2 Products

2.1 GHS-BUILT-S R2023A - GHS built-up surface spatial raster dataset, derived from Sentinel-2 composite and Landsat, multi-temporal (1975-2030)

The GHS-BUILT-S spatial raster dataset depicts the distribution of the built-up (BU) surfaces estimates between 1975 and 2030 in 5 year intervals and two functional use components a) the total BU surface and b) the non-residential (NRES) BU surface. The data is made by spatial-temporal interpolation of five observed collections of multiple-sensor, multiple-platform satellite imageries: Landsat (MSS, TM, ETM sensor) data supports the 1975, 1990, 2000, and 2014 epochs, while a Sentinel-2 (S2) image composite (GHS-composite-S2 R2020A) supports the 2018 epoch.

The sub-pixel built-up fraction (BUFRAC) estimate at 10m resolution is produced from the 10m-resolution Sentinel-2 image composite, using as learning set a composite of data from GHS-BUILT-S2 R2020A, Facebook settlement delineation, Microsoft, and Open Street Map (OSM) building delineation. The inferential engine is a multiple-quantization-minimal-support (MQMS) generalization of the symbolic machine learning (SML) approach (Pesaresi, Syrris, et al., 2016). The SML for the classification of the Sentinel-2 data uses in input both radiometric and multi-scale morphological image descriptors (Pesaresi, Corbane, et al., 2016), including functional (i.e. RES, NRES) delineation of the built-up areas. In particular, the multiscale decomposition of the image information it is supported by the characteristic-saliency-levelling (CSL) model (Pesaresi et al., 2012) from generalization of the image segmentation based on the derivative of the morphological profile (DMP) (Pesaresi & Benediktsson, 2001). The multi-scale CSL it is solved by using a computationally efficient approach (Ouzounis et al., 2012). The inference is computed in data tiles of 100×100 km size.

The non-residential (NRES) built-up surface domain is predicted from S2 data by observation of radiometric, textural, and morphological features in a multi-faceted image processing framework merging global unsupervised rule-based reasoning and inductive locally-adaptive methods leveraging on pixel-wise spectral indexes, textural assessments, and object-oriented shape analysis. Textural analysis is performed by multi-scale, anisotropic and rotation-invariant contrast measurements using increasing displacement vectors of the co-occurrence matrix selecting the areas where contrast of large objects dominate the textural contrast generated by smaller image structures (Gueguen et al., 2012, Pesaresi et al., 2008). The connected component ("object") image analysis is solved by a segmentation of salient image structure based on the watershed of the inverse of the saliency layer as defined in the "characteristics-saliency-levelling" CSL (Pesaresi et al., 2012).

As in previous GHSL releases (Corbane et al., 2019; Pesaresi, Ehrlich, et al., 2016), the multi-temporal (MT) process works stepwise from recent epochs to past epochs, deleting the BU information if the decision is supported by empirical evidences from satellite data of the specific epoch. By definition, the process can only decrease the amount of built-up surface going from recent to past epochs. In this release, a similar logic generalized to the continuous prediction domain is applied within an object-oriented image processing framework. Salient spatial units are delineated by the watershed of the inverse of the continuous BUFRAC function at 10m resolution. This is done in order to increase the robustness of the change detected by the system, vs. the changing sensor data geometry (origin of the grid, resolution, projection) of the supporting image data in the various epochs. For each evaluated epoch and available Landsat scene, the probability Φ that any specific sensor sample (pixel or grid cell) can be associated to the foreground "built-up" (BU) vs. the background "non-built-up" (NBU) information semantic is evaluated, by observing the statistical association between the combinations of the quantized reflectance values and the training data. This inferential process is solved by multiple-quantization-minimal-support (MQMS) generalization of the symbolic machine learning (SML) approach (Pesaresi, Syrris, et al., 2016). The semantic Φ extracted at the pixel level of the different Landsat scenes in arbitrary geometries is aggregated to the data segments using a surface-weighted average. The final prediction on the amount of built-up surface change for each segment is solved by a multiple decision support approach evaluating ensemble linear regression model predictions from the semantic association of Φ to the BU vs. NBU class abstraction hypothesis, stratified in different data domains characterized by different expected sensor bias in discrimination of BU vs. NBU classes, that are cumulated and maximized from all the available input satellite scenes in the various epochs. Finally, the predictions on built-up surface change are aggregated by averaging at a uniform sample size of 100m grid cells and are used to build the final raster data.

In the intermediate epochs not covered by direct satellite observations, in areas not covered by satellite imageries (i.e. satellite data gaps in the 1975 epoch), or in the future epochs 2025 and 2030, the BU prediction it is solved by spatial-temporal interpolation or extrapolation based on a rank-optimal spatial allocation method. This supporting spatial optimization function combines static and dynamical components: the static component

is determined by the observation of the empirical association between the occurrence of specific land form combinations (slope, elevation, water) and the occurrence of human settlement development from remotely sensed data. The dynamical component is based on the spatial dynamics of the BU surface in the observed epochs, decomposed in a change (growth, or shrink) vs inertial (i.e. unchanged) BU dynamical field components.

