This paper presents the analysis of Earth Observation data records collected between 1975 and 2014 for assessing the extent and temporal evolution of the built-up surface in the frame of the Global Human Settlement Layer project. The scale of the information produced by the study enables the assessment of the whole continuum of human settlements from rural hamlets to megacities. The study applies enhanced processing methods as compared to the first production of the GHSL baseline data. The major improvements include the use of a more refined learning set on built-up areas derived from Sentinel-1 data which allowed testing the added-value of incremental learning in big data analytics. Herein, the new features of the GHSL built-up grids and the methods are described and compared with the previous ones using a reference set of building footprints for 277 areas of interest. The results show a gradual improvement in the accuracy measures with a gain of 3.6% in the balanced accuracy, between the first production of the GHSL baseline and the latest GHSL multitemporal built-up grids. A validation of the multitemporal component is also conducted at the global scale establishing the reliability of the built-up layer across time.

1. Introduction

Growth in human settlements and related built-up areas are among the most profound challenges confronting climate, environmental and socio-economic systems at a global scale. Contemporary technological assets, such as earth observation (EO) and machine learning are enabling the analysis of human settlements with unprecedented detail. There are several EO-derived or EO-supported global maps reporting about physical features directly related to human settlement activities. Most of the maps were produced from low and medium resolution satellite imagery (300 m–10 km) such as MODIS (Schneider, Friedl, & Potere, 2009), Medium Resolution Imaging Spectrometer (MERIS) (Arino et al., 2012), and DMSP-OLS (Zhou et al., 2015). More recently, the availability of new sensors with long time-series triggered the proliferation of moderate to high
resolution (30–50 m) global datasets of human settlements: the Global Land Cover product (GlobeLand30) using Landsat and China Environmental Disaster Alleviation Satellite (HJ-1) (Chen et al., 2015), the Global Urban Footprint (GUF) generated from RADAR satellite constellation of TerraSAR-X and TanDEM-X (Esch et al., 2017) at a spatial resolution of 12 m and the latest multitemporal global urban product derived from Landsat images covering the period 1990–2010 with 5-years interval (Liu et al., 2018).

In line with these developments, the Global Human Settlement Layer (GHSL) project established in 2011 aims at producing global spatial information, evidence-based analytics and knowledge describing the human presence on Earth and its dynamics both in terms of spatial distribution of built-up areas and population.

The theoretical frame of the GHSL is defined inside the established foundation of the human settlement geography: “The suggested core of settlement geography is the buildings – where are they, and why are they there? The focus is on more than dwellings for people, implied as fixed, complete, solid and permanent installation. It includes tents, huts, barrack, barns, equipment sheds, … factories” (Stone, 1965).

In this context, the presence of a “building” is the evidence of the human presence on Earth. It is observable with Earth Observation technologies collecting decametric-scale data available in the open and free scientific domain. These sensors allow recognizing the spatial extent of the human settlements. In the GHSL methodology, satellite data records are translated to the information about the presence of buildings using an automated supervised classification procedure, the Symbolic Machine Learning (SML), that relies on available global land cover data as learning sets (Pesaresi, Syrris, & Julea, 2016). In the GHSL framework, the information about the presence of buildings, is subsequently used to downscale census population data to population grids (Freire, Doxsey-Whitfield, MacManus, Mills, & Pesaresi, 2016), for policy indicators (Corbane, Politis, Pesaresi, Kemper, & Siragusa, 2018) and for a consistent delineation of cities (Melchiorri et al., 2018).

The concept of “built-up area” applied in the GHSL is compliant with the definition of the “building” abstraction in the Infrastructure for Spatial Information in Europe (INSPIRE). The “built-up area” as defined in the GHSL framework is “the union of all the satellite data samples that corresponds to a roofed construction above ground which is intended or used for the shelter of humans, animals, things, the production of economic goods or the delivery of services” (M. Pesaresi et al., 2013).

The baseline information generated in the framework of the GHSL is delivered in an open and free policy context and made available on the JRC Open Data Catalogue (European Commission, Joint Research Centre, 2018a). The GHSL data contributes to the “Human Planet” Initiative of the Group on Earth Observation (GEO). It is part of the “GEOSS data-core” hosted by the Global Earth Observation System of Systems’ (GEOSS) Platform (GEOSS Portal) in conformity with the GEO Data Sharing Principles Implementation (Doldirina, 2015). The first publically released GHSL data were produced in 2014 and included a global multitemporal evolution of built-up surfaces derived from Landsat data collections organized in four epochs 1975, 1990, 2000 and 2014 (GHS_LDSMT_2015 – dataset 1) (M. Pesaresi, Ehrlich, et al. 2016; Pesaresi, Melchiorri, et al. 2016; Pesaresi et al., 2015). After the successful launch of the Sentinel-1 (S1), a proof of concept at global scale for a Sentinel-1-based GHSL data set was performed in 2016(Corbane et al., 2018; Syrris, Corbane, & Soille, 2017). The large-scale experiment
was leveraging on the JRC’s Earth Observation Data and Processing Platform (JEODPP) which is tailored for iterative optimization of big data machine learning methods and fine-tuning of their parameters at full scale (Soille et al., 2017). Noticeable improvements in terms of reduction of omission and commission errors were observed in the new global product derived from Sentinel-1 data (GHS_S1 – dataset 2) (Corbane, Politis, Syrris, & Pesaresi, 2018) in comparison to the Landsat derived built-up areas processed in 2014 (GHS_LDSMT_2015). Both the adaptability and generality of the SML architecture were key to the success of built-up areas extraction from Landsat and Sentinel-1 data. Moreover, the SML framework allowed incremental learning and improvement of the classification outputs as soon as new information on built-up areas became available.

The dataset presented here is a result of introducing the GHS_S1 as learning set for an improved reclassification of the historical Landsat data archive organized in the epochs 1975, 1990, 2000, and 2014. The methodological aspects related to the development of the built-up areas production workflow (GHS_LDSMT_2017- dataset 3) and the multitemporal compositing are detailed. A comparative analysis of the automatic information extraction results from the different workflows is also presented. This is complemented by a data validation protocol which includes a multitemporal accuracy assessment using a global reference dataset and a detailed comparison of the results of the most recent period 2014 against fine-scale digital cartographic reference data reporting the footprint of every single building for 277 sites around the globe. With the new GHS_LDSMT_2017 layer, the GHSL project has generated an up-to-date high resolution, validated and globally consistent dataset with the longest temporal coverage of human settlements available so far for global built-up products.

2. Methods

This section briefly describes the Symbolic Machine Learning (SML) classifier at the heart of the GHSL image classification workflow, the input imagery and the ancillary data used for extracting the multitemporal built-up area grids. The methodological steps followed for the classification of built-up areas at each single epoch and for the production of the final multitemporal grids are also presented.

2.1. The symbolic machine learning for large-scale data analytics

The information extraction tasks included in the GHSL production workflow builds on the SML method that was designed for remote sensing image classification allowing computationally efficient and model-free classification of large amount of satellite data (Pesaresi et al., 2016). The SML schema relies on two relatively independent steps:

- Reduce the data instances to a symbolic representation, also called unique discrete data-sequences;
- Evaluate the association between the unique data-sequences $X$ (input features) and the learning set $Y$ (known class abstraction derived from a learning set).

In the application proposed here, the data-abstraction association is evaluated by a confidence measure called Evidence-Based Normalized Differential Index (ENDI) which
is produced in the continuous $[-1, 1]$ range. The ENDI confidence measure $\Phi$ of data instances $X$ provided the $Y^+$ positive and negative $Y^-$ data instances from the learning set is defined as follows:

$$\Phi(X|Y^+, Y^-) = \frac{f^+-f^-}{f^+ + f^-}$$

where $f^+$ and $f^-$ are the frequencies of the joint occurrences among data instances and the positive and negative data instances, respectively.

To achieve a binary classification (i.e. two class datasets: built-up vs. non-built-up surfaces), a cut-off value of $\Phi$ is automatically estimated for assigning each data sequence to a single class. For the dataset presented here, the OTSU thresholding approach (Otsu, 1979) is used for the purpose of binarizing the ENDI output $\Phi$. The OTSU method chooses an optimal threshold by minimizing within-class variance and maximizing the between-class variance.

The SML automatically generates inferential rules linking the satellite image data to available high-abstraction semantic layers used as learning sets. In principle, any data thematically linked or approximating the "built-up areas" class abstraction with an exhaustive worldwide coverage can be used for deriving human settlements information from any satellite imagery. There is no need for full a priori spatial and temporal alignment between the input imagery, and the learning set nor calibration of the input data as the SML learning process is computationally efficient and can be executed “on-the-fly” for every input satellite scene. Details on the SML algorithm and its suitability for processing of big earth data are provided in (Pesaresi et al., 2016). In Pesaresi et al. (2015), the performance of the SML was compared to alternative supervised classification algorithms such Maximum Likelihood, Logistic Regression, Linear Discriminant Analysis, Naive Bayes, Decision Tree, Random Forest and Support Vector Machine. According to these experiments, at the parity of data conditions (same input image features and same quality or same noise level of the reference set), the SML approach was generally outperforming both parametric and non-parametric classifiers. Furthermore, the better performances were obtained with a much less expensive computational cost. Consequently, the SML classifier was evaluated as the best available solution for large-scale processing of big earth data.

