 (railways) in the post-processing step.

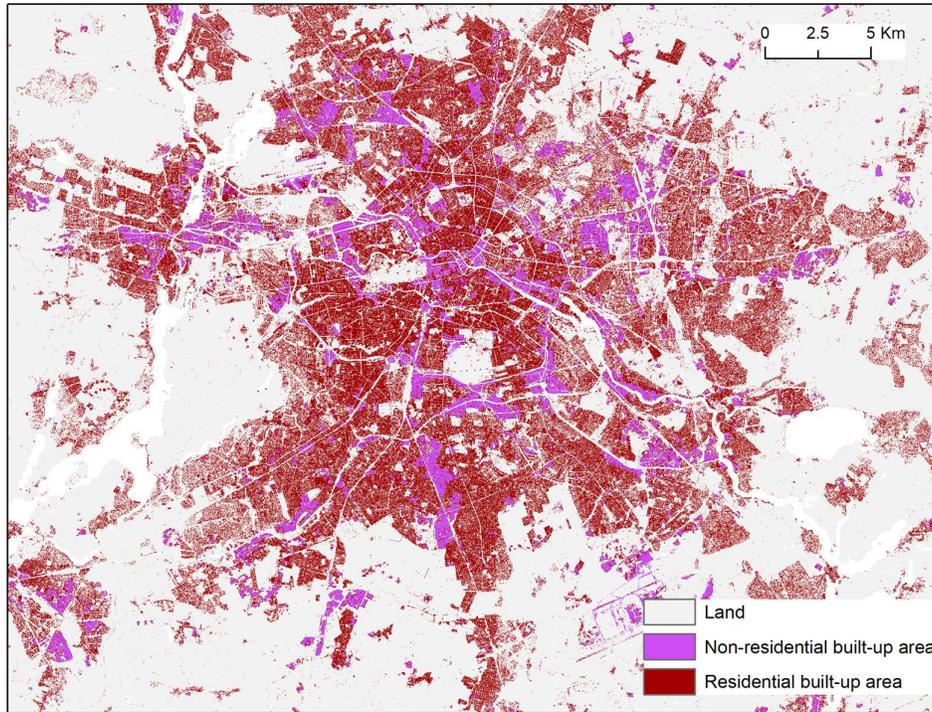


Figure 12. Discrimination between residential and non-residential built-up area for level 2 layer (10 m) in Berlin (Germany)

4.3 Level 3 classification results in terms of calculation of distances to green open spaces within urban centres at 10 meters

The level 3 layer delivers the following city-based layers: open spaces, a distance layer that contains the Euclidean distances from built-up pixels to open spaces and green in open spaces. The pixel size for level 3 layer is 10 m and the extent of each city is defined by the urban centres database. The level 3 layer is calculated for 888 cities. Open spaces are binary layers, while the distances to open spaces are continuous layers (float data type). The distance is expressed in number of pixels. To derive the distance in meters, it has to be multiplied by 10 corresponding to the pixel size in meters. The distance layers relative to open spaces are shown in figures 13 and 14 for two urban centres, Montpellier and Paris.