2.1.1 Definitions

2.1.1.1 The building

Since the initial concept of the GHSL (Pesaresi et al., 2013) the adopted definition is the same as the INSPIRE "building" abstraction (https://inspire.ec.europa.eu/id/document/tg/bu), limited to the above-ground case, and without the "permanent" characterization of the built-up structures, allowing to be inclusive to temporary settlements as associated to slums, rapid migratory patterns, or displaced people because of natural disasters or crisis . "*… Buildings are constructions above (and/or underground) which are intended or used for the shelter of humans, animals, things, the production of economic goods or the delivery of services and that refer to any structure (permanently) constructed or erected on its site…*" . In short, and taking in to account the remote-sensing technology characteristics and limitations, the implicit GHSL abstraction of the "building" can be summarized as: "*any roofed structure erected above ground for any use*".

2.1.1.2 The built-up surface (BUSURF)

The built-up surface is the gross surface (including the thickness of the walls) bounded by the building wall perimeter with a spatial generalization matching the 1:10K topographic map specifications, that it also informally called "building footprint".

2.1.1.3 The built-up fraction (BUFRAC)

The built-up fraction (BUFRAC) is the share of the raster sample (i.e. pixel or grid cell) surface that is covered by the built-up surface.

2.1.1.4 The residential (RES) domain

The RES domain is defined as the built-up surface dedicated prevalently for residential use. The residential use is defined as from INSPIRE: *"…Areas used dominantly for housing of people. The forms of housing vary significantly between, and through, residential areas. These areas include single family housing, multi-family residential, or mobile homes in cities, towns and rural districts if they are not linked to primary production. It permits high density land use and low density uses. This class also includes residential areas mixed with other non-conflicting uses and other residential areas…"*[7]

2.1.1.5 The "non-residential domain" (NRES)

The "non-residential domain" (NRES) is defined as the domain of the BUFRAC>0 complement of the RES domain. This can be worded also as "*any built-up surface not included in the RES class*". As a logical corollary of the fact that in the RES domain definition also mixture with other not conflicting uses is allowed, the complement NRES domain is characterized by uses not compatible with the human residence.

Examples:

Let assume a 100m resolution spatial raster dataset with a 100×100 = 10,000 square meters of surface per spatial sample (pixel, or cell grid) of this spatial raster dataset. Moreover, let be the built-up surface predicted at the sample X of this grid BUSURFx = 750 square meters.

The corresponding built-up fraction estimate will be : BUFRACx = 750 / 10,000 = 0.075

Let assume in the sample x of 100×100m resolution the total BUSURFx = 4380 square meters, while the NRES BUSURFx = 850 square meters. Then the residential built-up surface RES BUSURFx can be predicted as RES BUSURFx = 4380 – 850 = 3530 square meters.

