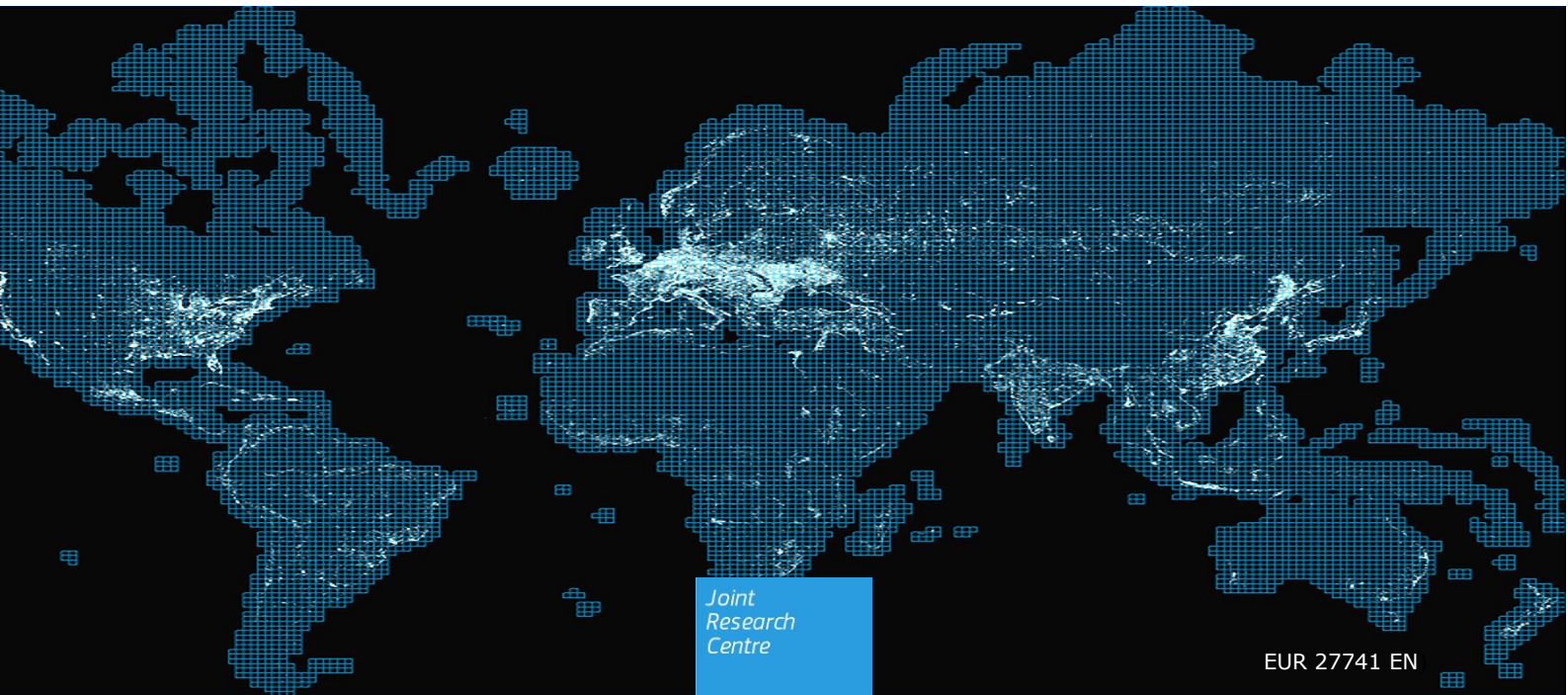
European Commission

JRC TECHNICAL REPORTS

Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014

Martino Pesaresi, Daniele Ehrlich, Stefano Ferri, Aneta J. Florczyk, Sergio Freire, Matina Halkia, Andreea Julea, Thomas Kemper, Pierre Soille, and Vasileios Syrris

2016



Joint Research Centre

EUR 27741 EN

Contact information
Name: Martino Pesaresi
Address: Joint Research Centre, Via E.Fermi 2749, TP267, I-21027 Ispra (VA), Italy
E-mail: martino.pesaresi@jrc.ec.europa.eu
Tel.: +39 0332 789524

JRC Science Hub
https://ec.europa.eu/jrc

JRC97705

EUR 27741 EN

ISBN 978-92-79-55012-6 (PDF)
ISBN 978-92-79-55013-3 (print)

ISSN 1831-9424 (online)
ISSN 1018-5593 (print)

doi:10.2788/253582 (online)
doi:10.2788/656115 (print)

How to cite: Martino Pesaresi, Daniele Ehrlich, Stefano Ferri, Aneta J. Florczyk, Sergio Freire, Matina Halkia, Andreea Julea, Thomas Kemper, Pierre Soille, and Vasileios Syrris; Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014; JRC Technical Report EUR 27741 EN; doi:10.2788/253582 (online)

Table of contents

Acknowledgements

We would like to thank Lewis Dijkstra, Ellen Hamilton, and Eduardo Moreno for the strategic support provided to the GHSL vision since the very early stages of the project. We are grateful to Shlomo Angel and Eugenie Birch for the numerous public endorsements of the early GHSL results and the encouragements provided during the long gestation of the Landsat processing methodology. Moreover, a valuable help in testing the GHSL results was provided by Shlomo Angel and his team by sharing their data of the Urbanization Project. Andrew J. Tatem and the WorldPop team were kindly sharing settlement geo-spatial data that was greatly contributing to the improvement of the training set used in the GHSL machine learning process. Finally, we warmly thank the partners of the GEO GHS working group who made a wonderful job in testing the intermediate results of the GHSL in real-world application scenarios. Their feedback and suggestions have been fundamental for the improvement of the method and of the results discussed here.

Abstract

A new global information baseline describing the spatial evolution of the human settlements in the past 40 years is presented. It is the most spatially global detailed data available today dedicated to human settlements, and it shows the greatest temporal depth. The core processing methodology relies on a new supervised classification paradigm based on symbolic machine learning. The information is extracted from Landsat image records organized in four collections corresponding to the epochs 1975, 1990, 2000, and 2014. The experiment reported here is the first known attempt to exploit global Multispectral Scanner data for historical land cover assessment. As primary goal, the Landsat-made Global Human Settlement Layer (GHSL) reports about the presence of built-up areas in the different epochs at the spatial resolution allowed by the Landsat sensor. Preliminary tests confirm that the quality of the information on built-up areas delivered by GHSL is better than other available global information layers extracted by automatic processing from Earth Observation data. An experimental multiple-class land-cover product is also produced from the epoch 2014 collection using low-resolution space-derived products as training set. The classification schema of the settlement distinguishes built-up areas based on vegetation contents and volume of buildings, the latter estimated from integration of SRTM and ASTER-GDEM data. On the overall, the experiment demonstrated a step forward in production of land cover information from global fine-scale satellite data using automatic and reproducible methodology.

1. Introduction

Human settlement information is used to develop indicators for modelling the *access* (to services, market, industrial infrastructure, food, water, land), the *exposure* (to natural / human hazards, disasters, pollution), and the *impact* of human footprint (to land, water, ecosystem). In fact, global human settlement information are in demand by a number of institutions operating globally that include the European Commission Services, the United Nations agencies and programs, the World Bank, as well as the donor countries that require quantitative variables to prioritize their humanitarian and development aid or their financial investments. Human settlements information will also be required for developing indicators for the four post-2015 international frameworks, namely Sendai framework for Disaster Risk Reduction (DRR), Sustainable Development Goals (SDG) with particular focus on Goal 11 (make cities and human settlements inclusive, safe, resilient, sustainable), Climate Change and the Global Urban Agenda (adopted in 2016). In fact, to monitor the implementation of the SDGs, it will be important to improve the availability and access to data and statistics to ensure that no one is left behind in the information gaps[1].

Human settlements can be mapped with Remote Sensing (RS) data which are independent, globally-consistent, updated regularly, providing a synoptic overview and can be considered objective. These data can be used to derive indicators that are action oriented, global in nature and universally applicable. Regular provision of remote sensing data may be one of the few ways to gather standardized information globally.

From the above perspective, open access to fine-scale global information is an important aspect to be taken in to consideration, together with the issues related to the scientific control and reproducibility of the information results as they are improving the objectivity of the derived assessments. The reproducibility of the information extraction model is linked to the level of automatic control of the information production workflow, the sustainability of the information gathering and the scalability to large continuous data streams as the ones produced by the Landsat and Sentinel satellite platforms.

This manuscript reports on processing 40 years of Landsat imagery for mapping the global built-up areas over this period. This work is conducted within the Global Human Settlement Layer (GHSL) project funded by the European Commission, Joint Research Centre (JRC). The aim of the GHSL is to provide scientific methods and systems for reliable and automatic built-up information gathering. The GHSL project contributes to the Group of Earth Observation (GEO) societal benefit SB-04-C1: *Global Urban Observation and Information*[2].

[1] Open Working Group proposal for Sustainable Development Goals `https://sustainabledevelopment.un.org/focussdgs.html`

[2] `http://www.earthobservations.org/ts.php?id=158`

In October 2014 the first (alpha) release of the global data was processed by JRC and shared inside a newly set group of experts (the GEO Global Human Settlement working group[3]) for early testing purposes. To date, the applications under test include impact assessment, disaster early warning and alerting, losses estimates, exposure and risk mapping and post-disaster need assessment (PDNA), spatial population modeling, census planning, urban and regional development, urban and global climate modeling, spatial epidemics analysis, ecological studies, environmental protection, agricultural fragmentation studies, and historical landscape protection. Their geographical scope may include national, regional/continental, and global coverages as well.

The experiment reported here aimed to answer to the following questions:

1. is it possible to increase the (spatial, thematic, temporal) detail of the available global information layers describing human settlements and using free and open satellite data in input?

2. is it possible to increase the level of automatic control and reproducibility of the information production workflow?

3. what are the minimal computational requirements needed for performing the tasks?

This manuscript will focus on answering the first question addressing the application domain. The second and third questions are related to the classification methodology and the computational issues. They will be addressed only partially in this manuscript by showing the specific solutions implemented in the Landsat data classification scenario. They are more developed in [43], where the core data classification strategy applied in this experiment is compared with other standard approaches for image data classification in remote sensing.

The report is structured as follows. The methodological context and the past experiences are summarized in the background section. Then, the general methodology, data sets and theoretical framework of our approach are presented in section 3. Section 4 details the basic recognized information components. The specific processing steps undertaken for Landsat 8 data are reported in section 5 and the description of the information fusion processes applied is included in section 6. Results are then discussed in section 7 and, finally, conclusions are drawn in section 8.

[3] http://www.earthobservations.org/ghs.php

2. Background

This section offers information about the methodological basis, the precedent experience and the innovative contribution of our approach.

2.1. Methodological issues

In GHSL, the *built-up area* class is defined as *the union of all the spatial units collected by the specific sensor and containing a building or part of it* as stated in [41]. Human settlements are made by population and physical infrastructures. The building is one essential part of the settlement infrastructure and is the basic sign of the human presence that can be physically observable by remote sensing technologies. The concept of Human Settlement adopted in this study relies on the classical notion inherited from the settlement geography, defined as *"...the description and analysis of the distribution of buildings by which people attach themselves to the land"* by [52].

Urban and more generally settlement areas observed by satellite sensors are characterized by large spectral mixture of different objects/entities. On average, the characteristic scale of the entities composing settlements is of 10 meters as stated in [50, 51]. Temporary settlements as refugees and internally displaced people (IDP) camps, camping areas, slums and some traditional rural settlements may be much smaller (up to 2 meters). On the other hand, large factories and commercial estates in contemporary settlements may exceed 100 meters.

The observed spectral variability is mainly produced by i) the relation between sensor spatial resolution and the characteristic scale of the settlement entities, ii) the presence of different materials in the settlement areas, their variable local spatial arrangements and mixture, and iii) the variable illumination conditions coupled with different building sizes causing variable shadowing largely affecting the observed radiometric behavior.

In [41], it is demonstrated the possibility to solve fine-scale recognition of human settlements by exploiting the fact that they systematically produce locally-heterogeneous spectral reflectance in the recorded image data. Local signal heterogeneity is formalized through a selected set of textural features [36, 40] and multi-scale morphological image descriptors [38, 42], both of them based on local contrast measures. In particular, the local density of highly contrasted square corners was found strongly correlated to the presence of human settlements [22]. These methods build upon early work of the '90s in the remote sensing community by [19, 20] exploring the use of structural image information for automatic discrimination of urban areas.

These textural and morphological image features demonstrated the capacity to discriminate built-up areas with optical sensors in the spatial resolution range of 0.5m-10m. The upper bound of the resolution range (10 meters) being consistent with the scale of the urban reflectance as estimated using spatial auto-correlation techniques in

[50]. They were successfully tested using 2.5m input sensor resolution for producing large national coverages in Brazil by [24] and in China by [28] using CBERS-2B panchromatic data, as well as in South Africa by [25] with SPOT-5 data. [15] and [16] used Spot-5,6 multispectral pan-sharpened data for producing a continental coverage in Europe.

In the experiment reported here, we test the possibility to extend these methods to global open Landsat satellite data that are beyond security and commercial restrictions. Free and open data access policy is facilitating the free sharing of the analysis results in the global scientific and user communities, thus augmenting their societal impact and benefit as stated by [55]. The Landsat free and open access to data policy, begun in 2008, offered the possibility to exploit the longest continuous global record of the Earth's surface dating back to '70.

2.2. Past experiences

Global human settlements have mostly been mapped from low to moderate resolution (300m - 1000m spatial resolution), as reported in [44] and [48]. Low and moderate resolution imagery has also been used to detect changes. For example, DMSP/OLS night time lights and SPOT-VGT data were used to detect changes between 1998 and 2008 in India by [34]. MODIS 500m resolution images were used by [32] to map urban areas in East Asia from 2000 to 2010.

Finer scale imagery has also been extensively used to map the built environment. Angel and his team [3] mapped 120 cities over 1990 and 2000. Taubenböck and his colleagues [53] conducted a systematic analysis of 27 current mega cities using multi-temporal Landsat data from 1975, 1990, 2000 and TerraSar-X data from 2010. The Terrasar-X was also used in the TanDEM-X mission to generate a Global Urban Footprint (GUF) for the years 2011-2013 by [29].

Global finer scale built-up areas mapping from Landsat was delivered by the Monitoring of Global Land Cover (FROM-GLC) project for the year 2006 as reported in [21]. In FROM-GLC, only one epoch (circa 2006) was processed, and the impervious surfaces resulted with not satisfactory classification accuracy as presented in [4] and [21]. Successive experimental activity tried to inject in FROM-GLC output the urban or impervious information derived from third-parts, low-resolution satellite-derived information sources as in [56]. Finally, a 30m resolution global land cover (GlobeLand30) was produced as reported in [8]. GlobeLand30 processes two global Landsat data collections (years 2000 and 2010), it integrates in the output several internationally available land cover products, and relies on large use of manual editing of the final information done by domain experts.

A global urban area map was also produced by [33] using an automated classification method based on ASTER data. The approach combine Learning with Local and Global Consistency (LLGC) and logistic regression with urban maps.

2.3. Innovation

Landsat historical data records range from 75m to 15m spatial resolution in multispectral and panchromatic imaging mode: consequently, textural and morphological image features used in [41] lose most of their capacity to discriminate built-up areas and must be complemented with multispectral image information.

The new method generalizes the single-variable single-training-set optimization techniques proposed by [41] in the machine learning phase, to the scenario where the combination of multiple variables in input are taken into consideration with a combination of multiple training set collections. The generalization to multiple variables allows exploiting the multi-spectral information contents of the input satellite data instances, while the generalization to multiple training sets allows the improving of broad-scale learning sets by adding fine-scale open information. This includes information collected from detailed but usually incomplete and heterogeneous geo-spatial sources describing settlements globally.

The volume, variety and partially unstructured nature of the data instances used in the experiment can be associated with the characteristics of Big Data Analytics as stated by [12], [31] and [37]. Today's state-of-the art machine learning techniques used in remote sensing applications show severe drawbacks if used in scenarios where large training sets are used together with large input data. Those algorithms have been designed and tested on small to medium data instances (if compared with fine-scale global RS data scenarios) with a moderate to large number of attributes or features and are difficult to scale to geo-spatial Big Data machine learning problems according to [54], [37]. Those drawbacks are amplified if the training set and input image data show inconsistencies, inaccuracies and calibration issues. The GHSL is in fact exploiting open source geospatial information as training set and multiple-sensor and historical satellite data series embedding large variability in metadata documentation. This is being addressed by a new general paradigm for remote sensing data classification introduced by [43], based on image data *sequencing* and *Symbolic Machine Learning* (SML) by association analysis (AA) techniques. Association Rule Analysis and Associative Classification are some of the most popular paradigms in mature data mining and knowledge discovery application areas, requiring the analysis of large databases. In particular, the Association Rules in genes (or Genetic Association), is extensively used in bio-informatics for revealing biologically relevant associations between different genes or between environmental effects and gene expression as in [5], [10], [18] and [1]. By analogy, the associative classifier (AC) proposed in this experiment searches for relevant, systematic relations between image data instances and spatial information encoded in the selected training sets.