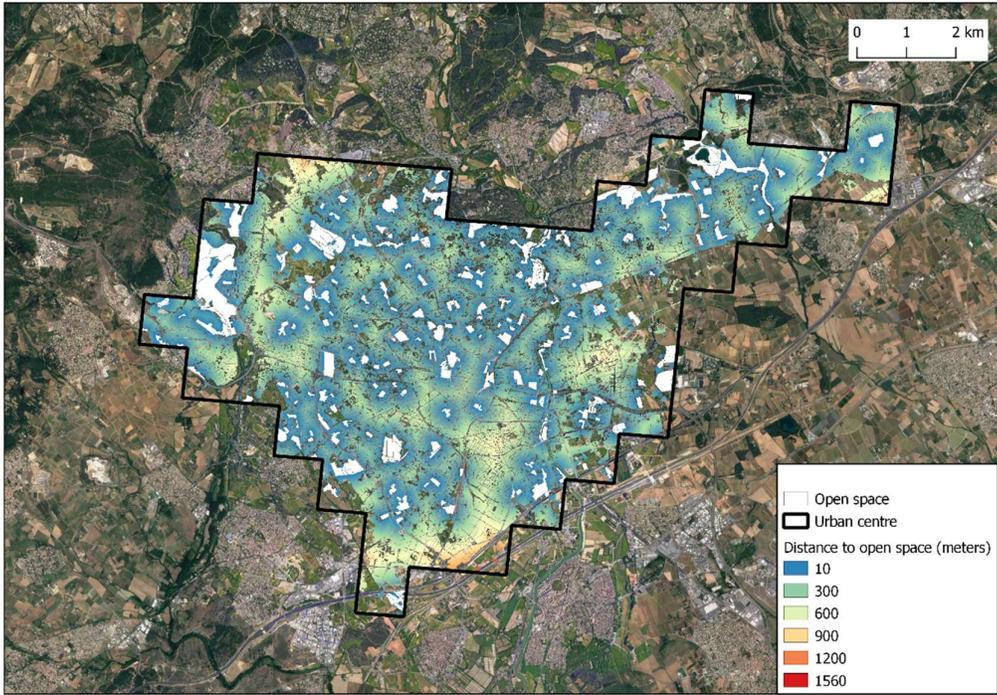


Figure 13. An example of open spaces and Distance to open space layers in Montpellier (France)

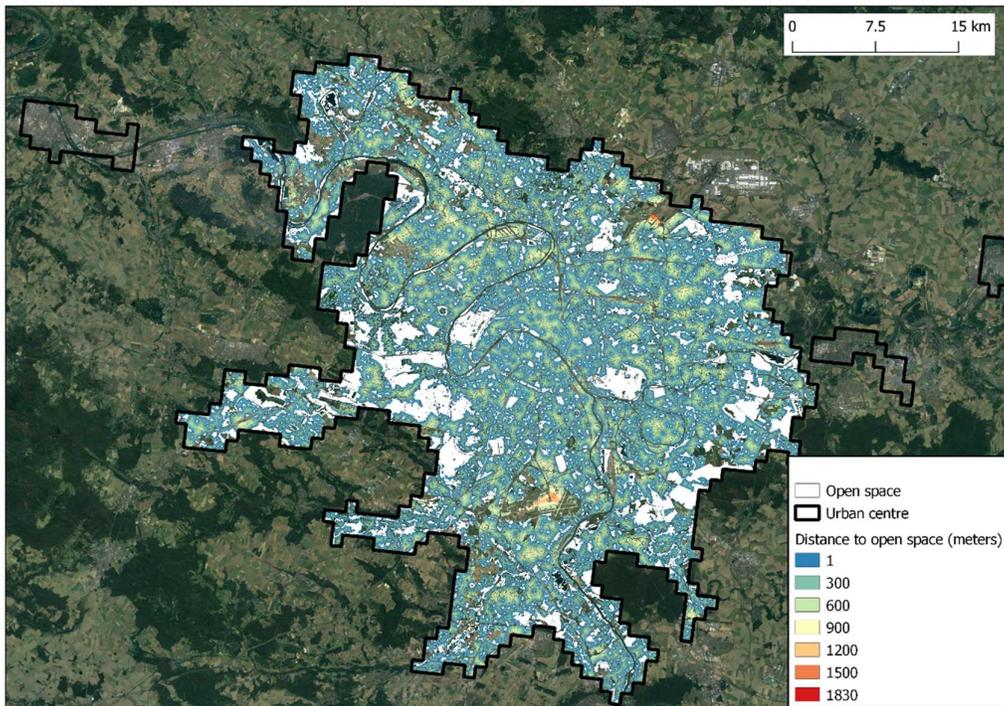


Figure 14. Distance layer relative to open spaces in Paris. Warmer colours indicate longer distances from BU to open space

5 Validation against LUCAS 2015

Points from the Land Use/Cover Area frame Statistical survey (LUCAS 2015) database were used in order to validate both the level 1 and level 2 layer at European scale. A total of 340,144 LUCAS points are distributed in 28 European Union countries. From the total number of points, 4275 belong to the built-up area classes (A11, A 12 or A13). LUCAS points were also used for the non-residential layer validation from level 2 since they contain a detailed land use. If the point belongs to the BU class (A11, A12 or A13) and the Land Use column (LU1) corresponds to non-residential buildings (industrial, commercial, etc.) then it was considered for the validation of non-residential BU versus all other classes (including residential BU). A buffer (circle with a radius of 2 m) was applied around LUCAS point in order to follow the theoretical definition of the points and to adapt it to the pixel size of the BU layers.