### 2.2. The Landsat data collections

For the production of the multitemporal grids of built-up areas, 33,202 images organized in four Landsat data collections centred at 1975, 1990, 2000 and 2014 (dataset 4) (Florczyk, Politis, Corbane, & Pesaresi, 2018) were processed with the SML classifier as follows:

- 7,597 scenes acquired by the Multispectral Scanner (MS) (collection 1975);
- 7,375 scenes acquired by the Landsat 4–5 Thematic Mapper (TM) (collection 1990);
- 8,788 scenes acquired by the Landsat 7 Enhanced Thematic Mapper Plus (ETM+) (collection 2000) and
- 9,442 scenes acquired by Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) (collection 2014).
The 1975, 1990, and 2000 collections use the Global Land Survey (GLS) datasets preprocessed by the Maryland University and described in Gutman, Huang, Chander, Noojipady, and Masek (2013).

The scenes from the collection 2014 were directly downloaded from the USGS website (United States Geological Survey, 2013). These input data show large heterogeneity in terms of completeness but also in the sensor characteristics evolving from Landsat-1 to Landsat-8. In this data universe, large differences in the radiometric and spatial resolutions (from 15 to 60 m) of the spectral bands, as well as the noise-to-signal ratio of the data, are observed (Table 1). The data availability in the different collections is shown in Figure 1: there are large data gaps in the northern part of South America and the whole Greenland is missing in the 1975 collection. Alike, large parts of Siberia are missing in the 1990 collection. Moreover, the incomplete metadata information for 16.6% and 32.8% of scenes of the 1975 and 1990 collections, respectively, was not allowing for the estimation of the top-of-atmosphere reflectance parameters and consequently directed us towards a classification strategy based on a per-scene machine learning as proposed in this study.

2.3. The learning data

The fully automatic satellite data supervised classification workflow builds on the data-driven SML learning from available ancillary data or “learning set” that are considered the

| Satellite | Landsat-8 | Landsat-7 | Landsat – 4/5 | Landsat 1–3/4–5 |
|-----------|-----------|-----------|---------------|-----------------|
| Thematic relevance of the spectral bands | Land Imager Thermal Infrared Sensor (OLI) | Enhanced Thematic Mapper (ETM) | Thematic Mapper (TM) | Multi Spectral Scanner (MSS) |
| Data type | Band and resolution | Wavelength (µm) | Band and resolution | Wavelength (µm) | Band and resolution | Wavelength (µm) | Band and resolution | Wavelength (µm) |
| Coastal Aerosol | 1 (30m) | 0.43–0.45 | | | | | | |
| Blue band | 2 (30m) | 0.45–0.51 | 1 (30m) | 0.45–0.52 | 1 (30m) | 0.45–0.52 | 4/1 (60) | 0.5–0.6 |
| Green | 3 (30m) | 0.53–0.59 | 2 (30m) | 0.52–0.60 | 2 (30m) | 0.52–0.60 | 5/2 (60) | 0.6–0.7 |
| Red | 4 (30m) | 0.64–0.67 | 3 (30m) | 0.63–0.69 | 3 (30m) | 0.63–0.69 | 6/3 (60) | 0.7–0.8 |
| NIR | 5 (30m) | 0.85–0.88 | 4 (30m) | 0.77–0.90 | 4 (30m) | 0.77–0.90 | 7/4 (60) | 0.8–0.11 |
| Cirrus | 9 (30m) | 1.36 – 1.38 | | | | | | |
| SWIR 1 | 6 (30m) | 1.57–1.65 | 5 (30m) | 1.55–1.90 | 5 (30m) | 1.55–1.75 | | |
| SWIR 2 | 7 (30m) | 2.11 – 2.29 | 7 (30m) | 2.09–2.35 | 7 (30m) | 2.08–2.35 | | |
| Thermal Infrared 1 | 10 (100m) | 10.60–11.19 | 6 (60m) | 10.40–12.50 | 6 (120m) | 10.40–12.50 | | |
| Thermal infrared 2 | 11 (100m) | 11.50 – 12.51 | | | | | | |
| Panchromatic | 8 (15m) | 0.50–0.68 | 8 (15m) | 0.52 – 0.90 | | | | |

Table 1. Characteristics of the different Landsat sensors used in the production of the multitemporal built-up grids. The bands used in the workflow are shown in bold.
best approximation of the target class abstraction to be detected in the satellite data, before the classification exercise. Moreover, the SML characteristics allow for incremental learning and improvement of the classification outputs as soon as new data on built-up areas become available. In other words, the outputs of the SML classifier in one global information extraction session can be used as a learning set in a subsequent classification, improving the overall stability and reliability of the global classification process. In order to provide a quantitative assessment of the improvement gained through the incremental learning, independent reference data (that were not used in the learning) on built-up areas are needed. The whole set of data describing built-up areas and used either for training the SML classifier or for the cross-comparison of the outputs of the classification produced in this study are referred to as “ancillary data”.

The ancillary data used for the generation of the GHS_LDSMT_2017 product include the following global scale grids of urban areas:

- **GHS_LDSMT_2015** (dataset 1) (Pesaresi et al., 2015) is the first publically released GHSL data produced in 2014 and includes global multitemporal evolution of built-up surfaces derived from the four Landsat data collections described in the above section (M. Pesaresi et al., 2016). The learning set used for training the SML classifier at the global level was very heterogeneous combining the best-available spatial information on human settlements at that time. They included various sources with different scales and diverse thematic definitions, completeness and reliability: MERIS Globe Cover artificial surfaces (Arino et al., 2012), LandScan population grids (Dobson, Bright, Coleman, Durfee, & Worley, 2000), Open Street Map (OSM) (OpenStreetMap contributors, 2017), Geonames, MODIS 500 global urban extents (Schneider et al., 2009).
GlobeLand30 (GLC-30) represents the first global open access landcover product at 30-m resolution from fine-scale Landsat imagery acquired in 2010 (Chen et al., 2015; Chen, Yifang, and Songnian, 2014). GLC30 was produced with operational human-operated photo-interpretation techniques using a hybrid pixel- and object-oriented image processing approach. It includes nine thematic classes with one associated to “artificial surfaces” mainly consisting of urban areas, roads, rural settlements and open pit mines and quarries. The “artificial surfaces” abstraction included in the GlobeLand30 (GLC30) is used here as a proxy (learning set) for the extracting human settlements from Landsat satellite image.

GHS_S1 (dataset 2) (Corbane, Politis, Syrris et al., 2018; Corbane, Lemoine, Pesaresi et al., 2018) corresponds to the first global map of built-up areas derived from Sentinel-1 produced in the framework of the GHSL at a spatial resolution of 20 m. The experiment was performed in 2016 and fully completed in 2017 exploiting over 7,000 Sentinel-1A and 1B scenes (acquired in 2016–2017). A dedicated workflow building on the SML classifier was developed for adapting the GHSL production workflow to the Sentinel-1 data. The information extraction of Sentinel-1 data at global scale is described in Corbane Politis, Syrris, et al. (2018). The learning data at the global level consisted of the union of i) the built-up layer GHS_LDSMT_2015 obtained from the Landsat imagery and ii) the GlobeLand30 (GLC30).

In addition to the learning data related to built-up areas, urban areas or artificial surfaces, the workflow for the production of the multitemporal built-up grids (GHS_LDSMT_2017) includes learning data reporting about the presence of water at the same scale of the data under processing. Information on water presence was retrieved from the Global Surface Water (GSW) layer produced by the JRC for the period 1984–2015 (Pekel, Cottam, Gorelick, & Belward, 2016). In particular, the water occurrence data have been used in the updated workflow as learning set input for the SML, aiming to extract the surface water mask per each Landsat scene. The water occurrence is a measurement of the water presence frequency (expressed as a percentage of the available observations over time actually identified as water). The provided occurrence accommodates for variations in data acquisition over time in order to provide a consistent characterization of the water dynamic over time.

Table 2 provides a summary of all ancillary datasets employed in the updated workflow presented in the next section. Both input data collections and ancillary data are included in this table together with their geometric, temporal, and thematic specifications.

### 2.4. Workflow for built-up extraction from Landsat data collection and for the generation of a multitemporal composite (GHS_LDSMT_2017)

Using the SML as the backbone of the GHSL production, the generation of the updated multitemporal grids of built-up areas is obtained through four main processing steps illustrated in the flowchart of Figure 2:

1) Per-scene cloud/shadow detection aimed at identifying the valid image data domain before the thematic information extraction;
2) Per-scene thematic information extraction using inductive reasoning operationalized by SML. The purpose of the inductive reasoning is to detect land vs. water surfaces and built-up surfaces, including alternative learning set options for the data corresponding to the most recent epoch. The outputs from the different learning set options are then merged following a majority voting approach;

3) Generation of single epoch thematic information mosaics;

4) Generation of a multitemporal thematic information composite.