In the proposed solution, five main new products are generated at the maximum spatial resolution allowed by the Landsat: i) the estimation of the global built-up area in 1975, 1990, 2000, and 2014,

ii) the estimation of the degree of built-up by sub-pixel information mixing, and iii) an experimental per-pixel multiple-class land cover classification.

3. Methodology

This section includes the description of the general data processing workflow, the input image data, the training and testing sets used in the experiment. The proposed method is based on a combination of data-driven and knowledge-driven reasoning by implementing supervised and unsupervised data classification processes. The supervised classification is done independently at the scene level, for all the scenes processed, using the same global training set. The unsupervised classification chains are implemented with the same criteria and the same parameters to the whole data set. During the processing of the satellite scenes, no manual tuning of model or parameters is applied per scene, per specific collection or per geographical area. The whole parameter sets are either decided at the beginning of the process, or optimized during the learning and classification phase. As a consequence, the whole process is fully transparent and reproducible with very limited manual intervention. Moreover, the drastic reduction of the free parameters to be tuned makes the method relatively easy to scale to large volumes of data processing by replication of parallel processes.

3.1. Workflow

As in [41], the general data processing workflow is based on supervised classification inserted in an evolutionary schema with a retro-action mechanism. Specifically, two main processing loops (α, β) are implemented for all the scenes of all the epochs, with the purpose of stabilization of the classification output (see Figure 1). In the first loop α, the learning set consists of coarse-scale global data (low-resolution land cover) available before the experiment and a multi-temporal $GHSL_{\alpha}$ mosaic at full resolution is produced. In the second loop β, the previous global learning set is augmented by the multi-temporal $GHSL_{\alpha}$ mosaic (the retro-action feed) and additional fine-scale data collected i) from unsupervised classification of the panchromatic channel of the Landsat 8 (L8) data, ii) from open source data such as Open Street Maps (OSM), or iii) made available by GHSL partners. The second loop ends with the $GHSL_{\beta}$ global mosaic that is discussed here. The α loop is a first approximation of the solution that is considered conservative and under-representing small and scattered settlement patterns given the scale of the used training set. The β loop is dedicated to the refinement and enlargement of the stable background information discovered in the $GHSL_{\alpha}$ set by risking new input fine-scale positive samples. Because some fine-scale training set sources (e.g. OSM) are arbitrarily available in some locations with unknown completeness and update characteristics, to rely only on these sources would introduce instability in the overall system reliability performances.
During the experimental design phase, textural, morphological and radio-metrical image features were tested with regard to the capacity to discriminate built-up areas with the Landsat input data char-

acteristics. The contribution of the textural and morphological image features was assessed as marginal if compared to the radiometric features, while they provide the major computational cost of the image feature extraction phase. For this reason, the textural and morphological image features play only an ancillary role. The input image data to the classification process includes most exclusively multispectral data provided by the different sensors in the four epochs considered. The panchromatic channel has been exploited exclusively for the improvement of the built-up area training sets through textural analysis when available.

The whole information production workflow (see Figure 1) is based on parallel single-scene learning and classification processes, without band combinations or spectral special feature extraction. Therefore, a direct re-quantization of the digital numbers (DN) recorded in the processed scenes was used in input for the symbolic machine learning phase. This choice dramatically contributes to decrease the computational complexity of the data preparation phase and it is made possible by the fact that during the experiment each satellite scene is classified individually and consequently no classification model transferability is required.

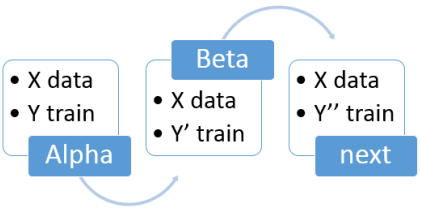


Figure 1: The general workflow applied in the experiment

At each process loop for each scene, three basic information (classes) are estimated with a fixed sequence: i) clouds and data mask, ii) surface water mask (and consequently land mask), and iii) built-up confidence. Clouds and valid data masks are extracted by deductive reasoning, while water and built-up areas are detected using symbolic machine learning and supervised classification. Clouds and water masks are Boolean outputs, while built-up is a continuous output in a standardized range. They are described in the next paragraphs. Only for the L8 collection of the epoch 2014-2015 a set of additional information layers was extracted. They include experimental land cover products made by integration and generalization of the extracted information. Only in the L8 collection, also textural image features were used for enhancing the classification of built-up areas. At the end of the per-scene processing phase, a global mosaic of the basic information contents was created for each of the 4 epochs, by fusing the information discovered in the same spatial unit by different

scenes of the same epoch. During this process, data gaps (as produced by clouds) are minimized, and redundant information was exploited for reduction of the overall error rate. After this phase of information aggregation by epoch, a second phase was implemented for the multi-temporal assessment. This phase includes the automatic mitigation of the systematic gain-bias effects that may result in the confidence calculated from different sensors and different geographical locations and consequently including different building practices, settlement patterns, and natural background characteristics. In this last phase, an expert-driven check was integrated in order to reduce the probability that macroscopic errors in the final classified product may be found.

3.2. Input Image Data

This report relies on the data of the Landsat programme, which is available since 1972 from six satellites and distributed with free and open access[4]. The satellites have been equipped with three primary sensors evolving over thirty years: MSS (Multispectral Scanner), TM (Thematic Mapper), and ETM+ (Enhanced Thematic Mapper Plus). Landsat supplies visible and infrared imagery, with thermal imagery and a panchromatic image also available from the ETM+ sensor. Details regarding the sensor characteristics are available from [30]. The RS data used in this experiment consists of 32808 scenes organized in four collections corresponding to the epochs 1975, 1990, 2000, and 2014, including 7588, 7375, 8756, and 9089 satellite scenes, respectively. The collections of the epochs 1975, 1990, and 2000 are made by the Global Land Survey (GLS) data pre-processed by the Maryland University and available for public use since 2008[5] and described in [23]. The collection of the epoch 2014 was made by a direct download of Landsat 8 scenes from the USGS website[6]. The main known issues related to the used GLS collections are (i) the incomplete metadata information for 16.6% and 32.8% of scenes of the 1975 and 1990 epochs, respectively, which does not allow to estimate the top-of-atmosphere (TOA) reflectance parameters, and (ii) large data gaps in the 1975 epoch. In particular, the northern part of South America, the whole Greenland, and large parts of Siberia are missing in the 1975 collection and consequently are not processed in this experiment. Figure 2 shows the amount of satellite data records available in the different epochs during the experiment.

3.3. Training Set

The proposed processing workflow makes large use of already available spatial information describing human settlements and that is available from various sources at different scales, thematic definition and completeness or accuracy conditions. This information is used

[4] http://landsat.usgs.gov/about_mission_history.php
[5] http://glcf.umd.edu/data/gls/
[6] http://landsat.usgs.gov/Landsat_Search_and_Download.php

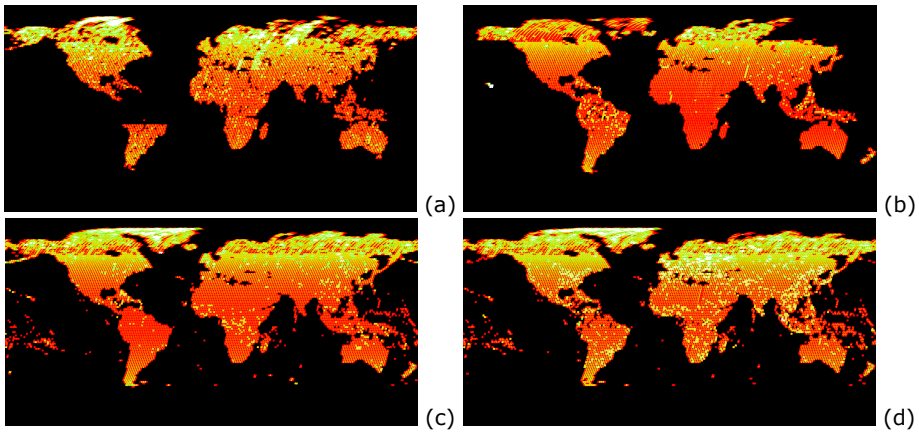


Figure 2: Global data availability by collections processed in the experiment. (a) 1975, (b) 1990, (c) 2000, and (d) 2014. Black = no data available, red = only one data record, yellow = two data records, green = three data records, white = four and more data records available.

as training set for the supervised classification process. The data are collected by open collaborative projects or, more structured, by academic or governmental organizations. The *fine-scale training sets* consists of information extracted from available sources with comparable or better scale with respect to the image data. They included i) Y_{osm} - the *Open Street Map* (OSM) data reporting about roads - $Y_{osm_{ws}}$, settlement places - $Y_{osm_{b1}}$, and urban cover - $Y_{osm_{b2}}$ extracted from the Geofabrik[7] source; ii) Y_{gnm} - the *Settlement location points,* extracted from the Geonames[8] source; and iii) Y_{nlc} - the settlement polygons extracted from fine-scale land cover information available at the moment of the experiment. In the Y_{nlc}, a set of settlement layers kindly made available by the WorldPop[9] project was also integrated. The *broad-scale training sets* were made by i) the MODIS global urban extents - Y_{mds} from [49, 48]; ii) the Meris Globcover - Y_{glc} from [6], and iii) the population density grid Landscan - Y_{lsc} from [13] (version for the year 2013 available from the website[10]).

The low-resolution training sets are more generalized than the input data under processing in this experiment. Especially if embedding categorical information such as Y_{mds} and Y_{glc}, they are expected to exhibit a systematic information bias that is estimated to over and under represent large/compact vs. small/dispersed settlements, correspondingly. On the other hand, they are expected to be consistent, complete, and with a known time attribute across the global spatial domain. The above holds with the exception of the Y_{lsc} source where

[7] http://www.geofabrik.de/
[8] http://www.geonames.org/
[9] http://www.worldpop.org.uk/
[10] http://web.ornl.gov/sci/landscan/

no information is available about the update, completeness and scale of the settlement spatial information embedded in the model. By definition, the fine-scale training sets are available at the same or finer scale than the input image data under processing: but the consistency, completeness and time attribute of these sources are typically unknown. The design of the GHSL processing workflow seeks a compromise between the goal of global consistency and the need of exploitation of finer-scale training sets if available.

The training set management is consistent with the overall design of the experiment that is evolutionary with a retro-action mechanism. The built-up areas detected in the $GHSL_{\alpha}$ are calculated by association analysis of the image data instances X using $BU_{ref}(Y_{mds}, Y_{lsc})$ as training set. As in [41, 39], BU_{ref} is made by adaptive learning of the best population thresholds discriminating built-up areas from Y_{lsc}, using Y_{mds} as training set. The built-up areas reported in $GHSL_{\beta}$ are also calculated by association analysis of the X data instances but using a more comprehensive training set including fine-scale data and the $GHSL_{\alpha}$ output classification (the retro-action mechanism). The harmonization of the input training set may be a very tedious labor-intensive task if including a large mass of heterogeneous partially-not-documented (completeness, update) fine-scale data. In the proposed approach the harmonization of the input training set is solved automatically by composition of the whole list of available training sets, augmented by the $GHSL^{e}_{\alpha}$ output, with e =1975, 1990, 2000 and 2014 corresponding to the epoch of the scene under processing. The process uses a voting schema that ends by setting two hypotheses of training masks $M : L \rightarrow \{0, 1\}$, respectively, *strict* M_{bu0} and *extended* M_{bu1} that are used as input by the $GHSL_{\beta}$ learning and classification procedure. They are calculated as follows:

$$Y_{lr} = M(BU_{ref}) + M(Y_{mds}) + M(Y_{glc}) \quad (1)$$

$$Y_{hr} = 2 * M(GHSL^{e}_{\alpha}) + M(\bigcup(Y_{osm_{b1}}, Y_{osm_{b2}}, Y_{gnm}, Y_{nlc})) \quad (2)$$

$$M_{bu0} = \begin{cases} 1, Y_{lr} + Y_{hr} > 3 \\ 0, otherwise \end{cases} ; M_{bu1} = \begin{cases} 1, Y_{lr} + Y_{hr} > 1 \\ 0, otherwise \end{cases} \quad (3)$$

with Y_{lr} and Y_{hr} denoting the low and high resolution training sets respectively. Each set Y_{lr} and Y_{hr} have three votes available. From the global point of view, the fine-scale sources are typically scarce, incomplete and missing the time attribute: consequently they are aggregated by the union operator contributing to Y_{hr} with one vote. The other two votes are derived from the fine-scale output of the $GHSL_{\alpha}$ process, introducing a retro-action with a temporal attribute. The whole process is done at the full spatial resolution of the specific scene under processing: a total of 6 votes are available for a total agreement between the training sources. The strict and large training set hypothesis are the sets consisting of the image elements where the agreement of at least 4 and 2 sources, respectively, can be found.

3.4. Testing Set

In the experiment, two families of test sets are introduced: i) validation sets, and ii) cross-comparison sets. The *validation* set includes the geo-information sources where an absolute better scale detail and thematic reliability was expected, when compared to the output of the GHSL Landsat. The *cross-comparison* set includes the geo-information sources that show some similarity with the GHSL Landsat in scale and thematic contents, but cannot be used as absolute reference. Both validation and cross-comparison sets were not included in the training set collections, thus independent from the supervised classification process. Due to the bounding conditions determined by the data and resources available for implementing the experiment, only the built-up area information extracted in the epoch 2014 was validated. Built-up areas of all the epochs (1975, 1990, 2000, and 2014) were tested by comparison with products that are similar to the GHSL outputs but embedding some methodological differences: consequently they can be taken only as cross comparison exercises.

Additionally, Open Street Map data have been evaluated as a potential source of building footprints (extracted from GeoFabrik.de in 2013). However, they have not be used for validation due to reported incompleteness and semantic inconsistency issues emerged in the phase of the experimental set design.

3.4.1. Validation

Two validation sources were used: a systematic field survey managed by the statistical office of the European Union (EUROSTAT) and a set of digital cartographic products with building footprints at a scale 1:10.000 or better.

The Land Use/Cover Area frame Survey (LUCAS)[11] is a harmonized in situ land cover and land use data collection exercise that extends over the whole of the EU's territory with an update period of three years. LUCAS sampling schema is based on a systematic grid with points spaced 2km apart in the four cardinal directions covering the EU's territory. The points are classified during direct field visit. Points above 1,500 meters of altitude and far from the road network are excluded form the survey in order to limit the cost. The last survey of 2012 used in this study covered all of the EU-27 member states and was based on 270,000 points/observations. In this [14] survey, eight main land cover categories are collected, and the "artificial land" class is divided into three subclasses listed below: **A11** - Buildings (low). Roofed constructions with one to three floors or less than 10 meters of height in total including: single-family houses, mobile homes, summer cottages, industrial or cultural buildings or stores, agricultural buildings, market halls, temporary constructions; **A12** - Buildings (high). Roofed constructions with more than three floors or more than 10 meters of height in total including: industrial buildings, stores, technical infrastructures, residential or cultural

[11] http://ec.europa.eu/eurostat/web/lucas/overview

buildings, agricultural buildings; **A13** - Greenhouses. Installations of glass, plastic or any other material which is translucent but impervious to water. For the validation of the GHSL, the three above mentioned classes were aggregated into a single reference *built-up* class. From the 270,276 LUCAS points available in the 2012 survey, 10,786 were excluded from the analysis because they included no data values in the GHSL data set.

The cartographic data is a collection of datasets which contains building footprints. The datasets were gathered from data portals of official web sites of public governmental institutions, such as counties, municipalities or city councils. The building footprints had been usually derived from aerial photographs, digital ortho-photography and also (if updated) urban planning projects. The temporal signature of the datasets varies between 2000 and 2010, and the data cover areas in USA and Europe mainly. In general, the data were available in a vector format. In order to increase the statistical interest of the test, the final collection of 27 datasets of available cities and regions was rasterised at 2.5m and then tiled following TMS grid notation (Tile Map Service - a global hierarchical quad-tree tiling schema based on Google Mercator projection) with the resolution of the GHSL MT dataset (i.e., 38m). As a result, 3826 sample raster tiles were created with the side size of approx. 10km (in Google Mercator metric). Only the tiles resulting with a valid data area greater than 50% of the tile surface were considered during the validation (i.e. 1505 tiles). In total, the validation tiles cover an area of 133,909 km^2 with a total built-up area of 4,656 km^2. Table 1 shows the distribution of built-up density in the tiles. We can observe that almost 50% of the tiles are characterized by built-up density lower than 1% and 11% of tiles have built-up density greater than 10%. There are only two tiles with the built-up density greater than 30%.

Built-up (%)	Number of tiles	Cumulative (%)
0 - 1	700	46.51
1 - 5	496	79.47
5 - 10	138	88.64
10 - 32	171	100.00

Table 1: Built-up density in the validation tiles.

3.4.2. Cross-comparison

The cross-comparison set included three sources of geo-information, namely the European Settlement Map (ESM), the results from the Urbanization Project of the New York University (NYU) and the Geonames dataset.

The European Settlement Map (ESM) represents an output generated with 2.5m and 1.5m resolution multispectral SPOT-5/6 images from the period 2011-2012 in input for the whole Europe as in [16]. The criteria for discrimination of built-up areas from SPOT data was based on textural and morphological image features as presented in

[41].