Statistics (True positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN)) from a point-based confusion matrix were derived and are shown in Table 1.

Table 1. Point based statistics from the confusion matrix

Layer	TP	FP	TN	FN
Level 1 (2 m)	2665	4117	323002	1610
Level 2 (10 m)	3736	11894	309961	582
Level 2 non-residential (10 m)	424	995	326142	589

The balanced accuracy (BA) was used as the main performance metric, because there is a high share of non-built-up areas compared to built-up. According to the performance metrics results given in the Table 2, the balanced accuracy of Level 2 layer is the highest (0.91). However, the commission error (CE) increased to 0.76 due to the aggregation from 2 to 10 m, that is the FP rate increased. Consequently, the omission error (OE) dropped to 0.13. Many of the FPs for all layers actually belong to the non-built-up area or linear features in the LUCAS dataset. These classes correspond to roads, railroads and parking lots in residential areas.

Table 2. Performance metrics calculated from the confusion matrix elements

Performance metrics	Level 1	Level 2	Level 2 non-residential
Balanced Accuracy	0.81	0.91	0.71
Omission error	0.38	0.13	0.58
Commission error	0.6	0.76	0.70

Level 2 non-residential BU validation provided the BA of 0.71 and CE and OE of 0.70 and 0.58, respectively. Similarly to the level 1 layer, the FPs of the non-residential layer also belong to the class roads as identified by the LUCAS set. In addition, the CE can be attributed to the confusion between non-residential and residential buildings in high BU density areas. It was noticed that issues still exist with the LUCAS dataset. For example, some of LUCAS BU points were located in agricultural fields, open spaces and green areas. This can deteriorate the performance metrics for all layers.

6 ESM_2015 mosaic and data access

The ESM_2015 is tiled and merged following the European reference grid (figure 15) tiling schema and delivered as a mosaic of 897 images. Tile dimensions are 50000x50000 pixels (100x100 km) for level 1 layer and 10000x10000 pixels for level 2 layer. Each tile is following a specific name convention which reflects its position in the ETRS89-LAEA coordinate reference system. The origin of the grid coincides with the false origin of the ETRS89-LAEA coordinate reference system ($x=0$, $y=0$). The grid orientation is south-north, west-east.