The main commonalities and differences between the workflows used for the production of the GHS_LDSMT_2015 and GHS_LDSMT_2017 from the same input

---

**Table 2. Learning data employed in the updated production workflow of Landsat imagery.**

| Learning sets              | Reference                                      | Producer       | Pixel size (meters) | Product time stamp | Imagery time stamp | Release date |
|----------------------------|------------------------------------------------|----------------|--------------------|--------------------|--------------------|--------------|
| GHS_LDSMT_2015             | Pesaresi et al. (2016)                          | EC JRC         | ~ 38               | 1975–1990-2000–2014| 1972–2014          | 2015         |
| GHS_S1                     | Corbane, Pesaresi, Politis, Syrris, Aneta et al. | EC JRC         | ~ 20               | 2016               | 2016              | 2017         |
| GlobeLand30 (GLC-30)       | Chen et al. (2015)                              | National Geomatics Center of China | ~ 30 | 2013 | 2000 – 2010 | 2014 |
| Global Surface water (water occurrence) | Pekel et al. (2016) | EC JRC | ~ 30 | 2015/2016 | 1984–2015 | 2016 |

---

**Figure 2.** Basic scheme of the production framework implemented to generate the new multitemporal Global Human Settlement Layer (GHS_LDSMT_2017).
data are summarized in Table 3. The updated workflow (from the pre-processing of input data until the generation of the final multitemporal product) has been designed as a fully automated processing chain suitable for being executed in a high-throughput computing facility and optimized according to the configuration of the JEODPP platform.

### 2.4.1. Per-scene cloud/shadows screening

In the workflow discussed here, the cloud detection methodology was revised compared to the previous GHS_LDSMT_2015 workflow. The cloud detection identifies the valid image data domain before the thematic information extraction. In the GHS_LDSMT_2015 production workflow, clouds were extracted using a cloud detection algorithm (Soille, 2011) for epoch 1975, 1990 and 2000, and using the Quality Assessment (QA) band of Landsat 8 for the epoch 2014. The latter actually captures, in addition to cloud contaminated pixels, also highly reflecting rooftops of e.g. new industrial/commercial buildings. Hence, highly reflecting large building structures have high probability to be erroneously classified as cirrus or thick clouds and are therefore excluded from the data domain. To deal with the problem of incorrect assignment of some valid pixels, the Fmask algorithm (Zhu et al., 2015) that was designed for Landsat 8 images was applied to the 9,089 satellite scenes constituting the data collection of 2014. The example shown in Figure 3 provides an indication of the added-value of re-
calculating the cloud mask using the Fmask methodology in comparison to the QA band of the Landsat 8 data.

2.4.2. Per-scene information extraction

In this study, the full training and classification workflow was applied independently to each individual satellite scene, without classification model transfer from one scene to other scenes, as it is common practice in similar data classification tasks (Gong et al., 2013). In this way, the automatic recognition model is optimized with respect to the specific data collection conditions of each processed scene. The results of built-up areas detection obtained in the GHS-S1 product in 2017 confirmed the interest of using Sentinel-1 data for the detection of sparse rural settlements and for a refined delineation of built-up areas in comparison to the first Landsat derived built-up areas (GHS_LDSMT_2015): notable improvements – in terms of reduction of omission and commission errors – were observed in this new global product derived from Sentinel-1 data (GHS_S1) (Corbane, Lemoine, Pesaresi et al., 2018). Hence, the results from the processing of Sentinel-1 enable incremental learning of the SML classifier for an enhanced delineation of built-up areas from historical Landsat collections. In the case of the most recent Landsat 8 input data collection, several parameter combinations for the image features' extraction and different options for the training of the SML classifier were tested by leveraging on the JEODPP infrastructure for big data processing. They are briefly summarized below:

- Radiometric and textural input image features: In terms of calculated image features, the workflow for processing Landsat multispectral imagery (i.e. Landsat 8 (collection 2014), ETM+ (collection 2000), TM (collection 1990) and MS (collection 1975)) relies essentially on radiometric features used as input to the SML workflow. However, in the case of Landsat 8 data, the availability of a panchromatic band at a spatial resolution of 15 m allows using texture features for a more refined

Figure 3. Comparison between Landsat 8 Quality band (b) and Fmask cloud detection (c) output for a sample Landsat 8 scene (a). The image shown in a) is a 3 band natural colour composition with Red: Band 4 (NIR), Green: Band 3 (RED), Blue: Band 2 (BLUE).
delineation of built-up areas. Hence, textural analysis was applied on the panchromatic band using the PANTEX methodology which is based on a set of rotation-invariant, anisotropic contrast texture measurements calculated on the Grey Level Co-occurrence Matrix (GLCM) of the input image (Pesaresi, Gerhardinger, & Kayitakire, 2008). For the purposes of this work, a textural feature was calculated with a window size of 75 m and a displacement vector of one pixel for the eight principal directions. While the use of texture can significantly reduce commission problems especially in sandy dunes, river beds and agricultural areas, it may lead sometimes to excessive refinement with under-detection of large industrial buildings areas and in very compact settlement patterns with narrow internal roads (thinner than the sensor resolution) such as historic (medieval) city centres or some large slum areas. Therefore, in the case of Landsat 8 collection, two options were included in the experiment leading to two built-up layers, with and without textural refinement, respectively.

• Alternative learning sets: In terms of the SML classifier training, we tested independently two learning sets in the case of Landsat 8; the artificial surfaces derived from GLC-30, and the built-up areas derived from Sentinel-1 (GHS_S1).

In practice, the following four parameters’ combinations were tested on the Landsat 8 collection:

• GLC-30 as a learning set and without textural refinement (output confidence measure denoted as $\Phi_{BU,GLC30}$);
• GLC-30 as learning set and using pantex textural refinement (output confidence measure denoted as $\Phi_{BU,GLC30,p}$);
• GHS_S1 as a learning set and without textural refinement (output confidence measure denoted as $\Phi_{BU,S1}$);
• GHS_S1 as a learning set and using pantex textural refinement (output confidence measure denoted as $\Phi_{BU,S1,p}$);

The binary outputs from the different learning set options ($\Phi_{BU,GLC30}, \Phi_{BU,GLC30,p}, \Phi_{BU,S1}, \Phi_{BU,S1,p}$) are then merged on the basis of majority voting approach for deriving the binary built-up layer $\Phi_{BU}$ for 2014 (the variables are expressed in a vector format, implying that $\Phi_{BU} = \Phi_{BU}(x)$, where $x$ denoting the pixel value; accordingly, the operators used have an element-wise functioning):

$$\Phi_{BU} = \begin{cases} \mathbf{1}, & \Phi_{BU,GLC30}^{T} + \Phi_{BU,GLC30,p}^{T} + \Phi_{BU,S1}^{T} + \Phi_{BU,S1,p}^{T} \geq 2 \\ \mathbf{0}, & \text{otherwise} \end{cases}$$

The remaining collections (collection 1975, collection 1990 and collection 2000) were all processed using only radiometric features and with GHS_S1 product as input learning set.

Compared to the procedure adopted for generating the first GHSL multitemporal product in 2014/2015 (GHS_LDSMT_2015), the improved workflow uses more detailed learning sets on built-up areas with a consistent global coverage. Besides the current implementation, an incremental learning has been tested in the SML framework. With
this approach, the SML is trained using built-up areas derived from Sentinel-1 (GHS_S1), which in turn were derived from the first GHSL multitemporal product, leading to a successive refinement (Figure 4).

This learning method has the advantage of achieving superior efficiency for training while minimizing the impact on accuracy. Another beneficial point is that the performance of the classifier can slightly improve. The results are improving after each iteration since the level of noise is progressively decreasing with each run of the GHSL workflow as shown in Pesaresi et al. (2015).

It is worth noting that the demonstrated robustness of the SML to noise related either to scale generalization, spatial displacement, and thematic content including temporal mismatch, suggests that in the case of temporal mismatch (i.e. using learning set from 2015–2016 to classify images from 1975 or 1990), the SML shows a compact behaviour, independent from the training set and displaying high performance despite high noise levels (Pesaresi, Syrris, & Julea, 2015).

2.4.3. Single-epoch tiling and mosaicking

As already mentioned, the GHSL information production strategy is based on a per-scene learning and classification approach, with no need to transfer the decision rules (linking satellite data instances with the class abstractions) learned in one scene to another satellite scene. The processing framework building on the SML, delivers for each scene of the four Landsat data collections the following outputs:

- **ENDI**: is the ENDI confidence measure $\Phi$ in the continuous range $[-1, 1]$ expressing strength of association between the image data instances and the class abstractions selected in the learning set. Values close to 1 indicate that the feature sequence is strongly associated with the image class of interest while values close to $-1$ indicate that the feature is strongly associated with the classes other than built-up.
• \(bu_{\text{binary}}\): is a binary layer corresponding to the binarized ENDI confidence measure following the Otsu thresholding approach (Otsu, 1979). Built-up areas are coded with the value 1 and non-built-up with 0.

• \(ENDI_{\text{water}}\): is the ENDI confidence measure of the water layer in the continuous range \([-1, 1]\).