The second dataset used for cross-validating our results was the baseline data kindly made available by the Urbanization Project of the New York University (NYU)[12] and described in [2]. The cities included in the NYU baseline data are an extension of the global stratified sampling methodology introduced by [3]. The methodology for built-up areas collection in the NYU data relies on human supervised classification of Landsat image data organized in three epochs 1990, 2000, and 2010. Per-scene expert tuning of the processing parameters and exhaustive visual checking of the output is included in the workflow of the NYU data.

The third member of our cross-comparison set is represented by the Geonames dataset that has been derived from the GeoNames project[13], which is an online free geographical database of location names (point data in WGS84). It was downloaded in April 2014. In total, the dataset contains more than 8.5 million of features, of which more than 3.2 million are classified as populated places.

3.5. Symbolic Machine Learning

The data classification experiment described here largely relies on a new supervised classification technique introduced by [43] as scalable solution to complex and large multiple-scene satellite data processing scenarios. The technique is based on Symbolic Machine Learning (SML) and consists of two main steps: i) image data *quantization-sequencing*, and ii) *association rule analysis*.

Let $D_{m\times n\times F}$ be a data set with $m \times n = mn$ spatial samples or pixels and F features or descriptors. Let $X_{mn\times F}$ be a 2-dimensional data matrix, $X = [\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_i}, \ldots, \mathbf{x_F}]$, with F expressing the number of used features and $\mathbf{x_i} \in \mathbb{Z}_+^{mn}$. Let $\widehat{X}$ be the set of all the unique data instances of X.

In the experiment, a uniform quantization approach was implemented during the data quantization-sequencing phase, the same for each feature of an epoch. $X(q) = [Round\,(x/q)]_{mn\times F}$ are the quantized data instances with $q = max(x)/s$ where $x \in \mathbf{x_i}$. Table 2 shows the parameters applied to the image data of the different epochs.

Year	1975	1990	2000	2014
q	2	4	8	1024
s	128	64	32	16
F	4	6	6	7 (+1)
$max\ x$	255	255	255	65535

Table 2: The values of quantization parameter q, number of symbols s, and number of attributes F applied in the different data epochs.

The q value was determined empirically by observing the mean support of $\widehat{X}$ estimated as $supp_\mu = |X| / \left|\widehat{X}\right|$ and by choosing a value

[12] http://urbanizationproject.org/

[13] http://www.geonames.org/

so that $supp_\mu$ is in the range of 10^2–10^3. These $supp_\mu$ orders of magnitude were tested as satisfactory for classification exercises incorporating noisy training sets, with noise characteristics similar to the ones experimented here [43]. The F number of available spectral attributes of X is dictated by the sensor characteristic in the specific epoch, while the s number of available symbols per each spectral attribute is derived by the combination of the raw DN encoding, the specific sensor data D and the q parameter. In the case of the epoch 2014 with L8 data, the image data sequencing included also the new Quality Assessment (QA)[14] band. This band encodes some surface, atmosphere, and sensor conditions that can affect the radiometric information collected by a given image element (pixel). Due to the symbolic nature of the SML classifier, it can be introduced seamlessly in the input data attributes used for classification. In the case of the QA attribute, because already encoding a classification output, no quantization was applied.

The confidence family measure named Evidence-based Normalized Differential Index (ENDI) Φ_E [43] is noted as $Z = \Phi_E(X, Y^+, Y^-)$ with X the quantized input image data instances, Y^+ the set of positive evidences and Y^- the set of negative evidences used during the training phase. It is worth noting that in the general Φ_E formulation the mutual exclusivity of the Y^+, Y^- training sets is not required and consequently $Y^+ \cap Y^- \geq \emptyset$. This property allows the exploitation of partially overlapping or inconsistent (by scale, by thematic contents) training sets like the ones handled by this experiment.

The Φ_E assumes values in the range $[-1, 1]$. They express the estimated association between the data instances and the target class or foreground information. Values close to 1 indicate data instances strongly associated with the presence of foreground information as summarized by the positive evidences (e.g. buildings), while values close to -1 are strongly associated with the presence of the complement of the foreground information (background, e.g. not built-up) as summarized by the negative evidences. Values of Φ_E close to 0 indicate that for the specific data instance, a comparable number of evidences can support both the foreground or background information hypothesis. Consequently, they are placed at the boundaries of the decision space.

[14] `http://landsat.usgs.gov/L8QualityAssessmentBand.php`

4. Information extraction

The information extracted for all the scenes include i) *clouds*, ii) *water*, and iii) *built-up*. The automatic recognition uses a hierarchical per-scene processing chain collecting the components with a precise order and propagating the information discovered at each step to the next. For recognition of the *water* and *built-up* components the procedure is designed in three main steps running for all data instances (scenes) X of all epochs: i) collection of the available training set at the geometry (resolution, projection) fitting the specific X scene, ii) automatic refinement and improvement of the training set based on the specific X data instances under processing and theory-driven reasoning, and iii) supervised classification of the specific X data instances based on data-driven reasoning (association analysis) and refined training set. The per-scene processing schema is supposed to reduce the problem of classification model transfer between scenes collected in different seasons and/or different geographical places, thus increasing the reliability of the automatic information extraction process. The theory-driven reasoning applied in the training set refinement step exploits known relations between image data and target information in order to automatically improve the reliability of the training set used in the subsequent supervised classification. The X data instances used for the classification are the ones recorded by the multi-spectral sensor available in the different epochs. The panchromatic channel was used exclusively for improving the training set used in the supervised built-up areas classification.

4.1. Clouds

Cloud detection identifies the valid image data domain that will be further processed. For each collection (1975, 1990, and 2000), we applied an adaptation of the Automatic Cloud Cover Assessment (ACCA) method, similar to that used in [41]. The 2014 Landsat 8 data set was processed using a direct query to the new Quality Assessment (QA) band including the coding of the cloud coverage was applied.

During the system design phase, the test of the information stored in the QA band showed that highly reflecting rooftops as in new industrial/commercial buildings have high probability to be erroneously classified as *cirrus* and *cloud*. In order to mitigate the risk to systematically miss classify highly reflecting rooftops in no data, a connected-component morphological filtering of the *cirrus* and *cloud* sets derived from the QA was performed. The filter is based on area criteria discriminating large patches contributing to the cloud mask M_{cloud} and small, scattered residual patches stored in the M_{high} mask.

4.2. Water surfaces

Water surfaces are recognized by associative classification using the Boolean mask M_w as training set and a Φ_E filtering by morphological

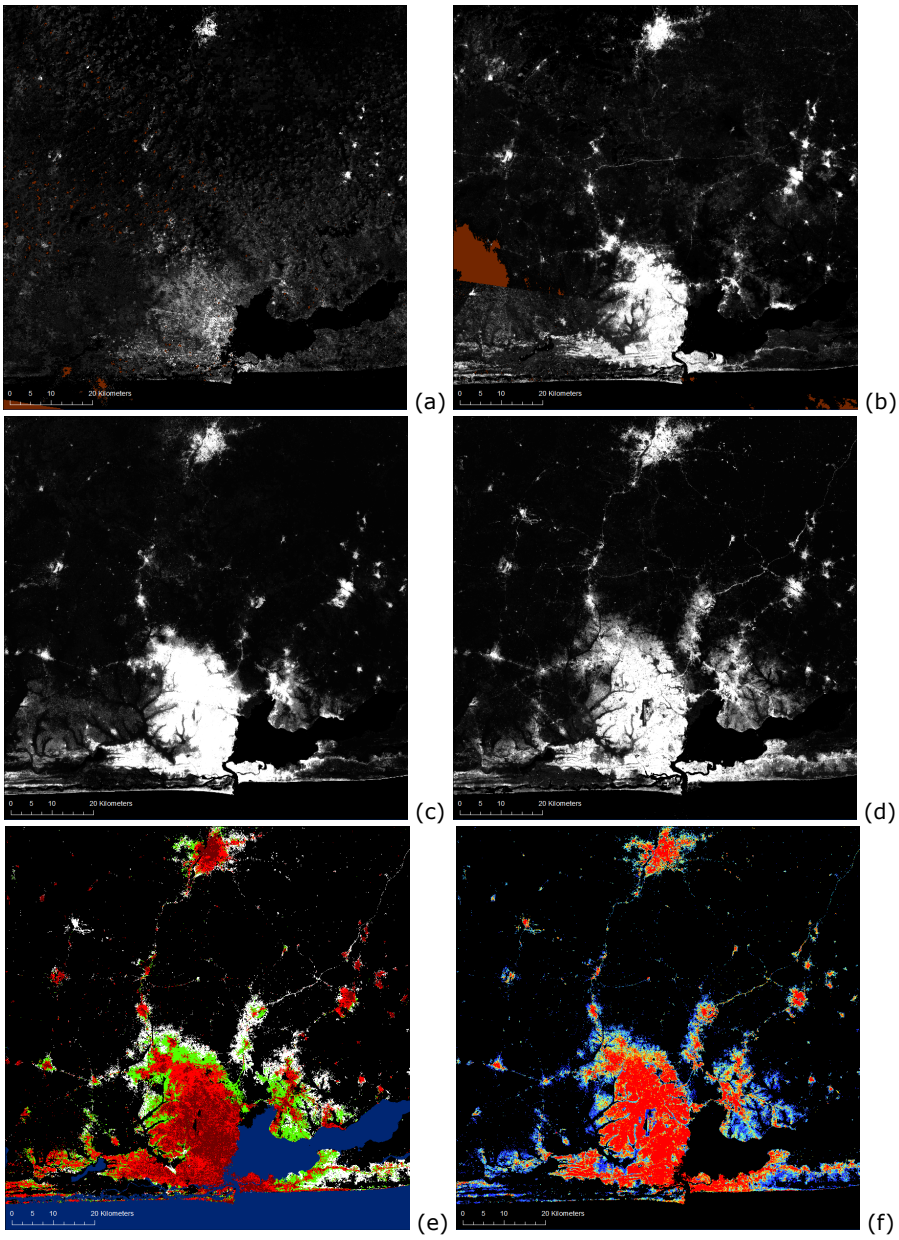


Figure 3: Lagos, Nigeria - Φ_E of the class *built-up* estimated from the Landsat data of the epochs 1975 (a), 1990 (b), 2000 (c) and 2014 (d) (brown = cloud mask). Multitemporal assessment (e): dark red = built before 1975, red = built from 1975 to 1990, green = built from 1990 to 2000, white = built from 2000 to 2014, blue = water bodies. (f): μ_{bumix} model output for 2014

reconstruction as in [47, 11]. At first instance, the M_{wlr} training set is derived by class code query from the low-resolution global land cover

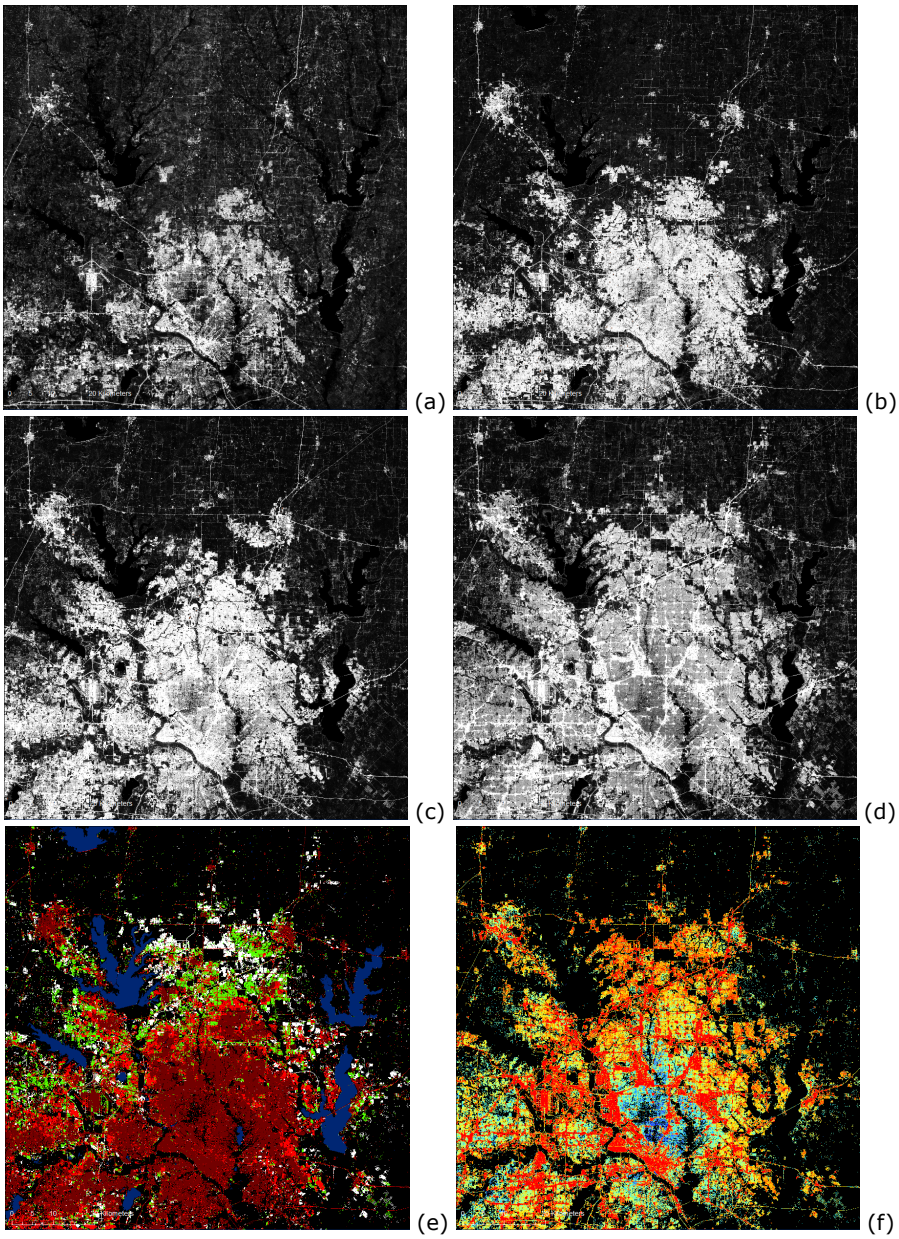


Figure 4: Dallas, US - Φ_E of the class *built-up* estimated from the Landsat data of the epochs 1975 (a), 1990 (b), 2000 (c) and 2014 (d). Multitemporal assessment (e): dark red = built before 1975, red = built from 1975 to 1990, green = built from 1990 to 2000, white = built from 2000 to 2014, blue = water bodies. (f): μ_{bumix} model output for 2014

source Y_{glc}. If a mid-infrared spectral band X_{mIR} is available in the specific scene under processing, then it is used in order to refine the

scale of M_{wlr}. Let μ and σ be the mean and standard deviation of the $X_{mIR} : x \in M_{wlr}$. Let M_{w^-} and M_{w^+} be defined as $M_{w^-} \in \{X_{mIR} : x < \mu + 2 * \sigma\}$ and $M_{w^+} \in \{X_{mIR} : x \leq \mu - 0.5 * \sigma\}$. The refined water training set is defined as $M_{whr} \leftarrow (M_{wlr} \cap M_{w^-}) \cup M_{w^+}$. Let $Z_{w0} = \Phi_E(X, Y^+ \in M_{whr}, Y^- \in \neg M_{whr})$ be the Φ_E score estimated for the X data instances using the M_{whr} training mask. And let M_2 and M_1 be defined as $M_2 \in \{x : Z_{w0} > 0.0\}$, $M_1 \in \{x : Z_{w0} > 0.1\}$. The output water mask M_{w0} is calculated according to [11] as reconstruction by dilation of the marker M_1 under the mask M_2. In case of L8 input data, a second hypothesis of water surface is made: $M_{w1} = M_{wlr} - M_{w0}$, including the samples that were labeled as water in the low-resolution data but not water in the high resolution mask derived from the image data instances. They are classified as *water interface* and a specific Φ_E is calculated accordingly as $Z_{w1} = \Phi_E(X, Y^+ \in M_{w1}, Y^- \in \neg M_{w1})$. Z_{w1} will be integrated in the final multiple-class land cover product.

4.3. Built-up areas

The confidence-like metric for the built-up areas in the $GHSL_\alpha$ are estimated by $\Phi_E(X)$ using $BU_{ref} \leftarrow (Y_{mds}, Y_{lsc})$ as training set. The confidence-like metric for the built-up areas in the $GHSL_\beta$ are also calculated by $\Phi_E(X)$ but using a more comprehensive training set aggregated by a voting schema in a M_{bu0} *strict* and M_{bu1} *extended* built-up training set hypothesis. The input X data instances used for the classification are the ones collected by the multi-spectral sensor of the Landsat satellite. The panchromatic channel - if available in the specific epoch - was used only for the improvement of the training sets through textural image analysis. Before the supervised classification, the built-up training sets are refined using image data instances and unsupervised reasoning. Let $Y \in \{BU_{ref}, M_{bu0}, M_{bu1}\}$ be any of the three training sets used for built-up detection, the refined set is defined as $Y_r = Y_r1 \cup Y_r2$, where $Y_r1 = Y \cap \neg\{M_{w0} \cup M_{ndvi0} \cup M_{ptx0}\}$ and $Y_r2 = \{M_{ptx1} \cap M_{ndvi1}\}$.