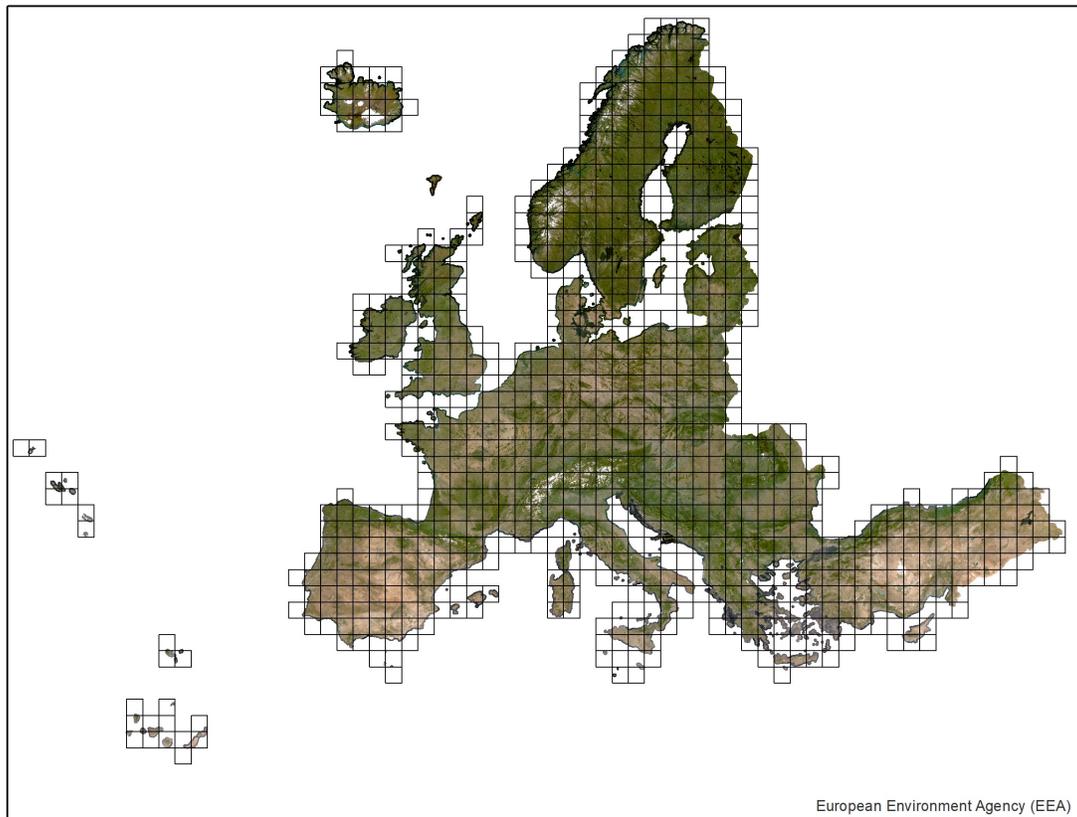


Figure 15. ESM tiling grid displayed on top of the EEA Sentinel 2 RGB mosaic.

The single tiles for the 2 meters and 10 meters datasets can be downloaded from the GHSL website:

<https://ghsl.jrc.ec.europa.eu/download.php?ds=ESM>

All ESM_2015 datasets are available on the JRC open data catalogue:

<https://data.jrc.ec.europa.eu/dataset/8bd2b792-cc33-4c11-afd1-b8dd60b44f3b>

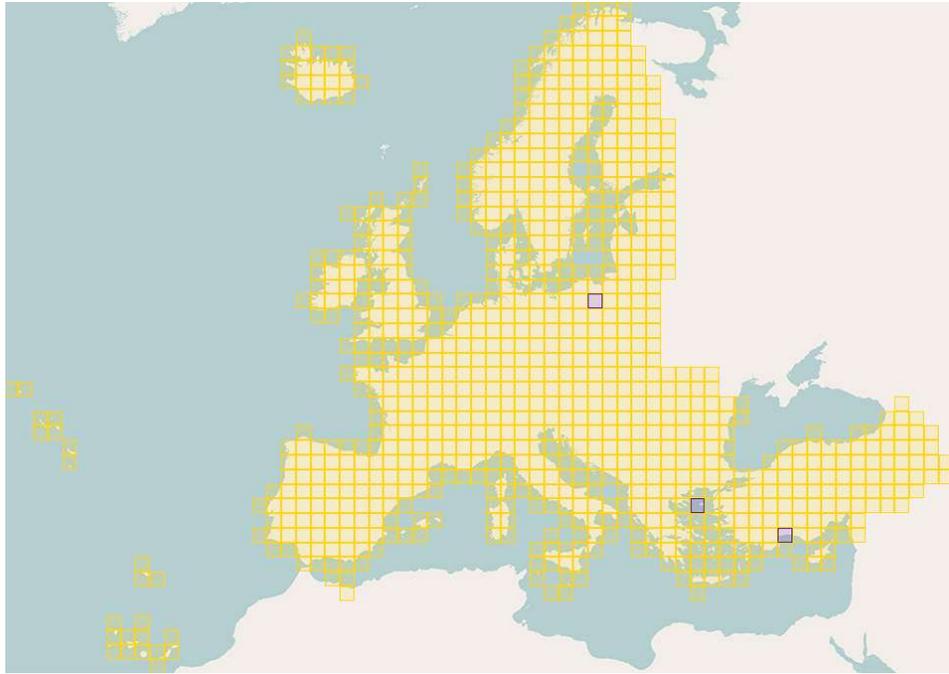


Figure 16. The ESM_2015 download interface from the GHSL website

<https://ghsl.jrc.ec.europa.eu/download.php?ds=ESM>

7 Datasets description and metadata

Product name: ESM_BUILT_VHR2015_EUROPE_R2019

The European Settlement Map (ESM_2015) is published in two layers:

- ESM_BUILT_VHR2015_EUROPE_R2019_3035_02_V1_0: classifies the built-up areas at a spatial resolution of 2 meters (EPSG:3035)
- ESM_BUILT_VHR2015CLASS_EUROPE_R2019_3035_10_V1_0: classifies the built-up areas into residential and non-residential at a spatial resolution of 10 meters (EPSG:3035)

7.1 ESM_BUILT_VHR2015_EUROPE_R2019_3035_02_V1_0 (level 1 layer)

Author: Christina Corbane, Filip Sabo; Joint Research Centre (JRC) European Commission

Spatial extent: Europe (EEA-39)

Temporal extent: 2014, 2015, 2016

Coordinate System: ETRS89-LAEA (EPSG: 3035)

Resolution: 2 m

Encoding: Multiclass layer [0,1,2,255]

Description:	built-up areas classification at a spatial resolution of 2 meters
Dataset name (size):	ESM_BUILT_VHR2015_EUROPE_R2019_3035_02_V1_0(8.14GB)
Legend:	0 = no data 1 = land 2 = water 255 = built-up area

Data organisation: VRT file (with TIFF files); pyramids, SHP file of the tile schema, color table file (clr), metadata file (txt).

The dataset is also distributed in .zip files in the form of tiles of 100 x100 km size.

7.2 ESM_BUILT_VHR2015CLASS_EUROPE_R2019_3035_10_V1_0 (level 2 layer)

Author: Christina Corbane, Filip Sabo; Joint Research Centre (JRC) European Commission

Spatial extent: Europe (EEA-39)

Temporal extent: 2014, 2015, 2016

Coordinate System: ETRS89-LAEA (EPSG: 3035)

Resolution: 10 m

Encoding: Multiclass layer [0,1,250,255]

Description:	built-up areas classification into residential and non-residential at a spatial resolution of 10 meters
Dataset name (size):	ESM_BUILT_VHR2015CLASS_EUROPE_R2019_3035_10_V1_0 (998MB)
Legend:	0 = no data 1 = land 250 = non-residential built-up area 255 = residential built-up area

Data organisation: VRT file (with TIFF files); pyramids, SHP file of the tile schema, color table file (clr), metadata file (txt).

The dataset is also distributed in .zip files in the form of tiles of 100 x100 km size.

7.3 How to cite

Dataset:

Corbane, Christina; Sabo, Filip (2019): European Settlement Map from Copernicus Very High Resolution data for reference year 2015, Public Release 2019. European Commission, Joint Research Centre (JRC) [Dataset] doi:[10.2905/8BD2B792-CC33-4C11-AFD1-B8DD60B44F3B](https://doi.org/10.2905/8BD2B792-CC33-4C11-AFD1-B8DD60B44F3B) PID: <http://data.europa.eu/89h/8bd2b792-cc33-4c11-afd1-b8dd60b44f3b>

Concept & Methodology:

F. Sabo, C. Corbane, P. Politis, M. Pesaresi and T. Kemper, "Update and improvement of the European Settlement map", 2019 Joint Urban Remote Sensing Event (JURSE), Vannes, France, 2019, pp. 1-4. doi: [10.1109/JURSE.2019.8808933](https://doi.org/10.1109/JURSE.2019.8808933)

Corbane, C., Sabo, F. Syrris, V., Kemper, T., Politis, P., Pesaresi, M., Soille, P. and Osé, K. , "Application of the Symbolic Machine Learning to Copernicus VHR Imagery: The European Settlement Map", IEEE Geoscience and Remote Sensing Letters, 10.1109/LGRS.2019.2942131