[7] https://inspire.ec.europa.eu/codelist/HILUCSValue/5_ResidentialUse

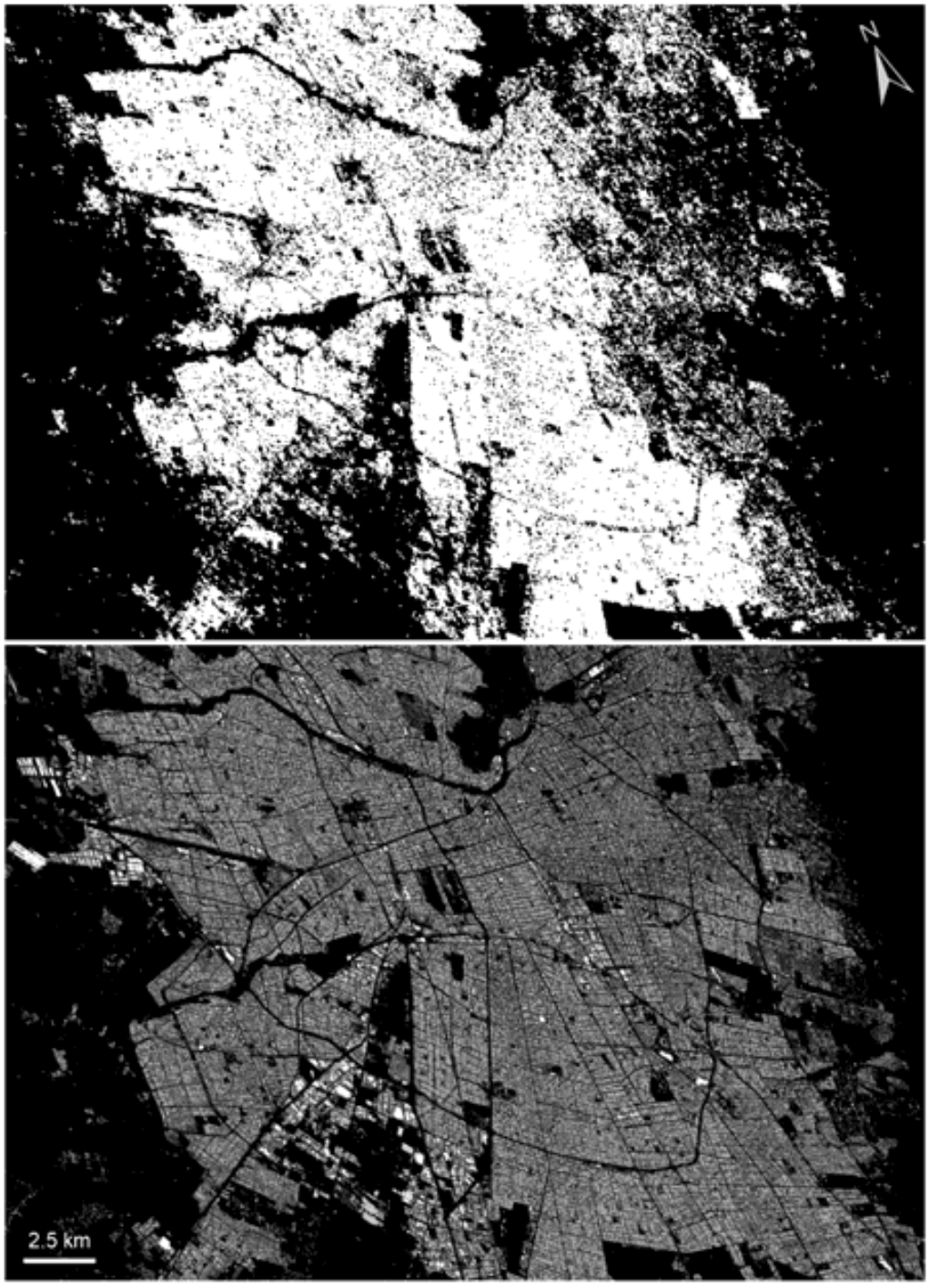

Figure 1 - Santiago de Chile: comparison of the built-up surfaces as assessed by the previous GHS_BUILT_LDSMT_GLOBE_R2018A for the epoch 2014 from Landsat image data with a Boolean 30m-resolution method (upper), vs the new GHS-BUILT-S_GLOBE_R2023A for the epoch 2018 from Sentinel-2 image data with a continuous 10m-resolution method (lower).

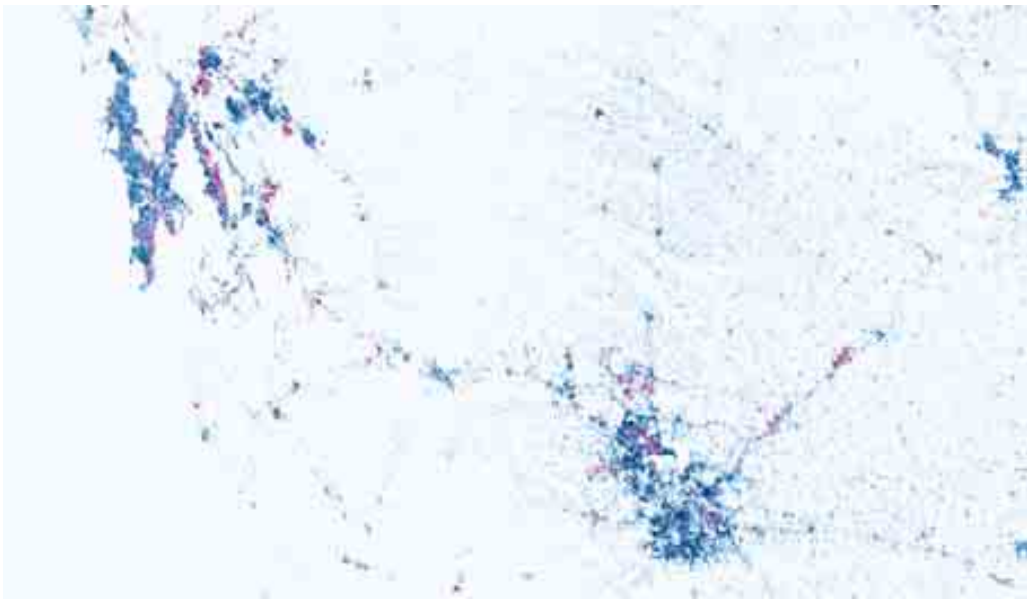

Figure 2 - Mumbai-Pune (India): residential (RES) and non-residential (NRES) components of the built-surfaces estimated for the GHSL 2020 epoch. RES and NRES are represented with blue and magenta, respectively.