The $M_{ndvi0,1}$ and $M_{ptx0,1}$ masks are calculated only in the 2014 collection where suitable data and meta-data were available. In particular, a normalized differential vegetation index X_{ndvi} feature was extracted using the top-of-atmosphere (TOA) reflectance values and used as input of the $M_{ndvi0,1}$ masks with $M_{ndvi0} \in \{X_{ndvi} : x < 0.65\}$ and $M_{ndvi1} \in \{X_{ndvi} : x < 0.3\}$. Moreover, a multi-scale textural analysis of the 15m resolution panchromatic channel of the L8 data was performed in order to calculate the $M_{ptx0,1}$ masks contributing to the improvement of the available training sets.

The textural analysis was made by application of anisotropic rotation-invariant contrast textural measures extracted using the grey-level co-occurrence matrix (GLCM) and the *pantex* methodology of [40] extended to a multi-scale domain.

Let X_{ptx} be the multi-scale *pantex* values calculated from the 15m resolution P data using a displacement vector of 1, 2 and 9 pixels (correspondingly 15, 30, 135 meters) and a sliding window size of 15x15 pixels (225 square meters). The X_{ptx} data were spatially

aggregated (by average) at the resolution of the multi-spectral channels and an association analysis was performed in order to estimate the index Z_{buptx} for the presence of built-up areas, based on textural information. Consequently, $Z_{buptx} = \Phi_E(X_{ptx}, Y^+ \in M_{bu1}, Y^- \in \neg M_{bu1})$. The cut-off points are determined by observation of the Φ_E index values in the training mask M_{bu1} and the complement $\neg M_{bu1}$. Let μ_0 and σ_0 be the mean and standard deviation of the index $Z_{buptx} : x \in \neg M_{bu1}$ and let μ_1 and σ_1 be the mean and standard deviation of the index $Z_{buptx} : x \in M_{bu1}$. The $M_{ptx0,1}$ masks are calculated as $M_{ptx0} \in \{Z_{buptx} : x > \max(-1, \mu_0 - \sigma_0)\}$ and $M_{ptx1} \in \{Z_{buptx} : x > \min(1, \mu_1 + 2 * \sigma_1)\}$.
The final Φ_E values of the built-up class for the $GHSL_{\alpha,\beta}$ loops, correspondingly $Z_{bu\alpha}$ and $Z_{bu\beta}$ were estimated as follows:

$$Z_{bu\alpha} = \Phi_E(X, Y^+ \in Y_r(BU_{ref}), Y^- \in \neg Y_r(BU_{ref})) \quad (4)$$

$$Z_{bu\beta} = \Phi_E(X, Y^+ \in \{Y_r(M_{bu0}) \cup Y_r(M_{bu1})\}, Y^- \in \neg Y_r(M_{bu0})) \quad (5)$$

The $Z_{bu\beta}$ estimation included a stratification of the X data instances based on the Y_{glc} prior knowledge. The stratification aimed to mitigate class-dominance effects that may be present in specific scenes under processing. In particular, strata with more than $|D|/5$ instances (pixels) were classified separately in the same scene. Moreover, a test on the number of samples falling in Y_{glc} water surfaces was made: a maximum number of water samples equal to the number of land samples was accepted.

Figures 3 and 4 present the Φ_E results of the built-up class for the four epochs in Lagos (Nigeria) and Dallas (US).

5. Landsat 8 specific processing

For the 2014 L8 data, a specific data processing chain was implemented. We tested the introduction of a multiple-class land cover classification scheme, and a new information unmixing model for a sub-pixel built-up areas classification scheme was proposed. They are described in the next paragraphs.

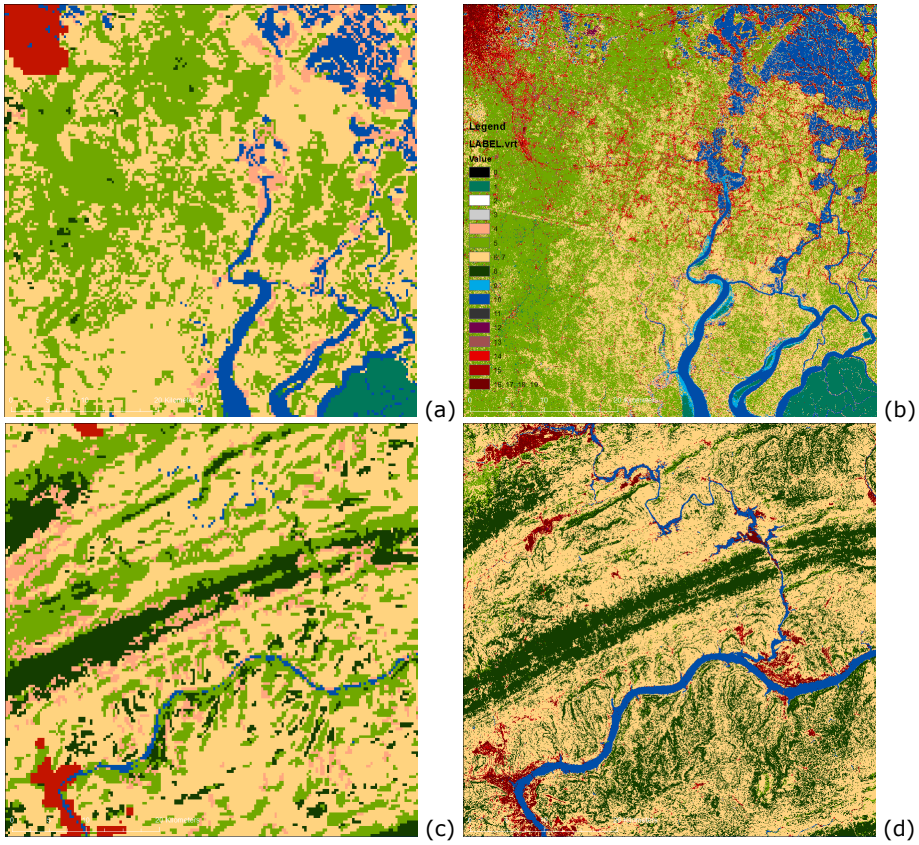


Figure 5: Examples of the multiple-class land cover outputs. (a,c) - SW of Calcutta (India) and the area of the Wanxian town (China) as represented by the learning set Y_{glc}. (b,d) - the output of the GHSL multi-class product in the same areas. For legend encoding see subsection 5.1.3.

5.1. Multiple-class land-cover/land-use

In this phase, available information extracted during the processing is used for a multiple-class land cover extraction by maximization of Φ_E index. The Z_{bu}, Z_{w0} and Z_{w1} scores estimated in the precedent steps are evaluated together with Φ_E related to a set of new training classes. They are extracted from the roads available in Y_{osm} repository and the low-resolution land cover Y_{glc}. In the applied schema,

the Φ_E of each training class is evaluated separately considering the specific set as positive evidence and using all the others training classes as background for negative evidence samples. Finally, the label of the class is evaluated by maximization of the Φ_E value. For the purpose of the experiment, the original Y_{glc} encoding is aggregated in a fewer number of classes as follows:

$$Forest = Y_{glc1} = y : Y_{glc} \in \{10 * z : z \in \mathbb{N}, 4 \leq z \leq 12\} \quad (6)$$

$$IrrigatedCropland = Y_{glc2} = y : Y_{glc} \in \{14\} \quad (7)$$

$$RainCropland = Y_{glc3} = y : Y_{glc} \in \{11\} \quad (8)$$

$$MosaicCroplandForest = Y_{glc4} = y : Y_{glc} \in \{20, 30\} \quad (9)$$

$$ShrubsGrassland = Y_{glc5} = y : Y_{glc} \in \{130, 140, 150\} \quad (10)$$

$$BareLand = Y_{glc6} = y : Y_{glc} \in \{200\} \quad (11)$$

$$IceSnow = Y_{glc7} = y : Y_{glc} \in \{220\} \quad (12)$$

$$Other = Y_{glc8} = y : y \notin \bigcup \{Y_{glc1}, \ldots, Y_{glc7}\} \quad (13)$$

5.1.1. Basic components

The Φ_E matrices extracted from the used training sets are stacked in a feature cube Z and then maximized. The following order was applied: $Z_1 = \{Z_{bu}, Z_{road}, Z_{w0}, Z_{w1}, Z_{glc1}, \ldots, Z_{glc8}\}$ reflecting an abstract relevance of the specific class from the point of view of the human settlement mapping. The first class is the built-up areas, while the last one encodes the residuals of any other class not considered in the classification schema. The classification procedure applies a two-stage approach with a retro-action mechanism aiming to handle the case where training sets may not sufficiently describe the data instances (presence of unexpected new classes) or data instances may not be able to discriminate with sufficient reliability the considered training sets. At the first trial the maximal Φ_E index is evaluated as $Z_{max} = \max(Z_1)$. The data domain where $Z_{max} < 0$ is considered problematic, because it is not showing strong association between data instances and any of the training sets classes. Consequently, the samples belonging to this data domain are added to the *other* training set class Y_{glc8}, a new Z_{glc8} score is evaluated again accordingly (the retro-action mechanism), and a new stack Z_2 is made. Finally, the second trial evaluates the class label as the class where the maximal Φ_E is found $L = \text{argmax}(Z_2)$. Figure 5 shows the results of the method for two examples in India and China.

5.1.2. Built-up areas characteristics

Let L_{bu} be the samples of L which are labeled as *built-up* by the supervised classification process described above. They are further reclassified by an unsupervised procedure with the purpose of discriminating some main characteristics of built-up areas that can be useful for the GHSL applications. In particular, two main characteristics of built-up areas are evaluated in this phase: i) the amount

of vegetation and ii) the amount of 3D roughness. The first characteristic is extracted from the normalized differential vegetation index X_{ndvi} calculated during the process. The second characteristic X_{dsm} is extracted by processing of digital surface model (DSM) input data derived from the Shuttle Radar Topography Mission[15] (SRTM) and the ASTER Global Digital Elevation Model[16] (GDEM). SRTM data characterization in urban areas was firstly introduced by [17] using the residuals of local average filtering. In the proposed approach, a new method was adopted allowing the fusion of the two DSM sources and reducing the false alarms in mountainous areas. A high-band-pass filter is made by the application of opening and closing morphological operators to both the SRTM and GDEM sources with a structuring element of 250 meters and observation of their residuals. The maximal opening/closing residual signal between the two sources is collected and subsequently processed by textural contrast analysis. In particular, the *pantex* procedure based on grey-level co-occurrence matrix (GLCM) was applied. The procedure aims to measure the presence of systematic local contrast in the DSM morphological residuals, reducing the false alarms and noise related to the observation of the residuals only.
The classification schema of built-up areas characteristics based on NDVI is made as follows:

$$VeryLightBU = L_{bu1} = L_{bu} : X_{ndvi} > 0.4 \quad (14)$$

$$LightBU = L_{bu2} = L_{bu} : 0.3 \leq X_{ndvi} < 0.4 \quad (15)$$

$$MediumBU = L_{bu3} = L_{bu} : 0.2 \leq X_{ndvi} < 0.3 \quad (16)$$

$$StrongBU = L_{bu4} = L_{bu} : X_{ndvi} < 0.2 \quad (17)$$

The classification schema of built-up areas characteristics based on DSM is only applied to the samples labeled as *strong built-up* L_{bu4} that are presumably minimizing the noise originated from trees canopy. The classes are defined as follows:

$$LowRiseBU = L_{bu5} = L_{bu4} : X_{dsm} < 25m \quad (18)$$

$$MediumRiseBU = L_{bu6} = L_{bu4} : 25m \leq X_{dsm} < 50m \quad (19)$$

$$HighRiseBU = L_{bu7} = L_{bu4} : 50m \leq X_{dsm} < 100m \quad (20)$$

$$VeryHighRiseBU = L_{bu8} = L_{bu4} : X_{dsm} \geq 100m \quad (21)$$

Figure 6 shows the results of the proposed urban characterization method in a sample of the Chicago (US), Tokyo (Japan), and Johannesburg (South Africa) cities.

5.1.3. Multiple-class GHSL encoding

The final legend encoding of the multiple-class GHSL schema is reported below:
0 = No data

[15] http://www2.jpl.nasa.gov/srtm/
[16] http://gdem.ersdac.jspacesystems.or.jp/

1 = Other
2 = Ice and permanent snow
3 = Bare soil and rocks
4 = Shrubs and Grassland
5 = Mosaic Croplands and Forest
6 = Rain Cropland
7 = Irrigated Cropland
8 = Forest
9 = Occasionally water / land-water interface
10 = Surface Water
11 = Roads
12 = Built-up with highly reflecting roof (associated to productive and commercial use)
13 = Very light built-up NDVI > 0.4
14 = Light built-up 0.3 < NDVI <= 0.4
15 = Medium built-up 0.2 < NDVI <= 0.3
16 = Strong built-up NDVI <= 0.2 and low rise buildings (3D roughness <25m)
17 = Strong built-up NDVI <= 0.2 and medium rise buildings (3D roughness 25-50m)
18 = Strong built-up NDVI <= 0.2 and high rise buildings (3D roughness 50-100m)
19 = Strong built-up NDVI <= 0.2 and very high rise buildings (3D roughness > 100m)

5.2. Degree of built-up

The estimation of the degree of built-up per cell is obtained by applying a fuzzy information unmixing model, the aim of the model being the estimation of the *percentage of the cell covered by a building*. The model uses as input the Φ_E index of built-up Z_{bu}, roads Z_{road}, and water Z_{w0}, together with the X_{ndvi} index. The membership functions μ are derived from linear rescaling of Z measurements inside the $[0,1]$ range by application of fixed min, max parameters. Let $y = rs(x,a,b)$ be the linear rescaling of x with a and b being the min and max parameters, respectively. The membership functions μ are defined as follows:

$$\mu_{bu} = rs(Z_{bu}, -0.3, 0.3) \tag{22}$$

$$\mu_{road} = rs(Z_{road}, -0.8, 0.8) \tag{23}$$

$$\mu_{green} = rs(X_{ndvi}, 0.1, 0.8) \tag{24}$$

$$\mu_{water} = rs(Z_{w0}, -0.5, 0.5) \tag{25}$$

The degree of built-up per cell (μ_{bumix}) was estimated by considering μ_{road}, μ_{green}, and μ_{water} as competitors of the μ_{bu} surfaces in a geometric fuzzy intersection equation as follows:

$$\mu_{bumix} = \mu_{bu} * (1 - \mu_{road}) * (1 - \mu_{green}) * (1 - \mu_{water}) \tag{26}$$

where, by definition, $\mu_{bumix} \leq \mu_{bu}$ and $\mu_{bumix} \in [0,1]$. The parameters applied in the model are tested in the experiment discussed here. They are derived from knowledge-driven reasoning and direct inspection of experts in few scenes taken as examples. Figure 7 shows the output of the proposed model μ_{bumix} in few examples including (a)

Cairo (Egypt), (b) Mexico City (Mexico), (c) Harare (Zimbabwe), (d) New Delhi (India), (e) London (UK), and (f) Minneapolis (US). Example of μ_{bumix} outputs are also included in the Figure 3f) and Figure 4f) showing the cities of Lagos (Nigeria) and Dallas (US). In contrast with the land cover categorical results (see Fig 3e), 4e) and 5), the output of the building density analysis highlights the strong differences in the structure of cities, which reflect the compactness as well as social or functional spatial segregation patterns, which however require further analysis.