References

- [1] M. Pesaresi *et al.*, "A Global Human Settlement Layer From Optical HR/VHR RS Data: Concept and First Results," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 5, pp. 2102–2131, Oct. 2013.
- [2] C. Corbane *et al.*, "Big earth data analytics on Sentinel-1 and Landsat imagery in support to global human settlements mapping," *Big Earth Data*, vol. 1, no. 1–2, pp. 118–144, 2017.
- [3] C. Corbane *et al.*, "Enhanced automatic detection of human settlements using Sentinel-1 interferometric coherence," *International Journal of Remote Sensing*, vol. 39, no. 3, pp. 842–853, Feb. 2018.
- [4] F. Sabo, C. Corbane, A. J. Florczyk, S. Ferri, M. Pesaresi, and T. Kemper, "Comparison of built-up area maps produced within the global human settlement framework," *Transactions in GIS*, Nov. 2018.
- [5] A. J. Florczyk *et al.*, "A New European Settlement Map From Optical Remotely Sensed Data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 5, pp. 1978–1992, May 2016.
- [6] M. Pesaresi, A. Gerhardinger, and F. Kayitakire, "A Robust Built-Up Area Presence Index by Anisotropic Rotation-Invariant Textural Measure," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 1, no. 3, pp. 180–192, Sep. 2008.
- [7] M. Pesaresi, G. K. Ouzounis, and L. Gueguen, "A new compact representation of morphological profiles: report on first massive VHR image processing at the JRC," 2012, pp. 839025–839025–6.
- [8] S. Ferri, M. Halkia, and A. Siragusa, "The European Settlement Map 2017 Release; Methodology and output of the European Settlement Map (ESM2p5m)," JRC105679, 2017.
- [9] P. Soille *et al.*, "A versatile data-intensive computing platform for information retrieval from big geospatial data," *Futur. Gen. Comput. Syst.*, 2017.
- [10] M. Pesaresi, V. Syrris, and A. Julea, "A New Method for Earth Observation Data Analytics Based on Symbolic Machine Learning," *Remote Sensing*, vol. 8, no. 5, p. 399, May 2016.
- [11] M. Pesaresi, C. Corbane, A. Julea, A. Florczyk, V. Syrris, and P. Soille, "Assessment of the Added-Value of Sentinel-2 for Detecting Built-up Areas," *Remote Sensing*, vol. 8, no. 4, p. 299, Apr. 2016.
- [12] K. Rosina, F. B. e Silva, P. Vizcaino, M. M. Herrera, S. Freire, and M. Schiavina, "Increasing the detail of European land use/cover data by combining heterogeneous data sets," *International Journal of Digital Earth*, vol. 0, no. 0, pp. 1–25, Dec. 2018.
- [13] M. H. F. Wilkinson, M. Pesaresi, and G. K. Ouzounis, "An Efficient Parallel Algorithm for Multi-Scale Analysis of Connected Components in Gigapixel Images," *ISPRS International Journal of Geo-Information*, vol. 5, no. 3, p. 22, Mar. 2016.
- [14] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, Jan. 1979.
- [15] L. Dijkstra and H. Poelman, "A harmonised definition of cities and rural areas: the new degree of urbanisation," Working Papers, Jun. 2014.
- [16] A. Florczyk *et al.*, *Description of the GHS Urban Centre Database 2015*. Publications Office of the European Union, 2019.
- [17] P. Soille, "Generalized geodesy via geodesic time," *Pattern Recognition Letters*, vol. 15, no. 12, pp. 1235–1240, Dec. 1994.