• \(water_{\text{binary}}\): is a binary layer corresponding to the binarized ENDI confidence measure of the water following the Otsu thresholding approach (Otsu, 1979). Pixels classified as "water" are coded with the value 1 and non-water with 0.

• \(land_{\text{binary}}\): is a binary layer corresponding to the logical complement of the \(water_{\text{binary}}\) layer. Pixels classified as "land" are coded with the value 1 and water with 0.

• \(datamask_{\text{binary}}\): is a binary layer corresponding to the valid pixels following the exclusion of the cloud and shadows contaminated pixels detected with the Fmask method.

In order to manage the redundancy among the thematic outputs obtained from overlapping individual scenes, a global reference grid system is required. In our study, each output scene was projected to the same global projection using an absolute spatial tiling schema in order to optimize the data management and I/O speed during the subsequent processing phases. In particular, a World Geodetic System 1984/ Pseudo-Mercator global projection (EPSG:3857) framework was applied, with a spatial indexing data based on regular tiles of 150 km x 150 km. A mosaicking procedure tailored to the type of output data (binary or continuous) merges overlapping tiles and generates homogeneous raster datasets organized into four epochs: 1975, 1990, 2000 and 2014.

The mosaics of the outputs at 30 m resolution represent the input data for the subsequent multitemporal data fusion step.

2.4.4. Multitemporal built-up grids

The new multitemporal built-up extraction consists in aggregating the results of the built-up layers obtained from the processing of individual Landsat data collections (Landsat8, collection 2000, collection 1990, collection 1975), previously tiled and mosaicked.

The fusion of built-up maps from different dates relies on the following general assumptions:

• The built-up areas of the most recent year (2014) are considered as the most reliable and the basis of production of the multitemporal image stack;

• An increase of built-up over time is assumed;

• The data gaps in any given epoch are filled with the closest information available from the earlier epochs;

• Redundant thematic information, if available, is used to decrease error rates by application of voting schemas.
The continuous growth in built-up is an assumption with a clear limitation, however, it allows us to make more robust the multitemporal information extraction against scene-specific satellite data collection parameters, and is considered valid for the large majority of the global built-up areas accounted for in the final multitemporal product (GHS_LDSMT_2017). Consequently, the data cannot describe processes that involve a destruction or removal of buildings. The multitemporal processing is performed in parallel for each tile of the global mosaics (i.e. corresponding to the merged overlapping tiles). The aim of this task is to fill remaining data gaps in the specific epochs, and to reduce as much as possible error rates by adopting a voting schema, if multiple thematic outputs are available for the same data sample (pixel).

The final GHS_LDSMT_2017 is a multi-class product encoding not only the multitemporal built-up maps, but also the water class, the land class and the no data class. The model developed for the multitemporal fusion and extraction of the different classes is described in detail per thematic class in the following sections.

### 2.4.4.1. Built-up classes for 1975, 1990, 2000 and 2014.

In this phase, we create first the best possible representation of the thematic information in the most recent epoch, and second we go stepwise back in time checking if there are data evidences supporting the thematic information change.

This first step aims at creating a “best possible” binary thematic information layer for the most recent epoch 2014 by adopting voting schemas for solving thematic disagreement and spatiotemporal reasoning for filling data gaps.

All the following equations refer to pixel values in a vector format, while the applied operators have an element-wise functioning.

Let \( bu_{binary} \) be the single binary built-up maps extracted for each of the four epochs. The binary layer for 2014 (\( base_{2014} \)) was calculated as:

\[
base_{2014} = bu_{binary_{2014}} \lor bu_{agg}
\]

where \( bu_{agg} = \begin{cases} 
1, & bu_{binary_{2000}} + bu_{binary_{1990}} + bu_{binary_{1975}} = 3 \\
0, & \text{otherwise}
\end{cases} \]

For the remaining dates (1975, 1990 and 2000), we adopted a step-wise hierarchical process progressively going back in time and starting from the built-up areas as reported in the ideal binary layer for 2014 (\( base_{2014} \)). At each step, we subtract from a given epoch \( t \), the built-up areas detected in epoch \( t-1 \) following the assumption of continuous growth in built-up. This is done through a progressive thresholding of the confidence layers \( ENDo_{bu} \) following a series of rules to determine the threshold value of the confidence layers. Those rules take into account the statistical distribution (estimated by the average \( \mu \) and the standard deviation \( \sigma \)) of the confidence layers in the domain corresponding to non-built-up pixels in the learning set. For instance, the confidence cut-off value \( T \) supporting the 2000 \( ENDo_{bu_{2000}} \) is defined as the mean of confidence layer \( ENDo_{bu_{2000}} \) within the non-built-up pixels of the learning set plus its standard deviation:

\[
T = \mu + \sigma
\]

where:
\[
\mu = \text{avg}(x_i : \forall x_i \in [ENDo_{bu_{2000}} \otimes \neg ref\_set])
\]
\[ \sigma = \text{std}(x_i : \forall x_i \in [\text{ENDI}_{bu,2000} \otimes \neg \text{ref}_{set}]) \]

where the operators \( \otimes \) and \( \neg \) denote the element-wise multiplication and logical negation, respectively. The \( \text{ref}_{set} \) is the learning set consisting of 0s for non-built-up pixels and 1s for built-up.

The resulting binary built-up area for 2000 \( \text{base}_{2000} \) is consequently extracted from \( \text{base}_{2014} \) by applying the following operator:

\[ \text{base}_{2000} = \text{bu}_{binary,2014} \land (\text{ENDI}_{bu,2000} > T) \] (5)

The adaptive thresholding approach tries to minimize the effects on the \( \text{ENDI}_{bu} \) scores generated by the mutual interaction of several factors not explicitly modelled in the production process: they include mainly i) the different sensor characteristics, ii) the data collection conditions (season, illumination, atmosphere) and iii) the geographic conditions (building practices patterns and materials, background land cover, orography) present in the mosaic. It is also worth noting that in case of data gaps this iterative thresholding approach propagates the information from the most recent available data records and hence allows filling data gaps.

2.4.4.2. Water, land and no-data classes. For the aggregated no-data class, we consider the union of all the binary data masks (dm) of all the four epochs. The non-valid pixels are calculated as belonging to the complement of the set of valid pixels observed in all four epochs:

\[ \text{Datamask} = \text{dm}_{binary,2014} \lor \text{dm}_{binary,2000} \lor \text{dm}_{binary,1990} \lor \text{dm}_{binary,1975} \] (6)

\[ \text{Nodata} = \neg \text{Datamask} \] (7)

For generating the water vs. land classes, we exploit the binary layers obtained from the processing of the single epochs and applied a voting schema:

Let \( \text{Water}_\Sigma \) and \( \text{Land}_\Sigma \) be the sums of the multitemporal binary layers of water and land respectively:

\[ \text{Water}_\Sigma = \sum_{t \in \{1975, 1990, 2000, 2014\}} \text{water}_{binary,t} \] (8)

\[ \text{Land}_\Sigma = \sum_{t \in \{1975, 1990, 2000, 2014\}} \text{Land}_{binary,t} \] (9)

The values of the sums can range from 0 to 4 (i.e. 0 corresponding to full agreement between all epochs on absence of water and 4 corresponding to full agreement between all epochs on the presence of water). Due to the dynamic nature of the water and land classes, the final classes are computed following the majority voting schema and taking into account the valid data per each epoch (all the operations work in an element-wise manner):

\[ \text{Water} = \text{Water}_\Sigma > 0.5 \times \sum_{t \in \{1975, 1990, 2000, 2014\}} \text{datamask}_{binary,t} > 0 \] (10)

\[ \text{Land} = \text{Land}_\Sigma > 0.5 \times \sum_{t \in \{1975, 1990, 2000, 2014\}} \text{datamask}_{binary,t} > 0 \] (11)
A description of the workflow is provided in Supplementary File 1, in the form of a pseudo-code describing the main processing steps.

2.5. Estimating the built-up surface evolution and the corresponding aggregated densities – first figures derived from the new multitemporal built-up grids

The quantitative assessment has been performed using built-up area density gridded at 1 km spatial resolution. The World Mollweide projection (EPSG:54,009) is selected in order to aggregate the global results from an equal area projection input.

The total built-up area reported in the GHS_LDSMT_2017 per epoch is as follow:

- In 2014, the built-up areas is 789, 385 km$^2$
- In 2000, the built-up areas is 655, 742 km$^2$
- In 1990, the built-up areas is 523, 333 km$^2$
- In 1975, the built-up area is 379, 552 km$^2$

Figure 5 depicts the density of built-up areas as obtained from the GHS_LDSMT_2017 product for the epochs 1975, 1990, 2000 and 2014. The original built-up gridded data were aggregated from 30 × 30 m to 100 × 100 km pixels for better display at a global scale. Overall, the map shows that the highest densities of built-up areas in 2014 can be seen in China, Europe and in East and West of the United States. Europe shows the highest proportion of land covered by built-up pixels with percentages greater than 1% of the 100 × 100 data tile for almost all the land surface.