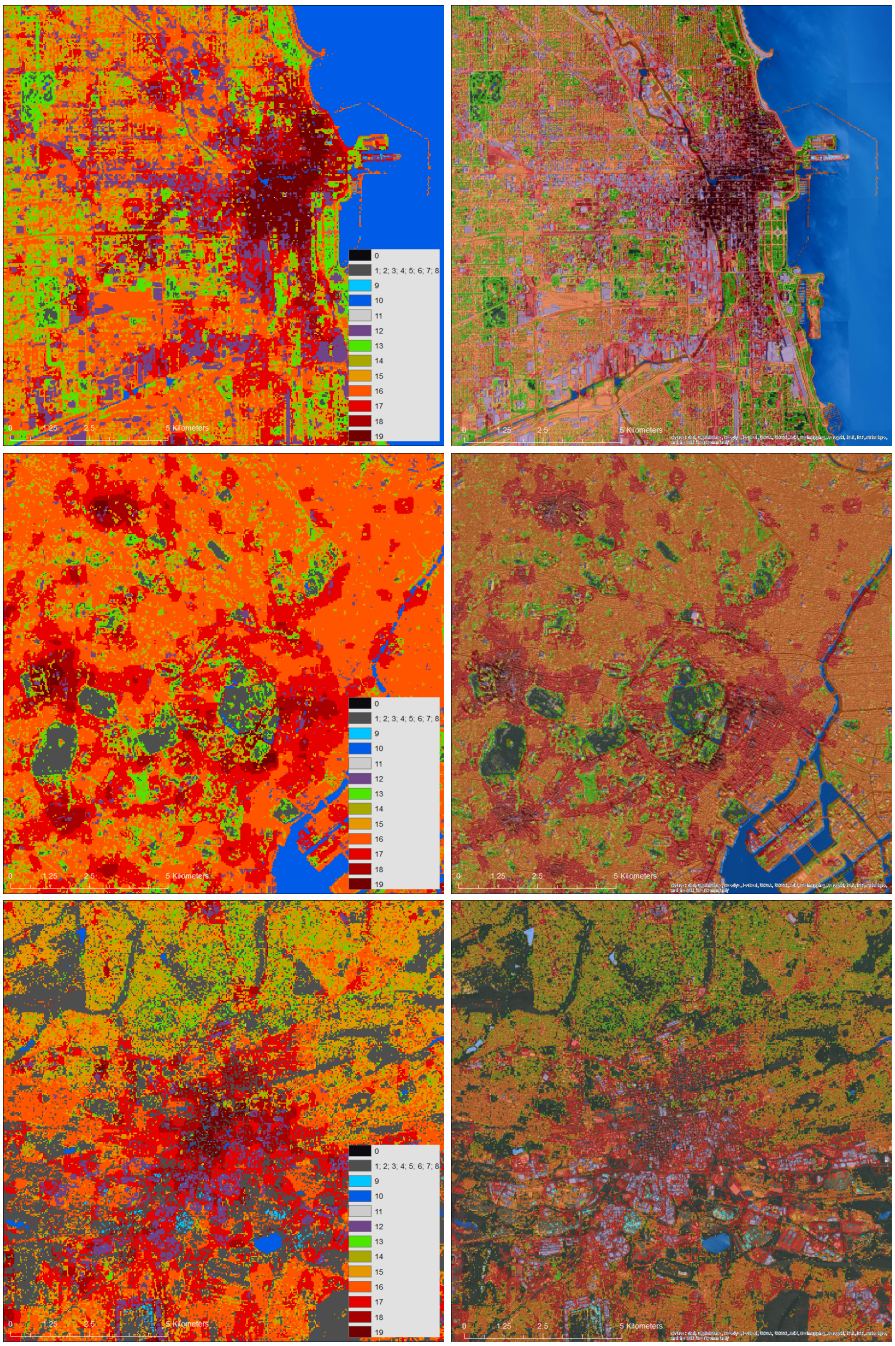


Figure 6: Examples of built-up areas characteristics as detected by the experiment. From top to bottom Chicago (US), Tokyo (Japan), and Johannesburg (South Africa). Left: the output of the classification, right: the same fused with background high resolution imagery (source: Bing) for visual reference. For legend encoding see subsection 5.1.3.

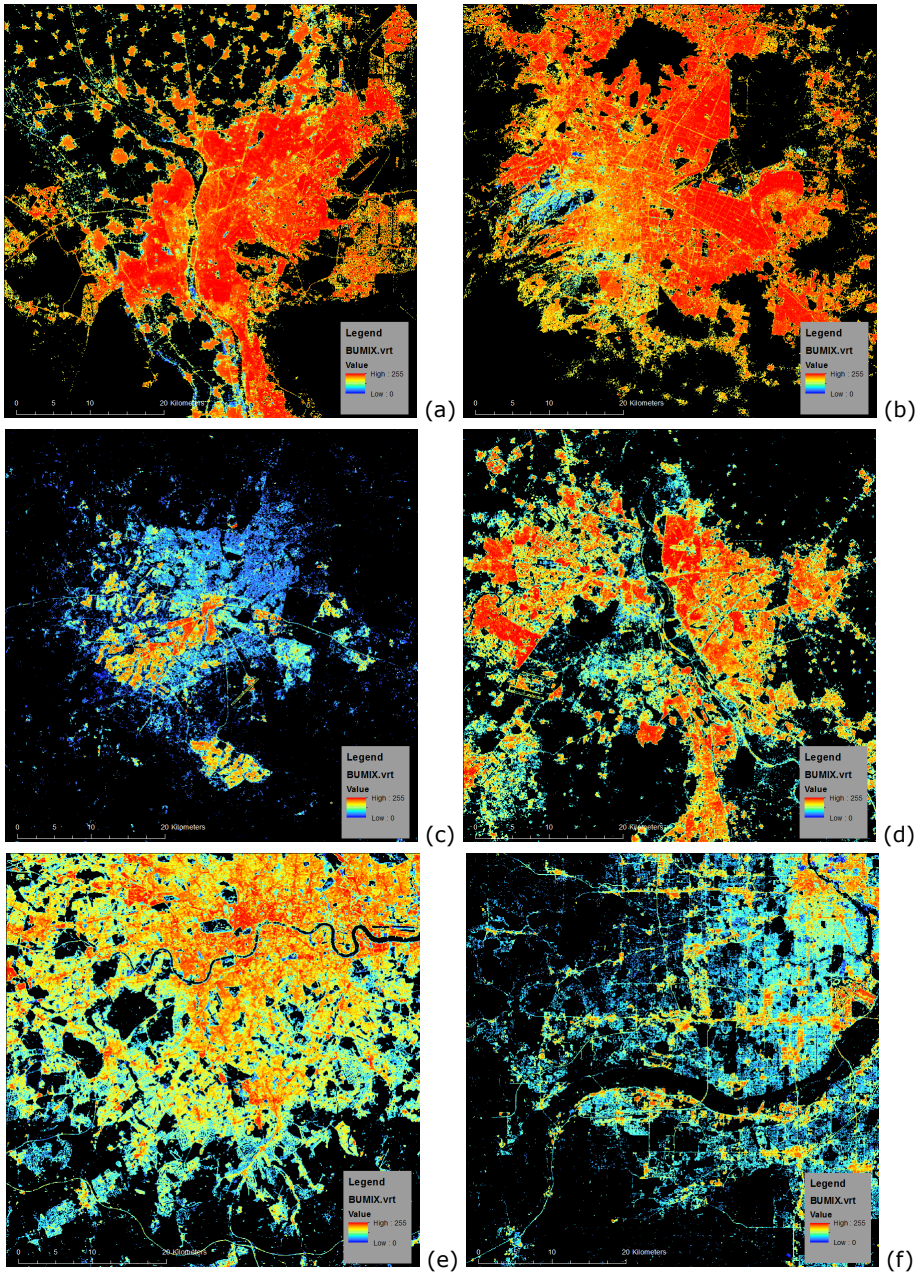


Figure 7: Examples of the μ_{bumix} model output in some cities. Same scale and same color code form blue (low), yellow-green (medium), red (high) values. (a) Cairo (Egypt), (b) Mexico City (Mexico), (c) Harare (Zimbabwe), (d) New Delhi (India), (e) London (UK), and (f) Minneapolis (US).

6. Multiple-scene information fusion

Information redundancy is common in large data classification projects. It is originated by the fact that the same sample on the ground can be mapped more than once from the same or different sensors at different times. This redundancy can be managed by information fusion processes in order to achieve explicit objectives as keeping specific information contents, reducing data gaps or reducing overall error rates. Figure 2 reports about the redundancy in each data collection processed and the table 3 shows the total redundancy in the experiment. The first column of the table contains the number of available input data scenes while columns 2 and 3 comprise the corresponding number of pixels and land areas covered by these superposed scenes. Finally, columns 4 and 5 display the percentages and cumulative percentages that the areas of column 3 represent from the global landmass.

Two main information fusion processes have been applied: one dedicated to the production of the global mosaic and the other to the multi-temporal information processing. They are described in the next sections.

Number of scenes	Number of pixels	Area (km^2)	Percentage (%)	Cumulative Percentage (%)
0	1,054,383	196,387.02	0.146	0.146
1	369,481	261,254.63	0.195	0.341
2	2,130,282	2,145,715.66	1.598	1,939
3	6,163,122	6,516,632.61	4.854	6.793
4	23,561,078	24,795,286.10	18.468	25.261
5	23,873,667	23,517,729.83	17.517	42.777
6	17,506,471	15,469,685.41	11.522	54.299
7	17,245,074	14,425,595.10	10.744	65.044
8	18,346,845	14,864,559.75	11.071	76.115
9	15,036,569	11,572,263.28	8.619	84.735
10	10,563,451	7,839,995.08	5.839	90.574
11	7,290,252	5,269,666.62	3.925	94.499
12	4,931,081	3,541,354.66	2.638	97.137
13	2,965,265	2,075,928.88	1.546	98.683
14	1,511,248	1,016,313.14	0.757	99.440
15	696,409	446,293.82	0.332	99.772
16	305,668	191,486.76	0.143	99.915
17	134,510	80,529.63	0.060	99.975
18	44,368	26,040.28	0.019	99.994
19	11,276	6,625.27	0.005	99.999
20	1,974	1,035.32	0.001	100.000
21	46	27.57	0.000	100.000
Grand Total	153,742,520	134,260,406.44	100.000	100.000

Table 3: Total dataset redundancy situation over land areas.

6.1. Global mosaic information fusion

As in [41], a TMS global hierarchical quad-tree tiling schema has been adopted for combining the whole information collected by the differ-

ent scenes in a single mosaic. TMS provides computational efficiency, facilitating the parallelization of the mosaic process. We adopted a nominal spatial resolution at the equator of 38.21m corresponding to the zoom level 12 of the TMS schema which best approximates the native 30m spatial resolution of the output. Four main mosaics are created with the built-up index Z_{bu} originated by the scenes belonging to the four different data collections corresponding to the epochs 1975, 1990, 2000, and 2014. Moreover, the mosaic of the degree of built-up μ_{bumix} was created for the epoch 2014 only, made of L8 data input. Figure 8 shows the aggregated results.

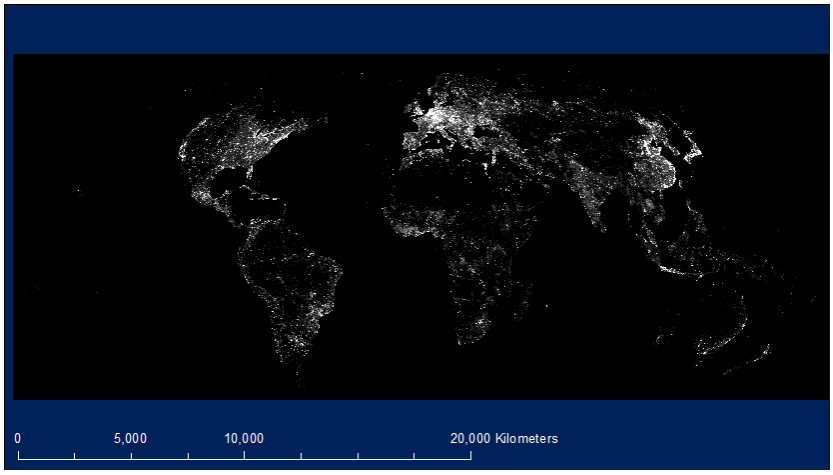


Figure 8: Global mosaic of the density of buildings as estimated by the μ_{bumix} model implemented in the experiment

Each epoch used 19,624 tiles covering the whole global landmass, each tile having a size of 4,096x4,096 elements, corresponding to 156.54 x 156.54 nominal square kilometers on the ground. Inside each specific collection, information gaps are filled with data from alternative scenes if available. If several valid information outputs were available in the same sample, a conservative composition rule based on average was adopted assuming to increase the consistency of the final Φ_E output.

6.2. Multi-temporal processing

The process is performed in parallel for each tile of the global mosaic. The aim of this phase is to fill remaining data gaps in the specific epochs, and reduce, if possible, error rates. Three main steps are performed i) determine final land/water surfaces ii) fill information gaps still remaining, iii) determine multitemporal information by cut-off of the Z_{bu} index collected in the different epochs.

At the first step a voting schema was applied. Let $M_{\Sigma} = M_{w1975} + M_{w1990} + M_{w2000} + M_{w2014}$ be the sum of the water masks collected in the four epochs. At the level of a spatial sample, each of these 4

water masks consists of a vector containing the assessments in the available scenes covering that specific sample. The possible assessments are water ("1"), non water ("0") and no data. Let M_{VD} be a layer containing the number of scenes from all the 4 epochs with valid data ("0" or "1" values). The final water mask has the value "1" in the samples in which the condition $M_{\Sigma} > 0.5 * M_{VD}$ is fulfilled.

The second step was solved by the creation of an ideal epoch 2014, made by filling the gaps present in the Z_{bu2014} with the most recent information available in the other epochs.

The third step was solved by a hierarchical process progressively going back in time and starting from the built-up areas as reported in the ideal epoch 2014. At each step, if strong evidences against the built-up hypothesis (strong negative Z_{bu} values) are collected in that epoch then these built-up areas are canceled becoming non-built-up. Similarly to [53], the model is applied under the assumption that built-up areas are monotonically spatially shrinking going back in time (respectively expanding from past to present). As a consequence, demolition and reconstruction dynamics in built-up areas creating oscillations in the Z_{bu} values over time are not assessed. While this assumption is a clear limitation of the model, it is considered true for the grand majority of the global built-up areas accounted in the final database. Furthermore, the assumption has the clear advantage to reduce significantly the computational cost and increase the statistical stability of the whole automatic change detection process. Figures 3f) and 4f) show the output of the procedure in the cities of Lagos (Nigeria) and Dallas (US). It is worth noting that in case of data gaps the above approach propagates the information from the most recent available data records. Figure 9 shows the final situation of temporal data availability after the multi-temporal processing. From left to right, the legend encodes the availability of data records (0 = unavailable, 1 = available) in the 2014, 2000, 1990, and 1975 collections.

The cut-off values determining the *strong evidences against the built-up hypothesis* are calculated dynamically for each tile and each epoch by observing first and second order statistics of the Z_{bu} samples falling in the spatial domain corresponding to the zero values of the low-resolution reference layer BU_{ref} (see details on this layer in section 3.3). The adaptive approach tries to minimize the effects on the Z_{bu} scores generated by the mutual interaction of several factors not explicitly modeled 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.

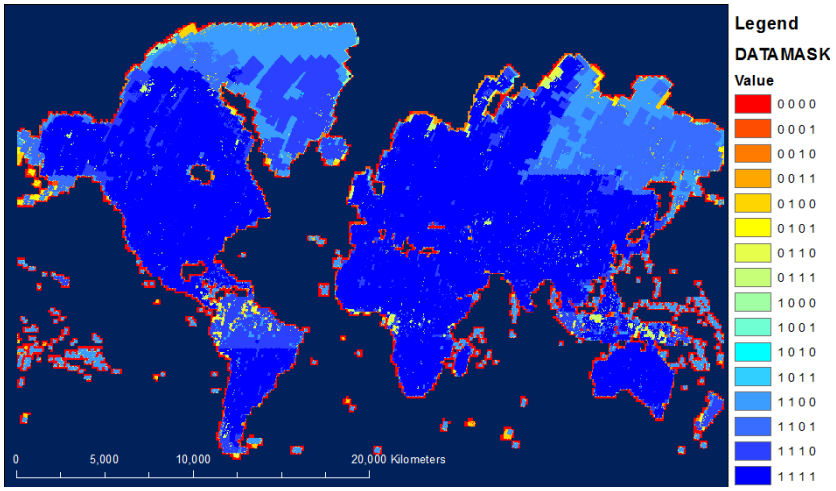


Figure 9: Global mosaic of the multi-temporal data availability mask as resulting of the mosaic process.

7. Results

This section summarizes the qualitative and quantitative assessments of the Landsat GHSL outputs in comparison to other available geo-information sources made by automatic satellite data classification. In addition, it includes also the presentation of some global and selected statistics.

7.1. Qualitative analysis

The qualitative assessments of the GHSL results in describing built-up areas generally show an increase of the spatial detail and absolute thematic accuracy as compared to other available global sources made by automatic satellite data classification. Figure 10 shows a typical behavior: from top to bottom reference topographic data 1:10K scale aggregated at 30m resolution, Landsat GHSL built-up areas detection, Meris GlobCover artificial surfaces, and MODIS urban layer. Omission error is still observable as compared to topographic mapping sources in detection of isolated built-up patches of 1-2 pixels. Furthermore, Landsat GHSL shows a tendency to saturate the information in built-up patches greater than circa 10 pixels (300 meters) as compared to reference topographic data. The phenomena is probably linked to the fact that mainly radiometric criteria are used as input for the classification and consequently the detection is strongly linked to the materials of the urban fabric more than to their morphology or texture. Therefore, a relatively high confusion between built-up (rooftops) and roads or other sealed surfaces placed around the building can be expected in this product.

The Φ_E confidence to the *built-up* class seems reasonably stable across sensor, seasonal, and geographical arbitrary conditions. Figures 3 and 4 show an example of Φ_E estimation in different years for cities of Lagos and Dallas, respectively. The most critical point seems to be the processing of the 1975 MSS data. Visual analysis by experts evidenced some cases of instability and low signal-noise ratio in the response of the classifier using this data input. The phenomena is geographical-related: it is linked to the specific contrast between the reflectance of the materials of the urban fabric and of the background. Typical examples of difficult situations are traditional settlement patterns made by natural materials (as mud or stones) and arid bushland background with large reflectance of the soil. Similarly, large presence of vegetation (especially large tree canopy) in both the settlement area and outside has been reported as creating instability in the classification. For this reason, it is advisable to check carefully the 1975 outputs before using it in the specific applications. While in many cities they provide a reliable information source, they are not considered sufficiently stable for global aggregated summaries.