List of figures

Figure 1 (a) Delivered imagery per contributing mission and (b) the imagery year distribution.	5
Figure 2. Html interface developed for the visual cloud check.....	6
Figure 3. The output from the Association analysis - Evidence-based Normalized Differential Index in Novi Sad (Serbia). Water is shown with blue colour.....	7
Figure 4. Example of a panchromatic image used for the calculation of the Pantex and the CSL in the area of New Belgrade (Serbia)	8
Figure 5. Textural features (pantex) derived from the panchromatic image (figure 4)	8
Figure 6. CSL2RGB model representation derived from the input panchromatic image (figure 4)	9
Figure 7. The ESM_2015 simplified workflow for level 1 and level 2 processing levels	11
Figure 8. Location and size (in terms of population) of urban centres in Europe	12
Figure 9. Level 3 processing workflow for extracting the distances from built-up areas to open spaces as defined in the Urban Atlas.	13
Figure 10. The full ESM_2015 level 1 resampled product at 5 km pixel size. Pixel values correspond to the BU densities since an aggregation operator was applied to the 2 m product	14
Figure 11. Built-up area detection for level 1 layer (2 m) in Berlin (Germany)	15
Figure 12. Discrimination between residential and non-residential built-up area for level 2 layer (10 m) in Berlin (Germany).....	16
Figure 13. An example of open spaces and Distance to open space layers in Montpellier (France)	17
Figure 14. Distance layer relative to open spaces in Paris. Warmer colours indicate longer distances from BU to open space	17
Figure 15. ESM tiling grid displayed on top of the EEA Sentinel 2 RGB mosaic.	19
Figure 16. The ESM_2015 download interface from the GHSL website.....	20
Figure 17. Including TomTom land use data in the map. A large majority of added polygons corresponded to the class 'Industrial and commercial units' (magenta patches in the box b). Legend for boxes a and c is provided in Figure 19 (a)	26
Figure 18. Breaking down the 'Urban fabric' class based on ESM building density (the grey shades in the box b. correspond to the four selected density intervals). Legend for box a and c is provided in Figure 19 (a and b), respectively	26
Figure 19. The original and refined colour-code matching	26
Figure 20. The target legend of the artificial classes expanded into the fourth hierarchical level (dark grey fill – classes expanded, light grey fill – classes not-changed)	27

List of tables

Table 1. Point based statistics from the confusion matrix	18
Table 2. Performance metrics calculated from the confusion matrix elements	18

Annexes

Annex 1. Nomenclature of ENACT refined Corine Landcover

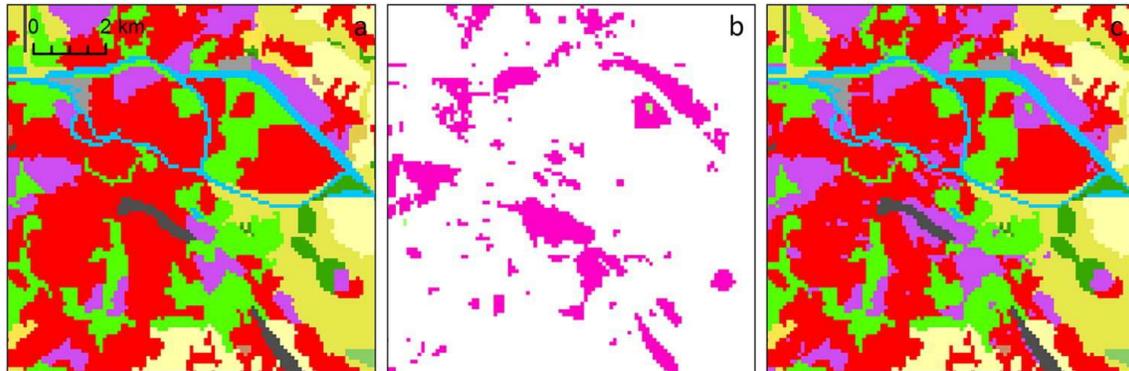


Figure 17. Including TomTom land use data in the map. A large majority of added polygons corresponded to the class 'Industrial and commercial units' (magenta patches in the box b). Legend for boxes a and c is provided in Figure 19 (a)