3. Validation and quality assessment

3.1. Overview of the validation approach and the reference data

The validation approach presented here aims at achieving a comprehensive and systematic description of the quality and validity of the GHSL built-up areas product:

- First, we explore the quality of the multitemporal composite GHS-LDMST_2017 compared to GHS-LDMST_2015 through a visual inspection of the results obtained over selected cities (section 3.2).
- Second, to rigorously assess the accuracy of the multitemporal composite GHS-LDMST_2017, we compare it against a global multitemporal reference dataset recently made available in Liu et al. (2018). This statistically representative reference data were originally produced for the assessment of the Landsat based multitemporal product developed by Liu et al. (2018) for the 1990–2010 period with a five – year interval (section 3.3).
- Third, we conduct a quantitative assessment of the performance of the GHS-LDMST _2017 compared to GHS-LDMST_2015 and GHS_S1 building upon detailed reference data and a selection of accuracy metrics. The purpose is to benchmark the three products by using accuracy metrics estimated at the sensor resolution and to establish an understanding of the accuracy of the mapped settlement with regard to the characteristics of the input satellite sensor (section 3.3).
Fourth, we perform a pattern-based assessment focusing on the relationship between density metrics derived from the independent reference data and the automatically derived built-up layers (section 3.5). This is implemented through a regression analysis which aims at evaluating the GHSL output from the perspective of using the satellite-derived information as input for population modelling (Freire et al., 2016) and derived policy indicators (Corbane et al., 2018). The knowledge of the systematic bias and gain parameters of the automatically-generated GHSL thematic products is a first step toward the thematic standardization of the GHSL built-up areas products and the minimization of systematic differences related to use of different satellite sensor data as input.

For the technical validation of the GHSL built-up areas product and in particular for the accuracy and regression analysis (sections 3.3 and 3.5), a reference spatial database including single building delineation derived from digital cartography at a nominal scale of 1:10,000 was developed. An inventory of available digital cartographic sources was performed in order to compile the most possible representative sample set from different cities around the globe, taking a best-effort approach given the available resources. The reference database gathered through this exercise consists of more than 40 million individual building polygons selected from 277 different areas of interest. These are mostly local administrative units covering specific cities or full counties (for the United States of America) and spread across the different continents. While not

Figure 5. Percentage densities of built-up areas in the four periods (1975, 1990, 2000 and 2014) as obtained from the GHS_LDSMT_2017 product aggregated to a spatial resolution of 100×100 km. No data is represented with white.
covering all the combinations of geographical, environmental, and cultural conditions that are determinant factors of the settlement patterns, the reference data correspond to very heterogeneous types of landscapes. The reference years for the collected reference data range between 2012 and 2018. This allows assessing only the most recent period 2014 of the GHSL.

The reference building footprints span over the whole spectrum of low density and high-density human settlement patterns, representing typical rural, suburban and urban spatial patterns (Supplementary File 2). Prior to their inclusion in the reference database, the sources were revised by experts in order to assess: i) their reliability, ii) their completeness including a manual spatial delineation of the valid data domain by trained remote sensing analysts, and iii) the date of last update. The revision was based on the comparison of the cartographic sources with very high-resolution satellite imagery as reported by Google, Bing, and ESRI global platforms. From an initial collection of 350 areas of interest with available building footprints, only those sites where all buildings were outlined were retained for the validation exercise. To our knowledge, the developed reference spatial database represents the most exhaustive fine-scale reference dataset used for the validation of satellite-derived spatial information on built-up areas.

In order to support the validation protocol, the reference data collected in vector format were converted into binary raster layers indicating the presence/absence of built-up areas. The rasterization of the vector cartographic data was performed at a spatial resolution of 1-m corresponding to the spatial tolerance admitted for a nominal scale of 1:10 000 (Freire et al., 2014).

As already introduced, in the GHSL framework the “built-up area” is defined as: “the union of all the satellite data samples that correspond to a roofed construction above ground which is intended or used for the shelter of humans, animals, things, the production of economic goods or the delivery of services” (Pesaresi et al., 2013). The operationalization of the above concept in the technical validation presented here is translated to the following rule: “any grid cell that overlaps (intersection ≠\emptyset) with any built-up or roofed structure is considered as belonging to the class built-up”.

In compliance with this definition, the gridded reference data were aggregated to the same spatial resolution of the GHSL information under test (i.e. GHS-LDMST_2017), hence to the spatial resolution of 30 m and projected to the Mollweide (EPSG: 54009) as an equal area projection (i.e. allowing to preserve areal relationships). In the same way, the other GHSL products (GHS_S1 and GHS-LDMST_2015) were reprojected to the Mollweide and resampled to the GHS-LDMST_2017 resolution to ensure compatibility between all the layers.

The common reference baseline grid was subdivided in spatial Statistical Units (SU-tiles) with a size corresponding to 900 × 900 m on the ground in order to support the continuous linear regression fit tests. The 900 × 900 m tile size was selected taking into account the following considerations:

- to align the tile sizes to a multiple of 30 m which corresponds to the spatial resolution of the satellite-derived information under assessment,
- to align with the scale of the policy indicators supported by the GHSL, including the “Degree of Urbanisation” urban vs. rural classification schema (Dijkstra & Poelman, 2014),
to allow to capture the internal variability of the test areas, providing support for
the stratification of the results by different built-up surface densities class in urban,
suburban and rural areas (Sabo, Christina Corbane, Florczyk, Pesaresi, & Kemper, 2018).

A total of 342,568 SU-tiles of 900 × 900 m each, derived from the vector reference
cartographic data were used during the technical validation covering a total surface area
of 277,480 square kilometres.

3.2. Qualitative assessment

The re-engineered and consolidated workflow for thematic information extraction from
Landsat data collections resulted in an enhanced multitemporal information layer reporting
about the presence of built-up areas. Visual inspection of the GHS_LDSMT_2017 layer
showed significant reduction of both over-detection and under-detection errors. The over-
detections were observed in sandy or rocky beaches, bright bare soils and riverbeds that
were misclassified as built-up areas in the first GHS_LDSMT_2015 product, especially in
Africa. In terms of omissions, Asia is the continent where the highest misdetections were
observed in the GHS_LDSMT_2015 product (Corbane et al., 2017). Those omission rates are
most probably related to the smaller sizes of settlements and stronger fragmentation mainly
in India and China. The visual comparison of the results of GHS_LDSMT_2015 and
GHS_LDSMT_2017 are a clear evidence of the refined built-up areas detection. Figure 6(a)
is an example of such enhanced capabilities covering the megacity of New Delhi and the
surrounding villages. It corresponds to a typical case of under-detection of small and
scattered settlements in rural areas in the first Landsat-derived GHSL product. The
GHS_LDSMT_2017 output clearly depicts the newly detected settlements in the rural fringes
of New Delhi and their temporal dynamic.

Another example highlighting the ability of the new workflow to delineate settle-
ments that remained undetected in GHS_LDSMT_2015, is shown in Figure 6(b) for
Shanghai. Most of the newly detected settlements are located on the islands of
Chongming, Changxing and Hengsha, which form the Chongming County. Differences
in the spatial patterns of built-up growth can be also observed between the two
products. The latter is related to the differences in the multitemporal processing work-
flow, driven essentially by differences in baseline layer for 2014 and the confidence
layers (ENDI) obtained for epochs 2000, 1990 and 1975. The example of Riaydh shown in
Figure 6(c) illustrates the problems of underdetection of built-up areas in 1975 and
confusion of spectral signatures among bare soils and the built-up areas that are
significantly reduced in GHS_LDSMT_2017 product. It is worth noting here that in the
GHSL framework no post-editing phase (either manual or automatic) is applied to the
output of the adopted fully automated and reproducible processing chain.

3.3. Multitemporal performance assessment

For the multitemporal accuracy assessment, we exploited a recent reference dataset made
available by Liu et al. (2018) and originally used for the validation of global urban land maps
based on Landsat images for the 1990–2010 period with five year-year internal. The
Figure 6. Examples from selected cities showing the reduction of over-detection and under-detection errors in the new multitemporal built-up layer. The results of GHS_LDSMT_2015 are compared against those of GHS_LDSMT_2017 respectively for the cities of New Delhi (India Figure 6(a)), Shanghai (China Figure 6(b)) and Riyadh (Saudi Arabia, Figure 6(c)).
reference sample has been selected following a stratified random sampling design that relies on LandScan 2010 data for stratifying the globe into urban and non-urban blocks of 6 × 6 km according to the criterion of >1000 persons/km² (= 36 000 per block). Then taking into account cost and precision considerations, 150 sample blocks were selected in the “urban blocks” stratum and 150 sample blocks in the non-urban stratum. In each sample block, urban land distributions was assigned using manual interpretation on the basis of the multitemporal Landsat images available in Google Earth. Details on the multitemporal reference dataset including the sampling design, the sample selection and spatial distribution, and the procedure of manual interpretation are described in Liu et al. (2018).

The reference dataset can be accessed through the link:
http://www.geosimulation.cn/GlobalUrbanLand.html.