Analogously to built-up baseline information, the visual inspection of the multiple-class land cover product shows a noticeable increase of detail respect to the spatial data used as training set. Figure 5 shows a typical setup in two examples: one area in the SW of Calcutta

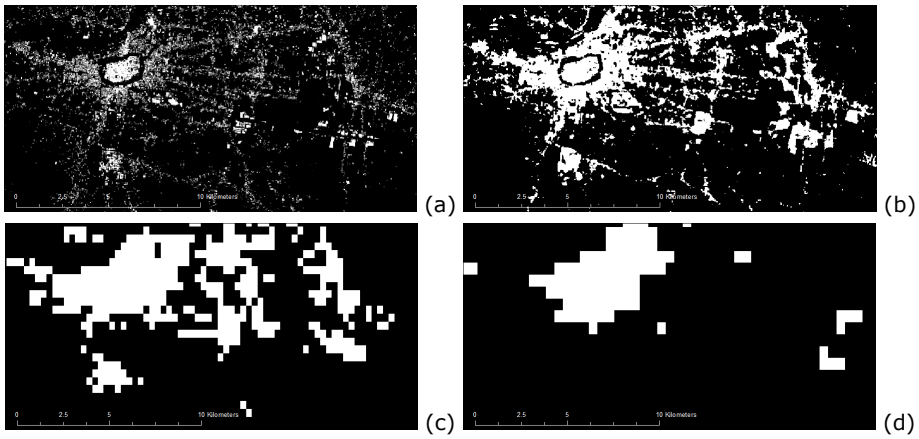


Figure 10: Town of Lucca, Italy. Comparison of the detail in the description of built-up areas available from different sources: a) topographic cartography at scale 1:10,000, b) Landsat GHSL c) Meris GlobCover, and d) MODIS urban layer.

(India) and an area in the surrounding of the Wanxian town (China). On the left the Meris GlobCover, on the right the output of the Landsat data classification.

Visual analysis of the global product showed that some classes are more stable than others. In particular, composite classes and classes incorporating a temporal component are apparently poorly performed by the classifier. They include the classes 4 Shrubs and Grassland; 5 Mosaic Croplands and Forest; 6 Rain Cropland; 7 Irrigated Cropland. On the other hand, the classes 10 Surface Water and 8 Forest are apparently very stable as performed by the classifier, together with the classes 2 Ice and permanent snow and 3 Bare soil and rocks.

Figure 6 shows some examples of the results of the classifier as regarding the built-up areas characterization. Output examples are shown in the cities of Chicago (US), Tokyo (Japan), and Johannesburg (South Africa): on the left the output of the classification, on the right the same fused with background high resolution imagery (Bing) for visual reference. Preliminary visual inspection by experts show promising results: the high-rise building areas of the cities seem to be correctly detected by the method implemented in the experiment, opening a new potential area of application for satellite-derived global digital elevation model products.

7.2. Quantitative accuracy assessment

In the experiment discussed here, the accuracy of the GHSL outputs are evaluated only for the built-up target class. Due to the availability of only recent cartographic reference, only the 2014 GHSL was assessed. The assessment is largely data-driven, in the sense that the scarcity of suitable reference data prevented the application of a statistical sampling procedure: as a consequence, the results cannot

be strictly considered as representative of the whole processed data universe. Geographically, the reference sets used in the experiment cover the entire Europe with a sampling schema and some arbitrary cities or regions where fine-scale digital cartographic data was available.

An absolute approach is adopted in order to be able to compare the performances of the Landsat GHSL outputs with other available global information sources, regarding the capacity to report about built-up areas in the available test sets. In this approach the reference data is maintained constant independently from the scale of the information product under test. The scale and reliability of the reference data is assumed to be better than any of the products under test. The adopted absolute approach does not adapt the reference and the target information definition to the characteristics of the sensor originating the information under test. The adaptation is typically performed by application of some kind of generalization procedure to the fine-scale reference data. Therefore, a more conservative assessment of the accuracy is expected in this test as compared to a more adaptive approach that is also common practice in remote sensing literature. For comparative purposes, the same performance statistics were calculated for GHSL and for other global or regional data sets, including the Landsat-based Corine2006, GlobalLand30, FROM-GLC, as well as the coarser resolution data sets MODIS-urban and GlobCover. In all cases, only the classes mostly related to the GHSL *built-up* were considered. As already introduced, the GHSL source used in the test is the output of the classification of the image data collection 2014.

7.2.1. Test with LUCAS reference data

The LUCAS data is described in section 3.4. The same data was used also for the validation of the European Settlement Map by [16] and hence allows a comparison with performance measures of this higher resolution layer. In the LUCAS test, a comparative analysis was implemented.

Table 4 shows the main results of the test. Following the considerations of [26], the reporting of quantitative accuracy metrics is based on the error matrix. The final performance statistics is reported with a set of quality measures. The total (or overall accuracy) represents the share of all correctly classified sample units of the entire error matrix. The Kappa analysis, introduced by [9], presents a discrete multivariate technique commonly used in accuracy assessment to statistically test if classifications are significantly different from the reference data, being a chance corrected renormalized form of accuracy. This very popular measure has nevertheless one disadvantageous characteristic especially in our case of imbalanced datasets: it is affected by prevalence. Consequently, we also use other two quality measures that are skew-insensitive, thus controlling better inflated performance estimates on imbalanced datasets. The first one is the balanced accuracy (BA) from [7] and it corresponds to the

Area Under the Curve from the ROC analysis in the case of a single parametrization. The second one is the informedness (named like this by [45] but introduced for the first time by [35]), which can be viewed as a chance corrected unbiased renormalized form of recall (True Positive Rate). The measure can be directly quantifiable as the probability that an informed decision is made rather than a random guess (chance), as stated by [46].

Layer	Total Accuracy	Kappa	Balanced Accuracy	Informedness
GHSL	0.9628	0.3233	0.7767	0.5534
Corine2006	0.9548	0.2701	0.7606	0.5212
GlobalLand30	0.9488	0.2692	0.7938	0.5877
GlobCover	0.9712	0.2072	0.6173	0.2346
FROM-GLC	0.9576	0.1843	0.6493	0.2986
MODIS	0.9624	0.1581	0.6099	0.2198

Table 4: Performance statistics for GHSL (2014) and other datasets compared to LUCAS (2012).

According to the LUCAS test in Europe, the total accuracy (TA) is very high for all data sets, well above 95% (Table 4). However, in this imbalanced two-class data set (built-up and non-built-up), the built-up accounts only for a small portion of the area (< 10%). Theoretically, just by classifying all areas as non-built-up, a total accuracy above 90% could be obtained. The other statistics show a more differentiated picture. The Kappa statistics is quite low for all data sets. The Kappa value for GHSL (0.3233) can be classified as 'fair' according to the proposal of [27]: nevertheless, is the best value obtained in the experiment with the worst being the performance of MODIS showing a Kappa value of 0.15. Kappa, BA and Informedness clearly show two groups of performances in the products under test. The GHSL, GlobalLand30 and the Corine2006 layers perform very similar with scores in Kappa, BA and Informedness greater to 0.27, 0.76, and 0.52, in that order. In this group, GHSL is top ranking in Kappa (0.3233), while GlobalLand30 is top ranking in both BA (0.7938) and Informedness (0.5877). In this group of top products, only GHSL has been produced with a highly automated procedure. The quality drops noteworthy for GlobCover, FROM-GLC and MODIS that instead are generated by automatic classification.

7.2.2. Test with fine-scale cartography

The reference test set consists of 1505 raster tiles with a surface of 10x10 km^2 each: they are derived from available vector cartographic sources at scale of approx. 1:10,000. Details of the test preparatory phase are included in section 3.4. The Table 5 shows the main results of the assessment comparing the GHSL (2014 collection) and other available global products with fine scale cartography. The average and standard deviation of the accuracy measures calculated per each tile are reported. The results are almost consistent with the results of the LUCAS test in Europe but in this case GHSL is top ranking in all

the accuracy performances showing average TA, BA, Informedness, and Kappa of 0.8996, 0.6718, 0.3435 and 0.3327, respectively.

Average	Total Accuracy	Balanced Accuracy	Informedness	Kappa	ComErr	OmErr
GHSL2014	0.8996	0.6718	0.3435	0.3327	0.5344	0.5414
ESA GC 2.3	0.8827	0.5368	0.0738	0.0837	0.5722	0.8701
FROM-GLC	0.8648	0.5524	0.1101	0.1054	0.6679	0.7506
GlobalLand30	0.8815	0.6495	0.2996	0.2678	0.6056	0.5627
MODIS	0.8610	0.5459	0.0919	0.0856	0.6817	0.7619
Standard Deviation						
GHSL2014	0.0964	0.1120	0.2241	0.1992	0.2215	0.3096
ESA GC 2.3	0.1278	0.0560	0.1122	0.1178	0.2302	0.2125
FROM-GLC	0.1429	0.0704	0.1424	0.1249	0.2361	0.3225
GlobalLand30	0.1146	0.1134	0.2267	0.1958	0.2026	0.3406
MODIS	0.1461	0.0688	0.1376	0.1220	0.2023	0.3185

Table 5: Performance statistics for GHSL (2014) and other datasets compared to fine scale building footprints.

Some of the applications of the GHSL require to estimate the absolute amount of built-up area per administrative units. It is thus interesting to test the capacity of the GHSL output to predict the real amount of built-up, being the information aggregated in spatial units greater than the Landsat pixel size. Figure 11 shows the results of a linear regression analysis between the per-tile totals of built-up surface in the GHSL outputs (X axis) and the built-up surface as estimated by the reference cartographic source in the same tile (Y axes). By setting the intercept to zero, the linear equation translating GHSL *built-up* surfaces to building footprints surfaces is estimated as $Y = 0.2164 \times X$. The coefficient of determination R^2 of the regression is equal to 0.9038: therefore, the automatic GHSL output seems to be a good estimator of the built-up surface as reported by the reference data. Both the input data characteristics (spatial resolution) and the information extraction methodology are factors strongly influencing the slope or gain coefficient of the equation: they are constant in the global processing experiment discussed here. Subsequently, according to these results the GHSL output seems to be a good candidate for estimation of the global built-up areas figures.

The same analysis can be performed using the sub-pixel estimation of the degree of built-up μ_{bumix} for the epoch 2014 as defined in the chapter 5.2. In this case, the linear equation translating GHSL μ_{bumix} surfaces to building footprints surfaces is estimated as $Y = 0.2267 \times X$, with the coefficient of determination R^2 of the regression equal to 0.9143. Both the gain and the R^2 parameters of the linear regression are slightly increasing. These results show that the GHSL μ_{bumix} model provides a better approximation of the built-up surfaces estimated by the fine-scale building footprints sources, as compared to the standard GHSL classification output. These are very preliminary results: further work must be done for systematic assessment of the parameters driving the equation 26 of the model.

Figure 12 shows the results of the analysis of the percent com-

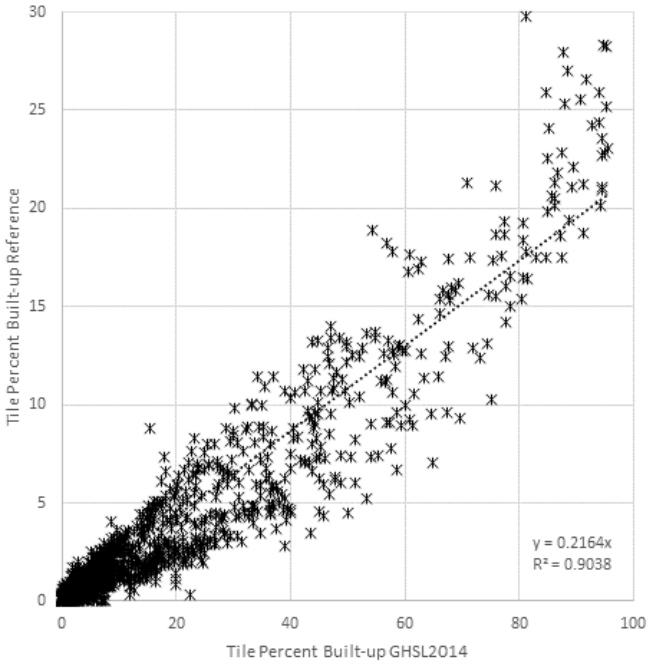


Figure 11: Comparison of BU per tile in the GHSL layer and in the reference.

mission and omission errors of the GHSL *built-up* class, respectively, using the building footprints as reference. The built-up percent error per tile (Y axis) is plotted together with the density of built-up as estimated by the reference data (X axis). Both commission and omission errors are very unstable in low-density tiles showing high variability of results in settlement patterns with built-up density < 5%: this is due to the presence of a random constant error in the GHSL output and the relative nature (percent respect to the total built-up surface) of the errors rates as calculated in this assessment. More interestingly, it can be observed that the commission error stabilizes at building densities greater than 5% with an average built-up relative commission rate of 0.34. This value is probably linked to the same factors influencing the systematic gain coefficient observed in the linear regression analysis previously discussed. Moreover, it is interesting to observe that the GHSL relative omission error rates are strongly (inversely) correlated with the density of the settlement patterns to be detected. As a consequence, according to these preliminary results the built-up surfaces associated to rural and scattered low-density settlement patterns are expected to be relatively under-represented in the GHSL output evaluated here. The capacity to discriminate such spatially scattered information is strongly linked to the resolution power of the input sensor: therefore, large margins of improvement are expected to be achieved from the adoption of

higher resolution data input for the production of the GHSL, as in a possible next GHSL option including Sentinel data input.

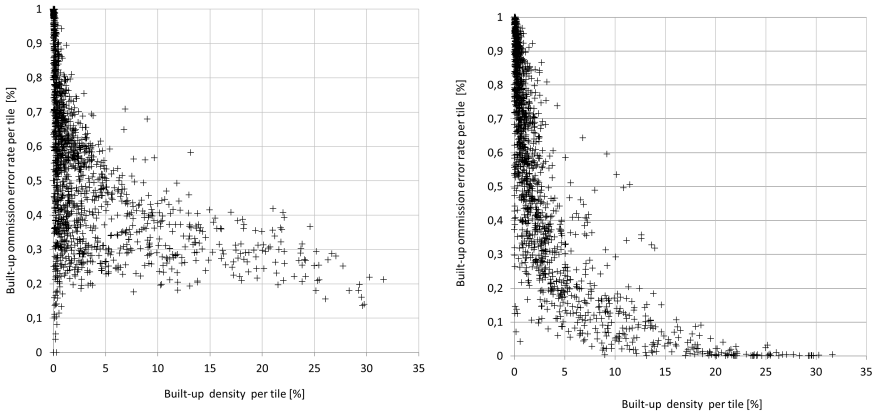


Figure 12: Relative built-up commission (left) and omission (right) error rate as a function of built-up density in the sample tile.

The GHSL results discussed here and the FROM-GLC share the same type of data input, the same scale and the same aim of automatic control of the production procedure. As reported in [21], the FROM-GLC experiment adopted the state-of-the-art classification techniques in the remote sensing community including the conventional maximum likelihood classifier (MLC), J4.8 decision tree classifier, Random Forest (RF) classifier and support vector machine (SVM) classifier. Thus FROM-GLC is a good candidate for a comparative analysis assessing the effects on accuracy of the new classifier implemented for the GHSL production. According to the preliminary evidences collected in the experiment, the introduction of the new symbolic machine learning (SML) approach and associative analysis techniques makes a significant step forward in the accuracy of the derived automatic products. From the Table 5, an increase of 0.0348 and 0.1194 of the mean TA and BA, respectively, are expected as consequence of the adoption of the new SML techniques for the GHSL production. Even more considerable are the estimated effects in the Informedness and Kappa measurements that show average values three times higher than in the case of FROM-GLC. As regarding the built-up commission and omission error percent rates, an average reduction of 0.1335 and 0.2092 are expected, in that order, as repercussion of the adoption of the new classifier. On the overall, these results are a confirmation of the findings collected in [43] during the experimental design phase of the new classifier.

7.3. Cross comparison

This section is dedicated to the results of the cross-comparison tests undertaken on three different datasets that show some similarity with

the GHSL Landsat in scale and thematic contents, but cannot be considered as absolute reference for validation.