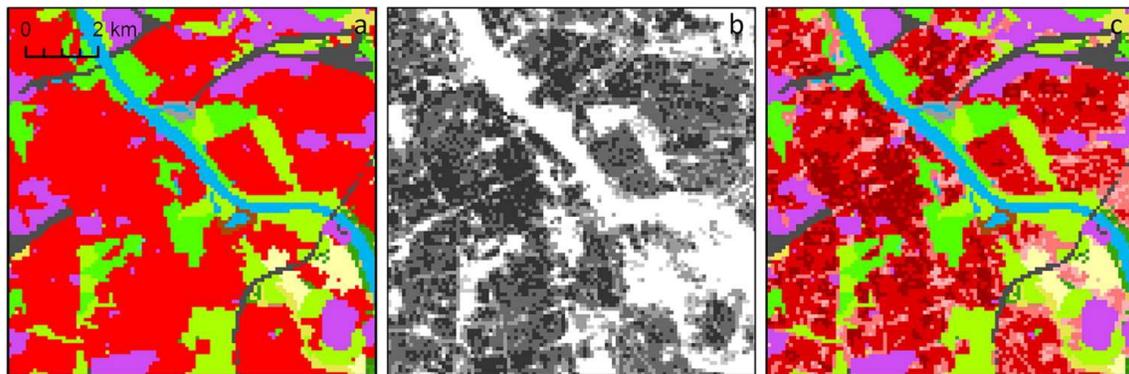


Figure 18. Breaking down the 'Urban fabric' class based on ESM building density (the grey shades in the box b. correspond to the four selected density intervals). Legend for box a and c is provided in Figure 19 (a and b), respectively

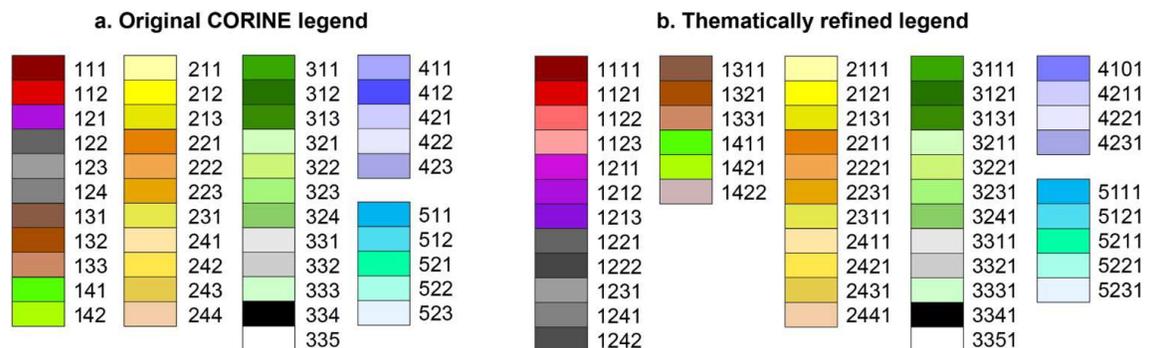


Figure 19. The original and refined colour-code matching

CLC1	CLC2	Level 2 Label	CLC3	Level 3 Label	CLC4	Level 4 Label
1 Artificial surfaces	11	Urban fabric	111	Continuous urban fabric	1111	Urban fabric high density
			112	Discontinuous urban fabric	1121	Urban fabric medium density
					1122	Urban fabric low density
					1123	Urban fabric very low density/isolated
	12	Industrial, commercial and transport units	121	Industrial and commercial units	1211	Production facilities
					1212	Commercial/service facilities
					1213	Public facilities
			122	Road or rail networks and associated land	1221	Road/rail networks and associated land
					1222	Major stations
					123	Port areas
	124	Airport areas	1241	Airport areas		
			1242	Airport terminals		
	13	Mines, dumps and construction sites	131	Mineral extraction sites	1311	Mineral extraction sites
			132	Dump sites	1321	Dump sites
			133	Construction sites	1331	Construction sites
	14	Artificial vegetated non-agricultural areas	141	Green urban areas	1411	Green urban areas
142			Sport and leisure facilities	1421	Sport and leisure green	
				1422	Sport and leisure built-up	

Figure 20. The target legend of the artificial classes expanded into the fourth hierarchical level (dark grey fill – classes expanded, light grey fill – classes not-changed)

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696, or
- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications from EU Bookshop at: <https://publications.europa.eu/en/publications>. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.



EU Science Hub

ec.europa.eu/jrc



@EU_ScienceHub



EU Science Hub - Joint Research Centre



EU Science, Research and Innovation



EU Science Hub



Publications Office
of the European Union

doi:10.2760/25824

ISBN 978-92-76-12184-8