Both the representativeness of the sample of global urban and non-urban land and the comparable masks of built-up extents in terms of input Landsat data, spatial resolution and time period, make this reference database suitable for the validation of the multitemporal component of the GHS_LDSMT_2017. In particular the reference sample blocks for the years 1990, 2000 could be directly compared to the built-up areas of 1990 and 2000 periods of the GHS_LDSMT_2017. Despite the time difference, the reference data for 2010 have been also considered for the validation of the 2014 mask of built-up areas of GHS_LDSMT_2017 given that collection 2014 of Landsat data includes a combination of data from 2013 and 2014.

Using this reference dataset, we performed a per-pixel accuracy assessment to validate the built-up areas of GHS_LDSMT_2017. Standard accuracy and error metrics derived from

![Figure 7](image)

**Figure 7.** Distribution of the balanced accuracy for the three periods 1990, 2000 and 2014 in the 300 samples of urban and non–urban blocks. The dashed-lines correspond to the average-balanced accuracy.
the confusion matrix were calculated (Congalton, 1991). Given the lack of a single universally accepted measure of agreement, we use a combination of three main performance metrics to give a complete picture of performance of the multitemporal dataset: the balanced accuracy (Brodersen, Soon, Stephan, & Buhmann, 2010) (Figure 7), the commission and the omission errors (Table 4). The results show a good agreement between the reference data and the automatically generated built-up masks. The values of the accuracy and error metrics are almost stable for the three periods highlighting the robustness of the built-up extraction workflow and the multitemporal compositing method. The slight decrease of the balanced accuracy in 2014 can be attributed to the time gap between the reference data collected in 2010 and the dates of the Landsat data in collection 2014.

3.4. Accuracy assessment using detailed building footprints

In this phase, the performance of the GHS-LDMST_2017 is compared to GHS-LDMST_2015 and GHS_S1 building on the basis of the detailed reference building footprints and a selection of accuracy metrics. Using the total of 342,568 SU-tiles of 900 × 900 m derived from the reference building footprints (section 3.1), per-pixel accuracy assessment was performed. The statistics associated with the confusion matrices of the three products under assessment were computed (Table 5), together with the three main performance metrics to give a complete picture of the differences between the products: the balanced accuracy the commission and the omission errors (Table 6).

The pixel-based accuracy assessment entails that all four layers (the three settlement layers under assessment and the reference layer) are projected to the same geometry, resampled to the same grids and tiled following a unique tiling schema.

The overall results of the validation among the GHS-LDMST_2015, the GHS_S1, and the GHS-LDMST_2017 built-up grids produced in the context of the GHSL are shown in Table 6. According to these results, the new GHS-LDMST_2017 provides the best balanced accuracy (BA), yielding BA = 0.86. Most importantly, we observe a gradual improvement in the accuracy measures starting from the first layer GHS-LDMST_2015 (BA = 0.83), the GHS_S1 layer (BA = 0.84) and the last GHS-LDMST_2017. The observed

| Product               | True positives | False positives | True negatives | False negatives |
|-----------------------|----------------|-----------------|----------------|-----------------|
| GHS-LDMST_2015        | 358,117,078    | 308,624,756     | 2,266,656,161  | 10,109,722      |
| GHS_S1                | 353,313,345    | 255,604,646     | 2,319,676,271  | 105,900,960     |
| GHS-LDMST_2017        | 375,437,040    | 269,400,098     | 2,305,880,819  | 83,777,265      |

Table 4. Performance metrics derived from the per-pixel accuracy assessment of the GHS_LDSMT_2017 using 300 reference sample blocks per date obtained from Liu et al. (2018).

| Performance metrics | 1990 | 2000 | 2014 |
|---------------------|------|------|------|
| Average Balanced Accuracy | 0.72 | 0.72 | 0.71 |
| Average Omission Error   | 0.27 | 0.26 | 0.28 |
| Average Commission Error | 0.46 | 0.42 | 0.36 |

Table 5. Statistics derived from the pixel-based confusion matrix.
trend of increasing BA rates in subsequent generations of GHSL data including the precedent releases as learning set, is an empirical confirmation of the positive impact of the incremental learning approach of the GHSL framework (Figure 4).

The disaggregation of the performance metrics per continent (Figure 8) shows that the best accuracy in the GHS-LDMST_2017 dataset was obtained in Oceania, with median BA of 0.82 (7,544 validated SU-tiles), followed by Africa (1,435 validated SU-tiles) and Europe (15,820 validated SU-tiles) with median BA of 0.79 and 0.78, respectively. The most noticeable improvements are observed in Asia (1,936 validated SU-tiles) followed by the Americas (315,833 validated SU-tiles) and Africa where the GHS-LDMST_2017 contributes to a prominent increase of the BA.

Despite the overall improvements observed in the GHS-LDMST_2017 product compared to its predecessors, a higher omission error was obtained with the most recent product GHS-LDMST_2017 in Africa compared to GHS-LDMST_2015 (median OE of 0.16 and 0.10, respectively). This increased omission errors can be explained by the use of the textural refinement in the workflow in an attempt to decrease overdetection of built-up areas. While the use of texture indeed reduced significantly commission errors, it resulted in excessive refinement with under-detection of compact settlement patterns with narrow internal roads (thinner than the sensor resolution) as in ancient-development cities and villages, and in some large slum areas.

Table 6. Performance metrics derived from the pixel-based accuracy analysis at a spatial resolution of 30 m.

| Performance metrics | GHS-LDMST_2015 | GHS_S1 | GHS-LDMST_2017 |
|---------------------|----------------|--------|----------------|
| Balanced Accuracy   | 0.83           | 0.84   | 0.86           |
| Omission Error      | 0.22           | 0.23   | 0.18           |
| Commission Error    | 0.46           | 0.42   | 0.42           |

Figure 8. Box plots of the performance metrics (balanced accuracy, omission and commission errors) of the three GHSL settlements products calculated per 900 × 900 m cell sizes for 277 cities and disaggregated by continent.
3.5. Regression analysis

A regression analysis between the densities of built-up areas derived from GHSL datasets and the reference built-up surface densities as derived from the reference cartographic data was performed at the 900x900m SU-tile level. This analysis allows establishing a better understanding of the systematic bias and gain parameters included in the GHSL products, namely the GHS_S1, GHS-LDMST_2015 and GHS-LDMST_2017 in relation to reference built-up density and structural characteristics of the different cities.

The results of the regression analysis between the reference densities of built-up areas in the SU-tiles of 900 × 900 m and the corresponding densities obtained from the three GHSL products are shown in Figure 9. This plot matrix shows on the extreme left, the scatterplots displaying the correlations of density measurements from GHSL products against the observed building reference densities. On the diagonal, are shown the distributions of the densities derived from building footprints as well as those obtained from GHS-LDMST_2015, GHS_S1 and GHS-LDMST_2017. On the extreme right column, the frequency plots of classified building densities are represented. The strength of the linear relation between the automatically-generated built-up surface estimation by the different GHSL products and the reference data is assessed through the Pearson correlation coefficient (r) complemented with the Lin’s concordance correlation coefficient (CCC) (Lin, 1989).

While the coefficient of correlation quantifies the share of the real world structural variability explained by the respective products, the concordance correlation coefficient

![Figure 9](image-url)
combines measures of both precision and accuracy to determine how far the built-up products deviate from the line of perfect agreement with the reference built-up densities. The CCC is the product of Pearson correlation coefficient, a measure of precision, and a bias correction factor (BCF) which reflects the degree that the linear association between two variables differs from 45 degrees through the origin.

From the scatterplots on the left side of Figure 9 it can be generally retained that all product layers significantly overestimate the reference building densities. This behaviour is also well described in the frequency plots of classified building densities (extreme right plots of Figure 9) which show increasing overestimation with increasing building densities. This relative overestimation is due to the joint effect of the sensor spatial resolution and to the GHSL-inherent semantic definition of built-up areas that does not comply with individual building footprints. The relative overestimation is more prevalent in the GHS-LDMST_2015 with a slope of 0.63, an intercept of 3.72 and the lowest value of CCC of 0.75 (at a confidence level of 95%). Although all layers show significant overclassification of built-up areas, especially in low-density classes, the correlation coefficients show that the GHS-LDMST_2017 is capable of explaining 89% of the structural variability followed by the GHS-S1 (86%) and GHS-LDMST_2015 (85%). This again demonstrates the positive trend in the refinement of the successive GHSL releases using the technique of the sequential training in the SML framework. The same positive trend can be observed in the increasing CCC and decreasing systematic bias (intercept) as derived by the linear regression analysis, from the 3.72 bias estimated for the GHS-LDMST_2015 to the 2.16 bias as estimated for the GHS-LDMST_2017. Moreover, it is worth noting that the globally systematic gain factor (slope) translating the built-up surface densities as derived from satellite data to built-up surface densities as derived from the reference cartographic data it is almost constant (0.59, 0.63 and 0.64) among the different GHSL products under test. These GHSL products were using radically different satellite sensor technology in input (optical and radar imagery) and different training sets for the statistical learning and classification phase. The robustness of the systematic gain factor versus changes in sensor technology and input training sets is a proof of robustness of the GHSL automatic data processing framework in providing consistent thematic information at global scale.