7.3.1. ESM

In this test, the GHSL outputs created from the Landsat data of the epoch 2014 are compared with the European Settlement Map (ESM) output. The objective of the study is to understand the relations between the two information outputs, which have differences in the sensor, in the scale, and in the image features used as input of the classification. In particular, a use scenario linked to policy support indicators is taken. From this perspective, the satellite-derived indicators are typically aggregated to spatial units corresponding to administrative boundaries, that are linked to specific decision making processes. In the test described here, the information is aggregated in 1342 regions corresponding to the level 3 of the Nomenclature of Territorial Units for Statistics (NUTS) defined by the European Union statistical office (EUROSTAT)[17]. Figure 13 shows the linear regression between the total of built-up estimated by the spot data (X axis) and the same information as estimated by the Landsat GHSL of the epoch 2014 (Y axis). According to these results a strong linear correlation between the two data sets can be observed. A coefficient of determination R^2 equal to 0.88 is obtained by a linear regression setting the intercept to zero. Moreover, the Landsat-made outputs seems to detect systematically more built-up surfaces than the outputs made by using textural and morphological criteria on high resolution Spot data. The gain of the linear regression equation is estimated as 1.92 for the whole data processed. Most presumably, this gain value is the resultant in the trade-off of two main factors originated by the different sensor and method characteristics: i) the decrease of the sensitivity (of Landsat vs Spot) in detection of scattered built-up structures and ii) the inflation of built-up surface totals (of Landsat vs Spot) in compact built-up areas. In the latter case, the Spot-generated built-up information differentiates between the buildings and roads or open spaces due to the higher sensor resolution (2.5 meters). This is done thanks to the possibility to exploit connected-component-based, morphological descriptors of the image. In the 30m resolution input Landsat data this is not applicable, and the same surfaces are aggregated to the built-up area class.

7.3.2. NYU Urbanization Project

In this subsection, the results of the cross-comparison between the Landsat GHSL outputs and the baseline data from the Urbanization Project of the New York University (NYU) [2] are presented. The main objective of the test was to collect some preliminary information about the behavior of the GHSL outputs produced by the experiment in the epochs 1990 and 2000, where no other reference data was available.

[17] http://ec.europa.eu/eurostat/web/nuts/overview

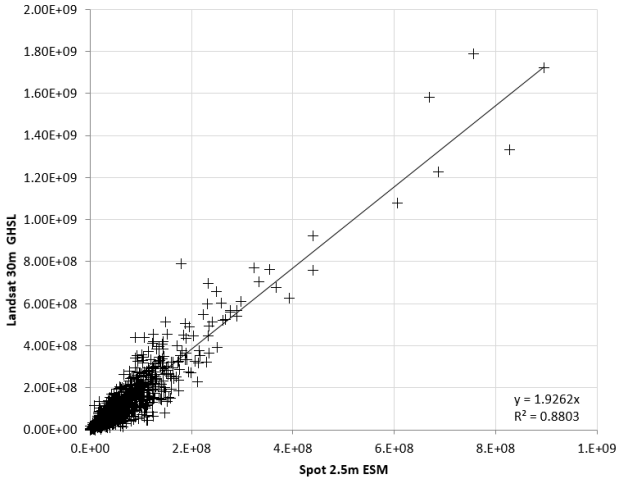


Figure 13: Estimated built-up surfaces in the ESM (2012) and GHSL (2014) outputs

The test reported here uses in input the $Z_{bu\beta}$ membership to the class *built-up* estimated by the GHSL workflow for each scene of each collection (see section 4.3 for details). The GHSL built-up mask used for calculation of the agreement with the NYU classified data is calculated with a neutral cutoff as $M = Z_{bu\beta} \geq 0$. A subset of 152 Landsat satellite scenes was selected for the test from the whole NYU data, by application of the following conditions: i) selection of the epochs 1990 and 2000, and ii) maximum 12 months of difference between the data collection time of the GHSL and the NYU scenes. The aim of the test is to control the consistency between the two outputs in the epochs 1990 and 2000 that have no other reference available in the GHSL assessment.

Dataset type	Epoch	Count of ID	Average	Stdv
Complete	1990	79	0.943	0.04923
	2000	73	0.933	0.04708
	Grand Total	152	0.939	0.04830
Excluding the worst 10 sample cases	1990	75	0.951	0.03248
	2000	67	0.943	0.03005
	Grand Total	142	0.947	0.03153

Table 6: Agreement of GHSL and NYU by GHSL epoch (total accuracy by scene - middle point ENDI threshold ("0" value))

Table 6 shows the agreement between the GHSL and the NYU classification outputs measured as total accuracy using the NYU data as reference. In the assessment a per-pixel approach is taken, and the GHSL data is reprojected in the NYU data that is maintained constant. In the 1990 and 2000 epochs, an average agreement of 0.943 ± 0.04 and 0.933 ± 0.04 is estimated, correspondingly, producing an overall

agreement of 0.939 ± 0.04. If the ten worst sample cases are excluded from the assessment, the overall agreement increases to 0.951 ± 0.03 and 0.943 ± 0.03 for the 1990 and 2000 epochs, in that order. It is worth noting that the two sources don't share the same geocoding (ortho-rectification) process and as a consequence no sub-pixel spatial consistency can be expected. This contributes to decrease the expected thematic agreement as measured by the per-pixel assessment. According to these results the thematic agreement between the two sources seems to be generally high, with very few exceptions that should be analyzed case by case and are not reported here.

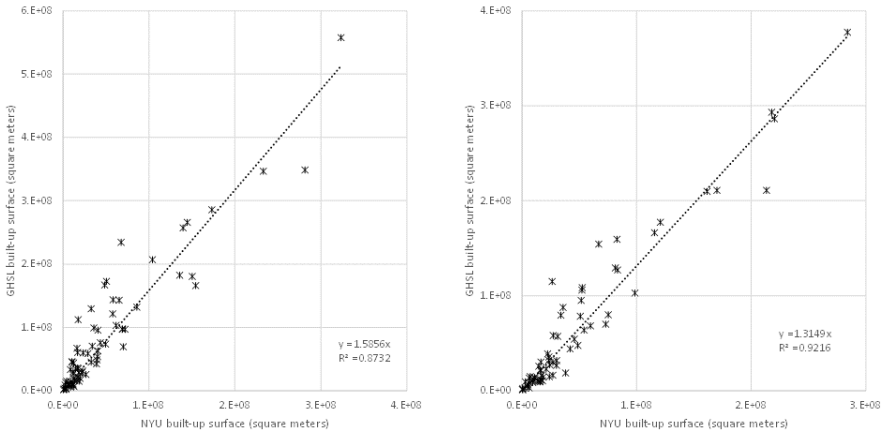


Figure 14: Estimated built-up surfaces in the NYU and GHSL outputs, (left) epoch 1990, (right) epoch 2000

Figure 14 shows the results of a linear regression analysis between the per-scene total built-up surfaces estimated by the NYU data (X axis) and the built-up surfaces estimated by the GHSL (Y axis), for the two epochs 1990 and 2000, respectively. A strong linear correlation can be observed in both 1990 and 2000 epochs, showing a coefficient of determination R^2 equal to 0.87 and 0.92, respectively. The 1990 epoch shows a relatively weaker correlation. The GHSL outputs seems to detect systematically more built-up surfaces than the NYU baseline. The gain of the linear regression equation is estimated as 1.58 and 1.31 for the 1990 and 2000 epochs, respectively.

7.3.3. GeoNames

A global omission rate test was done using the populated place names as derived from the GeoNames project. The populated place dataset has been cleaned to remove the spatially overlapping features or features whose spatial precision was assessed as low during the data ingestion phase. The resulted Geonames dataset consists in more than 2.5 million populated place points. No filtering was applied based on the population size of the settlement. Consequently all kind of settlements were included. The test was done by measuring how many

populated places points receive a *not-built-up* class label from the satellite-derived assessment, and comparing it with the total number of points in the reference data base. There are no information about the global completeness of the specific source: consequently no commission test can be done. Nevertheless, for some applications including population spatial modeling the omission ratio is already an interesting quality measurement. For a total of $2,665,314$ populated places included in the global assessment, a total of $1,263,944$; $133,580$, and $187,343$ were labeled as *built-up*, *urban* or *artificial surfaces* by GHLS2014, MODIS, and FROM-GLC source, in that order. This provides an estimated omission ratio of 0.52578, 0.94988, and 0.92971 of the GHSL2014, MODIS, and FROM-GLC outputs, respectively. The results show a clear increase of the precision in detecting populated places using automated satellite classification processes as regarding GHSL2014 in comparison with MODIS and FROM-GLC. While for MODIS a scale and generalization issue can play a role, the FROM-GLC data is at the same scale of the GHSL2014. On the other side, the test shows the challenges that still need to be addressed for the complete mapping of the whole populated places of the globe using satellite data input. This can be an example of composite indicators including satellite-derived information and potentially supporting the monitoring of SDGs.

7.4. Aggregated results

In this section, some global and selected statistics are presented that relate the built-up areas given by Landsat GHSL with some social-economic variables (as population level and Gross Domestic Product (GDP) per capita).

7.4.1. Global statistics

Figure 15 shows the global trends on built-up areas and population as assessed by the experiment using Landsat data input. The black points are derived from the GHSL assessment of the epochs 1990, 2000, and 2014. The global built-up area assessment of the epoch 1975 is considered underestimating the built-up areas as detected in the successive epochs because of two main factors: i) a worse sensor characteristics, and ii) the presence of large data gaps (no data available) in the 1975 collection. According to this assumption, only the data of the epochs 1990, 2000, and 2014 have been used for estimating the parameters of the curve linking global population and built-up areas: a polynomial 2-degree curve fitting was used in the example. Given the trends learned in the 1990, 2000, and 2014 epochs, the built-up areas of 1975 have been estimated using the global population in that year, while the built-up areas in 2050 have been extrapolated by using the value of 9.6 billion of people projected for 2050 by UN [18]. In the Figure 15 the difference between the global

[18] World Population Prospects: 2015 Revision http://www.un.org/en/development/desa/population/events/other/10/index.shtml

built-up areas detected and estimated in the 1975 collection is noted as *sensor gap* that is quantified to be the 23.78% of global built-up surface as estimated for this epoch. The Table 7 reports about the final assessments of population and built-up surfaces done during the experiment. The global population figures in column (2) are derived from the World Bank (WB) source. Column (3) shows the total of built-up surface accounted by the Landsat data processing. Column (4) displays the built-up surface as estimated by the polynomial curve fitting on the epochs 1990, 2000, and 2014. Column (5) reports about the average built-up surface per capita as estimated in the four epochs. The surfaces reported in the columns (3,4,5) are expressed at Landsat scale. The columns (6) and (7) report about the global built-up surfaces and built-up per capita expressed at the scale of building footprints with a generalization of $1:10,000$. These numbers have been calculated by application of the corrector coefficient equal to 0.2164 to the Landsat surfaces, as estimated by the regression analysis reported in the section 7.2.2.

Year	WB Population [persons]	GHSL Landsat BU [m^2]	estimated Landsat BU [m^2]	estimated Landsat BU [m^2 per person]	estimated C10K BU [m^2]	estimated C10K BU [m^2 per person]
1975	4.03946E+09	3.08779E+11	4.05165E+11	100.30	8.76777E+10	21.71
1990	5.25428E+09	5.32197E+11	5.32197E+11	101.29	1.15167E+11	21.92
2000	6.07504E+09	6.29508E+11	6.29508E+11	103.62	1.36225E+11	22.42
2014	7.09653E+09	7.74530E+11	7.74530E+11	109.14	1.67608E+11	23.62
2050	9.60000E+09		1.29988E+12	135.40	2.81293E+11	29.30

Table 7: Global assessments of population and built-up surfaces

7.4.2. Selected Statistics

Figure 16 shows different development patterns of some top GDP ranking countries as they can be observed by the data extracted during the experiment. On the horizontal axis, there are the population counts for the years 1990, 2000, and 2014 (source: WB). Correspondingly, in the vertical axis there is the estimated Landsat built-up surface per capita as estimated by the experiment. Vertical patterns are originated by built-up area expansion without comparable population growth: it can be associated to mature economies. According to these empirical results this is the pattern of France, Germany, Italy, Japan, and Russia. UK is a particular case of this cluster showing an almost stable behavior in both built-up area per capita and population. A completely different development pattern can be noticed in another cluster including Brazil, India, and China. In these cases a dominant horizontal or slightly positive sloped behavior can be observed. It can be associated to recent dynamical economies, with significant population growth rates and increasing average housing standards expanding the built-up surface per capita. US is standing apart in the top left of the picture, showing the highest built-up

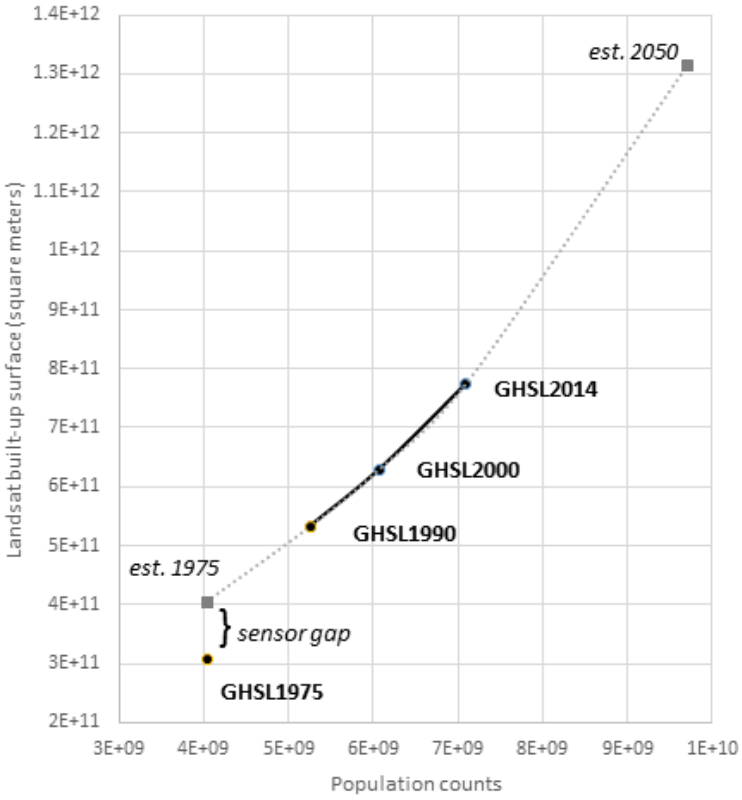


Figure 15: General trends on global built-up areas as measured during the experiment

surface per capita associated to significant population growing rates conditions.

Figure 17 shows an example of possible integration between social-economical variables and satellite-derived measurements. The countries with population greater than 30 millions of inhabitant have been selected in the example. On the vertical axis and the horizontal axis, correspondingly, the GDP per capita (source: WB 2013), and the built-up area per capita (as derived from GHSL epoch 2014) are represented. A clear correlation between the two measures can be observed, with interesting deviations. For example, Ukraine and South Africa display as outliers showing much higher built-up surface per capita than what could be expected by their GDP. In Ukraine, the explanatory factors can be related to large emigration rates producing abandoned built-up structures, and consequently increasing anomalously the average built-up surface per capita. In South Africa, the anomaly in the model can be interpreted as the impact of governmental social settlement programs dedicated to low-income population and increasing the average housing standards.

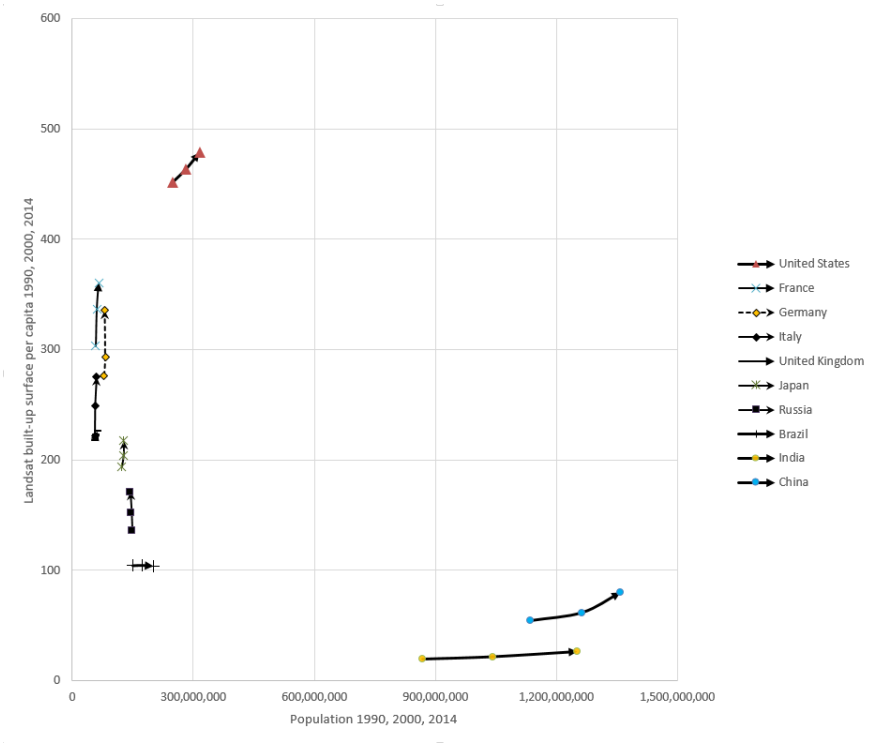


Figure 16: Patterns of development trends in the top 10 Countries as GDP (WB 2013). In the horizontal axis the population counts, while vertical axis the Landsat built-up surface per capita. Years 1990,2000 and 2014 are assessed from the corresponding GHSL 30m resolution data collections

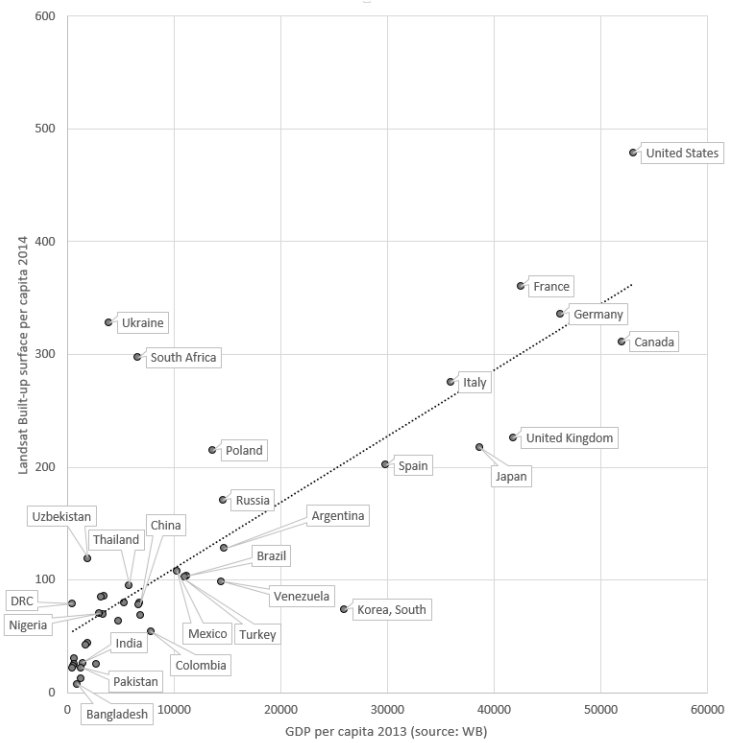


Figure 17: Relation between GDP per capita (X axis) and Landsat built-up surface per capita (Y axis). The GDP source is WB 2013. The built-up surface per capita 2014 was assessed from the GHSL 30-m-res data collection and population data from the WB source

8. Conclusions

A new global human settlement map was extracted from Landsat satellite data and is referred to as Global Human Settlement Layer (GHSL). The Landsat GHSL is the most spatially global detailed data available today dedicated to human settlements, and it shows the greatest temporal depth. More generally, the GHSL includes several innovative features shortly listed below: i) it largely relies on a new classification method based on symbolic machine learning, ii) it is the first known attempt to classify global MSS data for any use, and iii) it is the first known attempt to derive global settlement volume using digital elevation model data.