These results suggest that both the GHS_S1 and to a larger extent the GHS-LDMST_2017 products are suitable for characterizing built-up densities, especially in medium and high-density settlements. The values of slopes and intercepts, of the first order linear regression between the reference densities, and the corresponding densities obtained from the three GHSL products, allow the users to model, retro-fit and compare the results obtained from the different GHSL products. The high correlation coefficients vs. reference cartographic data allow using the GHSL data as estimator of the density of buildings in statistical spatial units of 900 × 900 m size or greater.

4. Usage notes and code availability

The global grids reporting about the presence of built-up areas in the past 40 years are produced in the context of the Global Human Settlement Layer (GHSL) project. They are made for supporting researchers and decision-makers in monitoring the implementation of international frameworks requiring evidence-based, globally-consistent, multitemporal, open and
free information generated from universally-applicable methods. These information grids can be combined with other environmental, physical or socio-economic datasets in order to derive actionable metrics and indicators such as those related to Sustainable Development Goals (e.g. indicator 11.3.1 on Land Use Efficiency (Melchiorri, Pesaresi et al., 2019). Among the different applications that involve built-up gridded data as backbone, population mapping applications generally use information on human settlements either for disaggregating census data to uniform cell grids (Freire et al., 2016), or as covariate in Random Forest modelling approaches (Linard, Tatem, & Gilbert, 2013; Sorichetta et al., 2015).

In the framework of GHSL, the built-up surface grids were combined with multitemporal census data to produce population grids (GHS-POP) depicting globally the population distribution and its density at 250 m and 1 km spatial resolution (Freire et al., 2016). The combined use of built-up and population grids can be beneficial for delineating city boundaries globally following a consistent approach. This was the purpose behind the Global Human Settlement Model (GHS-SM) which uses the GHSL baseline data for porting the Degree of Urbanisation classification (Dijkstra & Poelman, 2014) and hence deriving a globally consistent delineation of city boundaries at a spatial resolution of 1 km.

Several other applications benefit directly or indirectly from the multitemporal built-up grids. The assessment of indicators in the context of the sustainable development goals and the analysis of changes of exposure in support to the Sendai Framework for Disaster Risk Reduction, represent the most exhaustive applications requiring uniform baseline data on human settlements. Several examples of such indicators and their usage in policy guidance are available in the Atlases of the Human Planet released in 2016, 2017 and 2018 (Pesaresi et al., 2016, 2017; European Commission, Joint Research Centre, 2018b).

Limitations of the built-up grids include the potential of omission of settlements due to unavailability of suitable satellite data (i.e. data gaps in particular in 1975 and 1990). The binary built-up, non-built-up masks are a simplification of the reality and cannot effectively address the problem of mixed pixels problems of Landsat images. This explains the inclusion of other artificial surfaces such as parking lots, pavements, roads and man-made impenetrable cover types in the built-up mask in some high built-up density areas.

The general workflow for built-up areas extraction from remote sensing data is available in the MASADA tool (MASADA stands for Massive Spatial Automatic Data Analytics) (Politis, Corbane, & Pesaresi, 2017) that can be downloaded from the GHSL website (https://ghsl.jrc.ec.europa.eu/). The tool builds on the Symbolic Machine Learning (SML) classifier. The image classification workflow incorporates radiometric, textural and morphological features as inputs for information extraction. The tool supports several types of multispectral optical imagery. It includes ready-to-use workflows for specific sensors, and allows the parametrization and customization of the workflow by the user. Version 1.3 includes predefined workflows for SPOT-5, SPOT-6/7, RapidEye and CBERS-4, but it was also tested with various high and very high resolution sensors like GeoEye-1, WorldView-2/3, Pléiades and Quickbird. A new version of MASADA (v.2) supporting Sentinel-1, Sentinel-2 has been recently developed.

**Acknowledgments**

The authors wish to acknowledge all the team members of both projects, the Joint Research Centre Global Human Settlement Layer (GHSL) and the JRC Big Data Analytics project for their contribution to the production of the global datasets.
Data Availability Statement

All output datasets described in this article are publicly and freely available through JRC Open Data Catalogue. They represent the datasets produced at the time of writing, and will be preserved in their published form. A visualization of the layers is given online at https://ghsl.jrc.ec.europa.eu and https://cidportal.jrc.ec.europa.eu/services/webview/jeodpp/databrowser/

[dataset 1] GHS built-up grid, derived from Landsat, multitemporal (1975-1990-2000-2014), R2015B. Joint Research Centre Data Catalogue, PID: http://data.europa.eu/89h/jrc-ghsl-ghs_built_ldsmt_globe_r2015b (2015), reference number (Pesaresi et al., 2015).

[dataset 2] GHS built-up grid, derived from Sentinel-1 (2016), R2018A. Joint Research Centre Data Catalogue, doi:10.2905/jrc-ghsl-10,007, PID: http://data.europa.eu/89h/jrc-ghsl-10007 (2018), reference number (Corbane et al., 2018).

[dataset 3] GHS built-up grid, derived from Landsat, multitemporal (1975-1990-2000-2014), R2018A. Joint Research Centre Data Catalogue, doi:10.2905/jrc-ghsl-10,008, PID: http://data.europa.eu/89h/jrc-ghsl-10008 (2018), reference number (Corbane, Florczyk, Pesaresi, Politis, & Syrris, 2018).

[dataset 4] GHS built-up grid input data, Landsat multitemporal collections (1975-1990-2000-2014), R2018A. Joint Research Centre Data Catalogue, PID: http://data.europa.eu/89h/jrc-ghsl-10009 (2018), reference number (Florczyk et al., 2018).

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Arino, O., Perez, J. J. R., Kalogirou, V., Bontemps, S., Defourny, P., & Van Bogaert, E. (2012). Global Land Cover Map for 2009 (GlobCover 2009). PANGAEA - Data Publisher for Earth & Environmental Science. doi:10.1594/PANGAEA.787668

Bhaduri, B., Bright, E., Coleman, P., & Dobson, J. (2002). LandScan: Locating people is what matters. Geoinformatics, 5(2), 34–37.

Bontemps, S., Defourny, P., Radoux, J., Van Bogaert, E., Lamarche, C., Achard, F., … Arino O. (2013). Consistent Global Land Cover maps for climate modelling communities: current achievements of the ESA’ Land Cover CCI. In ESA Living Planet Symposium, (Vol. 722, pp. 62). Noordwijk, NL: European Space Agency. http://adsabs.harvard.edu/abs/2013ESASP.722E..62B

Brodersen, K. H., Soon, O. C., Stephan, K. E., & Buhmann, J. M. (2010). The balanced accuracy and its posterior distribution. In 2010 20th International conference on pattern recognition (pp. 3121–3124). doi:10.1109/ICPR.2010.764.

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., … Mills, J. (2015). Global Land Cover Mapping at 30 m resolution: A POK-based operational approach. ISPRS Journal of Photogrammetry and Remote Sensing, Global Land Cover Mapping and Monitoring, 103(May), 7–27.

Chen, J., Yifang, B., & Songnian, L. (2014). Open access to earth Land-Cover Map: China. Nature, 514 (7523), 434.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment, 37(1), 35–46.

Corbane, C., Politis, P., Pesaresi, M., Kemper, T., & Siragusa, A. (2018). Estimation of land use efficiency from the Global Human Settlement Layer (GHSL). In N. Baghdadi, C. Mallet, & M. Zribi (Eds.), QGIS and applications in territorial planning (pp. 39–52). Hoboken, NJ, USA: John Wiley & Sons, Inc. doi:10.1002/9781119457121.ch2

Corbane, C., Florczyk, A., Pesaresi, M., Politis, P., & Syrris, V. (2018, July). GHS built-up grid, derived from landsat, multitemporal (1975-1990-2000-2014), R2018A. doi:10.2905/jrc-ghsl-10007
Corbane, C., Lemoine, G., Pesaresi, M., Kemper, T., Sabo, F., Ferri, S., & Syrris, V. (2018). Enhanced automatic detection of human settlements using sentinel-1 interferometric coherence. *International Journal of Remote Sensing*, 39(3), 842–853.

Corbane, C., Pesaresi, M., Politis, P., Syrris, V., Aneta, J., Florczyk, P. S., … Kemper, T. (2017, November). Big earth data analytics on sentinel-1 and landsat imagery in support to global human settlements mapping. *Big Earth Data*, 1–27. doi:10.1080/20964471.2017.1397899

Corbane, C., Pesaresi, M., Politis, P., Syrris, V., Florczyk, A. J., Soille, P., Maffenini, L., … Kemper, T. (2017). Mass processing of sentinel-1 and landsat data for mapping human settlements at global level. In S. Pierre & M. P. Giorgio (Eds.), *Proceeding of the BiDS’17* (pp. 52–55). Luxembourg: Publications Office of the European Union. doi:10.2760/383579

Corbane, C., Politis, P., Syrris, V., & Pesaresi, M. (2018). GHS built-up grid, derived from Sentinel-1 (2016), R2018A. doi:10.2905/jrc-ghsl-10008

Di Giorgio, A., & Latham, J. (2009). Land use, land cover and soil sciences – Vol. I - AFRICOVER land cover classification and mapping project. In W. H. Verheye (Ed.), *Land use, land cover and soil sciences* (pp. 236–251). Belgium: UNESCO-EOLSS. National Science Foundation Flanders-Belgium and Geography Department, University of Gent.