The first assessments of the results seem to be encouraging and showing an improvement in terms of spatial detail and thematic contents, respect to the open and free information available today at global level. Some examples of use of the new information as baseline for the design of new development indicators are provided showing a promising new area of application of remote sensing data analysis. Only the GHSL core information about the *built-up* class was quantitatively assessed in the experimental setup presented here. The accuracy of the GHSL in detection of built-up areas derived from building footprints at scale 1:10.000 was always best ranking respect to available global alternatives made by automatic satellite data classification approaches. Cross comparison with the historical spatial information (epochs 1990, 2000) produced in an extensive stratified sample of global cities showed high degree of agreement with the automatic GHSL multi-multitemporal output. The core methodology for image information extraction developed in the experiment is designed to be scalable to massive data scenarios requiring distributed parallel processing infrastructures. The design of the core classification method is pixel-based and is well compatible with the multi-scale image data cubes as proposed by some advanced satellite data processing dedicated infrastructures.

According to the results collected here, a substantially positive answer can be provided to the first two basic questions reported in the Introduction and placed in the background of the experiment: i) the possibility to increase the detail of the available global information layers describing human settlements and using open satellite data input and ii) the possibility to improve the automatic control (repeatability, cost sustainability) of the information production process.

Nevertheless, the information products described in this manuscript remain mostly experimental. From an operational use perspective, some main challenges must be addressed. The main ones are listed below:

1. The insufficient availability of global test sets with the right scale, temporal attribute and reliability of the thematic information, that can be suitable for a statistically correct validation and as support for the improvement of the product;
2. The insufficient availability of global multi-sensor comparative

studies focused on human settlement and allowing to understand the systemic inter-sensor bias and gain functions linking the information results obtained from different input imagery.

3. The insufficient availability of image data in the 1975 collection processed in the experiment (data gaps, stability of results), despite the strategic importance of the information of the epoch 1975 for understanding global trends;

4. The unavailability of global digital elevation models having a temporal dimension, thus allowing the measurement of volumetric changes in the cities at the scale tested in the experiment.

These critical points will design the road map of the priority areas for improvement of the GHSL information with the help of the remote sensing community.

References

[1] M. Anandhavalli, M. K. Ghose, and M. Gauthaman. Association Rule Mining in Genomics. *International Journal of Computer Theory and Engineering*, 2(2):1793–8201, April 2010.

[2] S. Angel, A. M. Blei, D. M. Civco, P. Lamson-Hall, J. Parent, N. Galarza Sanchez, and K. Thom. *Atlas of Urban Expansion - The 2015 Edition*. 2015. forthcoming 2015.

[3] S. Angel, S. Sheppard, D. L. Civco, R. Buckley, A. Chabaeva, L. Gitlin, A. Kraley, J. Parent, and M. Perlin. *The dynamics of global urban expansion*. World Bank, Transport and Urban Development Department Washington, DC, 2005.

[4] Y. Ban, P. Gong, and C. Giri. Global land cover mapping using earth observation satellite data: Recent progresses and challenges. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103(0):1–6, Feb 2015.

[5] C. Becquet, S. Blachon, B. Jeudy, J.-F. Boulicaut, and O. Gandrillon. Strong-association-rule mining for large-scale gene-expression data analysis: a case study on human SAGE data. *Genome Biology*, 3(12):1–16, 2002.

[6] S. Bontemps, P. Defourny, E. V. Bogaert, O. Arino, V. Kalogirou, and J. R. Perez. GLOBCOVER 2009-Products Description and Validation Report. *ESA Technical Report*, 2011.

[7] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann. The balanced accuracy and its posterior distribution. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 3121–3124. IEEE, 2010.

[8] J. Chen, J. Chen, A. Liao, X. Cao, L. Chen, X. Chen, C. He, G. Han, S. Peng, M. Lu, W. Zhang, X. Tong, and J. Mills. Global land cover mapping at 30m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103(0):7 – 27, Oct 2014.

[9] J. Cohen. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960.

[10] C. Creighton and S. Hanash. Mining gene expression databases for association rules. *Bioinformatics*, 19(1):79–86, 2003.

[11] J. Crespo, J. Serra, and R. W. Schafer. Theoretical aspects of morphological filters by reconstruction. *Signal Processing*, 47(2):201–225, Nov 1995.

[12] Y. Demchenko, C. Ngo, and P. Membrey. Architecture Framework and Components for the Big Data Ecosystem Draft Version 0.2. SNE technical report SNE-UVA-2013-02, System and Network Engineering, Sept 2013.

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

[14] EUROSTAT. LUCAS - Land use and land cover survey. online, 2015. [Accessed: 03 August 2015].

[15] S. Ferri, V. Syrris, A. Florczyk, M. Scavazzon, M. Halkia, and M. Pesaresi. A new map of the European settlements by automatic classification of 2.5m resolution SPOT data. In *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*, pages 1160–1163, July 2014.

[16] A. J. Florczyk, S. Ferri, V. Syrris, T. Kemper, M. Halkia, P. Soille, and M. Pesaresi. A New European Settlement Map From Optical Remotely Sensed Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015. accepted.

[17] P. Gamba, F. Dell'Acqua, and B. Houshmand. SRTM Data Characterization in Urban Areas. *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, 34(3/B):55–58, 2002.

[18] E. Georgii, L. Richter, U. Rückert, and S. Kramer. Analyzing microarray data using quantitative association rules. *Bioinformatics*, 21(Suppl 2):ii123–ii129, 2005.

[19] P. Gong and P. J. Howarth. The Use of Structural Information for Improving Land-Coverover Classification Accuracies at the Rural-Urban Fringe. *Photogrammetric Engineering and Remote Sensing*, 56(1):67–73, January 1990.

[20] P. Gong and P. J. Howarth. Land-use classification of SPOT HRV data using a cover-frequency method. *International Journal of Remote Sensing*, 13(8):1459–1471, 1992.

[21] P. Gong, J. Wang, L. Yu, Y. Zhao, Y. Zhao, L. Liang, Z. Niu, X. Huang, H. Fu, S. Liu, et al. 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, 2013.

[22] L. Gueguen, P. Soille, and M. Pesaresi. A new built-up presence index based on density of corners. In *Proc. Int. Symp. on Geoscience and Remote Sensing (IGARSS)*, Munich, 2012.

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

[24] T. Kemper, X. Blaes, D. Ehrlich, F. Haag, and M. Pesaresi. On the feasibility to map the settlements of Brazil with the CBERS-2B satellite. In *Urban Remote Sensing Event (JURSE), 2013 Joint*, pages 078–082, April 2013.

[25] T. Kemper, N. Mudau, P. Mangara, and M. Pesaresi. Towards an automated monitoring of human settlements in South Africa using high resolution SPOT satellite imagery. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-7/W3:1389–1394, 2015.

[26] M. Klotz, T. Kemper, C. Geiß, T. Esch, and H. Taubenböck. Mapping spatial settlement patterns on a global scale: Multi-scale cross-comparison of new and existing global urban maps. *Remote Sensing of Environment*, 2015. Submitted.

[27] J. R. Landis and G. G. Koch. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159–174, 1977.

[28] L. Lu, H. Guo, M. Pesaresi, P. Soille, and S. Ferri. Automatic recognition of built-up areas in China using CBERS-2B HR data. In *Urban Remote Sensing Event (JURSE), 2013 Joint*, pages 065–068, 2013.

[29] M. Marconcini, T. Esch, A. Felbier, and W. Heldens. Unsupervised high-resolution global monitoring of urban settlements. In *Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International*, pages 4241–4244, July 2013.

[30] B. L. Markham and D. L. Helder. Forty-year calibrated record of earth-reflected radiance from Landsat: A review. *Remote Sensing of Environment*, 122:30–40, 2012. Landsat Legacy Special Issue.

[31] V. Mayer-Schönberger and K. Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt, 2013.

[32] C. M. Mertes, A. Schneider, D. Sulla-Menashe, A. J. Tatem, and B. Tan. Detecting change in urban areas at continental sscale with MODIS data. *Remote Sensing of Environment*, 158:331–347, 2015.

[33] H. Miyazaki, X. Shao, K. Iwao, and R. Shibasaki. An automated method for global urban area mapping by integrating ASTER satellite images and GIS data. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 6(2):1004–1019, 2013.

[34] B. Pandey, P. Joshi, and K. C. Seto. Monitoring urbanization dynamics in India using DMSP/OLS night time lights and SPOT-VGT data. *International Journal of Applied Earth Observation and Geoinformation*, 23:49 – 61, 2013.

[35] C. S. Peirce. The numerical measure of the success of predictions. *Science*, ns-4(93):453–454, 1884.

[36] M. Pesaresi. Texture analysis for urban pattern recognition using fine-resolution panchromatic satellite imagery. *Geographical and Environmental Modelling*, 4(1):43–63, 2000.

[37] M. Pesaresi. Global fine-scale information layers: the need of a paradigm shift. In P. Soille and P. G. Marchetti, editors, *Conference on Big Data from Space (BiDS'14)*. JRC, 2014.

[38] M. Pesaresi and J. Benediktsson. A new approach for the morphological segmentation of high-resolution satellite imagery. *Geoscience and Remote Sensing, IEEE Transactions on*, 39(2):309 –320, Feb. 2001.

[39] M. Pesaresi and S. Freire. BUREF producing a global reference layer of built-up by integrating population and remote sensing data. Backgroud paper, Global Assessment Report on Disaster Risk Reduction 2015, 2014.

[40] M. Pesaresi, A. Gerhardinger, and F. Kayitakire. A robust built-up area presence index by anisotropic rotation-invariant textural measure. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 1(3):180–192, Sept. 2008.

[41] M. Pesaresi, G. Huadong, X. Blaes, D. Ehrlich, S. Ferri, L. Gueguen, M. Halkia, M. Kauffmann, T. Kemper, L. Lu, M. Marin-Herrera, G. Ouzounis, M. Scavazzon, P. Soille, V. Syrris, and L. Zanchetta. A Global Human Settlement Layer From Optical HR/VHR RS Data: Concept and First Results. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 6(5):2102–2131, Oct 2013.

[42] 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. In S. S. Shen and P. E. Lewis, editors, *Proceedings of SPIE 8390, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII, 839025*, volume 8390, Baltimore, USA, April 2012.

[43] M. Pesaresi, V. Syrris, and A. Julea. Benchmarking of the Symbolic Machine Learning classifier with state of the art image classification methods - application to remote sensing imagery. JRC Technical Report EUR 27518 EN, European Commission, Joint Research Centre, Institute for the Protection and Security of the Citizen, 2015.

[44] D. Potere, A. Schneider, S. Angel, and D. L. Civco. Mapping urban areas on a global scale: which of the eight maps now available is more accurate? *International Journal of Remote Sensing*, 30(24):6531–6558, 2009.

[45] D. M. W. Powers. Recall & Precision versus The Bookmaker. In *International Conference on Cognitive Science*, pages 529–534, 2003.

[46] D. M. W. Powers. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *International Journal of Machine Learning Technology*, 2(1):37–63, 2011.

[47] P. Salembier and J. Serra. Flat Zones Filtering, Connected Operators, and Filters by Reconstruction. *IEEE Transactions on Image Processing*, 4(8):1153–1160, 1995.

[48] A. Schneider, M. Friedl, and D. Potere. Mapping global urban areas using MODIS 500-m data: New methods and datasets based on 'urban ecoregions'. *Remote Sensing of Environment*, 114(8):1733–1746, Aug. 2010.

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

[50] C. Small. High spatial resolution spectral mixture analysis of urban reflectance. *Remote Sensing of Environment*, 88(1-2):170–186, 2003.

[51] C. Small. A global analysis of urban reflectance. *International Journal of Remote Sensing*, 26(4):661–681, 2005.

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

[53] H. Taubenböck, T. Esch, A. Felbier, M. Wiesner, A. Roth, and S. Dech. Monitoring urbanization in mega cities from space. *Remote Sensing of Environment*, 117:162–176, 2012.

[54] T. van Zyl. *Big Data: Techniques and Technologies in Geoinformatics*, chapter 7 Machine Learning on Geospatial Big Data, pages 133–148. CRC Press, 2014.

[55] M. A. Wulder and N. C. Coops. Satellites: Make earth observations open access. *Nature*, 513(7516):30–31, 2014.

[56] L. Yu, J. Wang, X. Li, C. Li, Y. Zhao, and P. Gong. A multi-resolution global land cover dataset through multisource data aggregation. *Science China Earth Sciences*, 57(10):2317–2329, 2014.

List of abbreviations

AA – Association Analysis
AC – Associative Classifier
ACCA – Automatic Cloud Cover Assessment
ASTER – Advanced Spaceborne Thermal Emission and Reflection Radiometer
ASTER-GDEM – ASTER Global Digital Elevation Model
BA – balanced accuracy
BU – built-up
CBERS – China–Brazil Earth Resources Satellite program
DMSP/OLS – Defense Meteorological Satellite Program / Operational Linescan System
DN – digital number
DRR – Disaster Risk Reduction
DSM – digital surface model
ENDI – Evidence-based Normalized Differential Index
EO – Earth Observation
ESM – European Settlement Map
ETM+ - Enhanced Thematic Mapper Plus sensor
EUROSTAT – European Union statistical office
FROM-GLC – Finer Resolution Observation and Monitoring of Global Land Cover
GDP – Gross Domestic Product
GEO – Group of Earth Observation
GEO GHS - GEO Global Human Settlement working group
GHSL – Global Human Settlement Layer
GLCM – grey-level co-occurrence matrix
GLS – Global Land Survey
IDP – Internally Displaced People
JRC – Joint Research Centre
L8 – Landsat 8 data
LLGC – Learning with Local and Global Consistency method
LUCAS – Land Use/Cover Area frame Survey
MLC – maximum likelihood classifier
MODIS – Moderate Resolution Imaging Spectroradiometer
MSS – Multispectral Scanner sensor
NDVI – Normalized Differential Vegetation Index
NUTS – Nomenclature of Territorial Units for Statistics
NYU – New York University
OSM – Open Street Maps
PDNA – Post-Disaster Need Assessment
QA – Quality Assessment
RF – Random Forest
ROC – receiver operating characteristic curve
RS – Remote Sensing
SDG – Sustainable Development Goals
SML – Symbolic Machine Learning
SPOT – Satellite Pour l'Observation de la Terre
SPOT-VGT – SPOT Vegetation

SRTM – Shuttle Radar Topography Mission
SVM – support vector machine
TA – total accuracy
TM – Thematic Mapper sensor
TMS – Tile Map Service
TOA – top-of-atmosphere
UK – United Kingdom
UN – United Nation
US – United States
USGS – United States Geological Survey
WB – World Bank

List of Figures

List of Tables

LB-NA-27741-EN-N

JRC Mission

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new methods, tools and standards, and sharing its know-how with the Member States, the scientific community and international partners.

Serving society
Stimulating innovation
Supporting legislation

doi:10.2788/253582

ISBN 978-92-79-55012-6