Dijkstra, L., & Poelman, H. (2014). *A harmonised definition of cities and rural areas: The new degree of urbanisation* Working Papers. Regional Working Paper 2014. Retrieved from http://ec.europa.eu/regional_policy/en/information/publications/working-papers/2014/a-harmonised-definition-of-cities-and-rural-areas-the-new-degree-of-urbanisation.

Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C., & Worley, B. A. (2000). LandScan: A global population database for estimating populations at risk. *Photogrammetric Engineering and Remote Sensing*, 66(7), 849–857.

Doldirina, C. (2015). Open data and earth observations the case of opening up access to and use of earth observation data through the global earth observation system of systems. *Jipitec*, 6(1), 73–85.

Esch, T., Heldens, W., Hirner, A., Keil, M., Marconcini, M., Roth, A., … Strano, E. (2017). Breaking new ground in mapping human settlements from space – the global urban footprint. *ISPRS Journal of Photogrammetry and Remote Sensing*, 134(December), 30–42.

European Commission, Joint Research Centre. (2018a). *Atlas of the human planet 2018, A world of cities*. Luxembourg: Publications Office of the European Union. doi:10.2760/124503

European Commission, Joint Research Centre. (2018b, July 18). “JRC data catalogue.” *Global Human Settlement Layer Collection*. Retrieved from https://data.jrc.ec.europa.eu/.

European Environment Agency (EEA). (2007). *CLC2006 technical guidelines*. EEA Technical report No 17/2007. EEA.

Florczyk, A., Politis, P., Corbane, C., & Pesaresi, M. (2018, October). GHS built-up grid input data, landsat multitemporal collections (1975-1990-2000-2014), R2018A. Retrieved from http://data.europa.eu/89h/jrc-ghsl-10009.

Freire, S., Doxsey-Whitfield, E., MacManus, K., Mills, J., & Pesaresi, M. (2016). Development of new open and free multi-temporal global population grids at 250 m resolution. In T. Sarjakoski, M. Y. Santos & L. T. Sarjakoski (Eds.), *Proceeding of the 19th AGILE Conference on geographic information science*, Vol. 250. Helsinki, Finland.

Freire, S., Santos, T., Navarro, A., Soares, F., Silva, J. D., Afonso, N., … Tenedório, J. (2014). Introducing mapping standards in the quality assessment of buildings extracted from very high resolution satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 90 (April), 1–9.

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., … Chen, J. (2013). Finer resolution observation and monitoring of global land cover: First mapping results with landsat TM and ETM+ Data. *International Journal of Remote Sensing*, 34(7), 2607–2654.

Gutman, G., Huang, C., Chander, G., Noojipady, P., & Masek, J. (2013). Assessment of the NASA-USGS Global Land Survey (GLS) Datasets. *Remote Sensing of Environment*, 134, 249–265.

Lin, L. I.-K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45 (1), 255.
Linard, C., Tatem, A. A. J., & Gilbert, M. (2013). Modelling spatial patterns of urban growth in Africa. Applied Geography, 44(October), 23–32.

Liu, X., Guohua, H., Chen, Y., Xia, L., Xiaocong, X., Shaoying, L., … Wang, S. (2018). High-resolution multi-temporal mapping of global urban land using landsat images based on the google earth engine platform. Remote Sensing of Environment, 209(May), 227–239.

Melchiorri, M., Florczyk, A., Freire, S., Schiavina, M., Pesaresi, M., & Kemper, T. (2018). Unveiling 25 years of planetary urbanization with remote sensing: Perspectives from the global human settlement layer. Remote Sensing, 10(5), 768.

Melchiorri, M., Pesaresi, M., Aneta, J., Florczyk, C. C., & Kemper, T. (2019). Principles and applications of the global human settlement layer as baseline for the land use efficiency indicator –SDG 11.3.1. ISPRS International Journal of Geo-Information, 8(2), 96. doi:10.3390/ijgi8020096.

OpenStreetMap contributors. (2017). Planet Dump Retrieved from https://Planet.Osm.Org.

Otsu, N. (1979). A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics, 9(1), 62–66.

Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. Nature, 540(7633), 418–422.

Pesaresi, M., Ehrlich, D., Ferri, S., Florczyk, A., Freire Sergio, M. C., Halkia, S., … Syrris, V. (2016). Operating procedure for the production of the global human settlement layer from landsat data of the Epochs 1975, 1990, 2000, and 2014. Luxembourg: Publications Office of the European Union. Retrieved from http://publications.jrc.ec.europa.eu/repository/handle/111111111/40182

Pesaresi, M., Ehrlich, D., Florczyk, A., Freire, S., Julea, A., Kemper, T., … Syrris, V. (2015 January). GHS built-up grid, derived from landsat, multitemporal (1975, 1990, 2000, 2014). Retrieved from http://data.europea.eu/89h/jrc-ghsl-ghs_built_ldsmt_globe_r2015b.

Pesaresi, M., Ehrlich, D., Kemper, T., Siragusa, A., Florczyk, A. J., Freire, S., & Corbane, C. (2017). Atlas of the human planet 2017: Global exposure to natural hazards. EUR 28556 EN. Luxembourg: European Commission, Joint Research Centre, Institute for the Protection and Security of the Citizen. Retrieved from https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/atlas-human-planet-2017-global-exposure-natural-hazards.

Pesaresi, M., Gerhardinger, A., & Kayitakire, F. (2008). 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, 1(3), 180–192.

Pesaresi, M., Huadong, G., Blaes, X., Ehrlich, D., Ferri, S., Gueguen, L., … Zanchetta, L. (2013). 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, 6(5), 2102–2131.

Pesaresi, M., Melchiorri, M., Siragusa, A., & Kemper, T. (2016). Atlas of the human planet - mapping human presence on earth with the global human settlement layer. JRC103150. Publications Office of the European Union. Luxembourg (Luxembourg): European Commission, DG JRC.

Pesaresi, M., Syrris, V., & Julea, A. (2015). Benchmarking of the symbolic machine learning classifier with state of the art image classification methods - application to remote sensing imagery, EUR 27518. JRC Technical Report EUR 27518. JRC Technical Report. Publications Office of the European Union. doi:10.2788/638672.

Pesaresi, M., Syrris, V., & Julea, A. (2016). A new method for earth observation data analytics based on symbolic machine learning. Remote Sensing, 8(5), 399.

Politis, P., Corbane, C., & Pesaresi, M. (2017). MASADA User Guide. EUR 28609 EN. Luxembourg: Publications Office of the European Union. http://ghsl.jrc.ec.europa.eu/documents/MASADA_1.3.zip.

Sabo, F., Christina Corbane, A. J., Florczyk, S. F., Pesaresi, M., & Kemper, T. (2018). Comparison of built-up area maps produced within the global human settlement framework. Transactions in GIS, 22(6), 1406–1436.

Schneider, A., Friedl, M. A., & Potere, D. (2009). A new map of global urban extent from MODIS satellite data. Environmental Research Letters, 4.

Soille, P. (2011). IMAGE-2006 Mosaic: SPOT_HRVIR and IRS-LISSIII cloud detection. EUR 23636 EN. Luxembourg: Publications Office of the European Union: European Commissio Joint Research Centre. doi:10.2788/49355
Soille, P., Burger, A., De Marchi, D., Kempeneers, P., Rodriguez, D., Syrris, V., & Vasilev, V. (2017). A versatile data-intensive computing platform for information retrieval from big geospatial data. *Future General Computing System*. doi:10.1016/j.future.2017.11.007

Sorichetta, A., Hornby, G. M., Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). High-resolution gridded population datasets for Latin America and the Caribbean in 2010, 2015, and 2020. *Scientific Data, 2*(September), 150045.

Stone, K. H. (1965). The Development of a Focus for the Geography of Settlement. *Economic Geography, 41*(4), 346–355.

Syrris, V., Corbane, C., & Soille, P. (2017). A global mosaic from Copernicus Sentinel-1 data1. In S. Pierre & M. P. Giorgio (Eds.), *Proceeding of the BiDS’17* (pp. 268–271). Luxembourg: Publications Office of the European Union.

United States Geological Survey. (2013, January 12). USGS landsat mission. Retrieved from https://landsat.usgs.gov/.

Zhou, Y., Smith, S. J., Zhao, K., Imhoff, M., Thomson, A., Bond-Lamberty, B., … Elvidge, C. D. (2015). A global map of urban extent from nightlights. *Environmental Research Letters, 10*(5), 054011.

Zhu, Z., Wang, S., & Woodcock, C. E. (2015). Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for landsats 4–7, 8, and Sentinel 2 images. *Remote Sensing of Environment, 159*(March), 269–277.