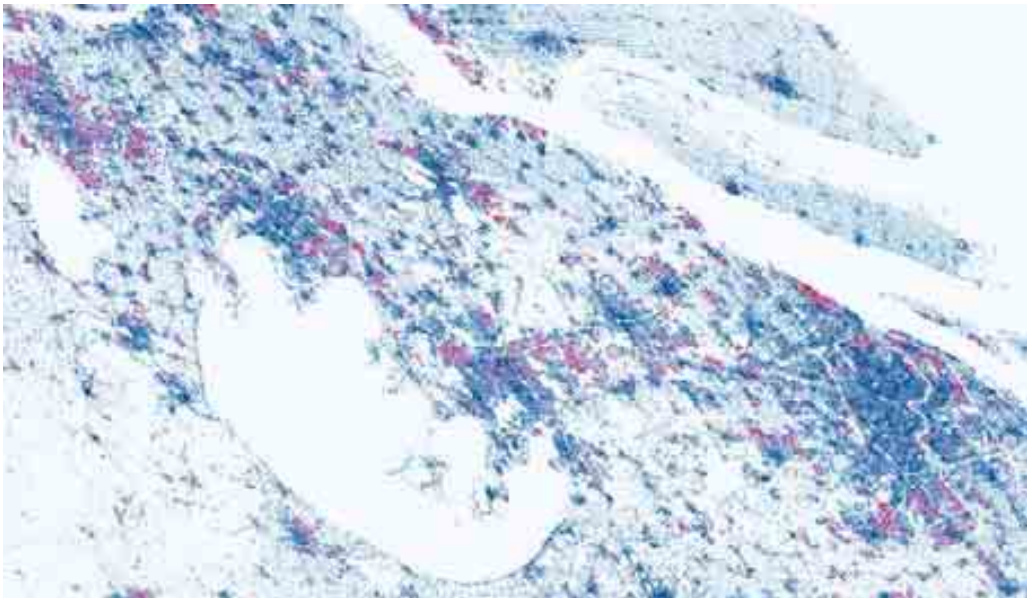

Figure 3 - Shanghai-Changzhou (China): residential (RES) and non-residential (NRES) components of the built-surfaces estimated for the GHSL 2020 epoch. RES and NRES are represented with blue and magenta, respectively.

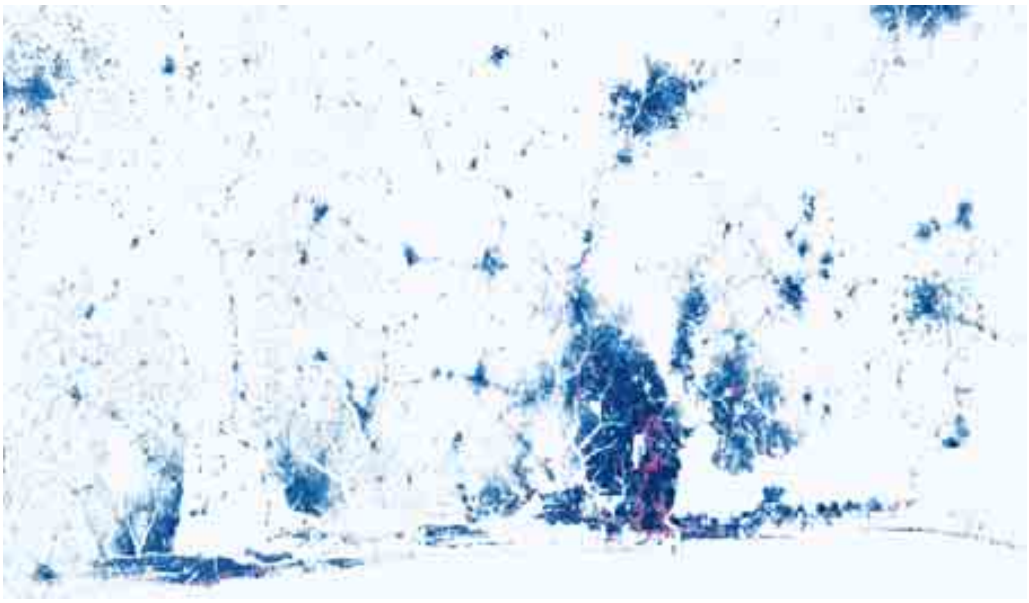

Figure 4 - Lagos-Porto Novo-Abeokuta (Nigeria): residential (RES) and non-residential (NRES) components of the built-surfaces estimated for the GHSL 2020 epoch. RES and NRES are represented with blue and magenta, respectively.

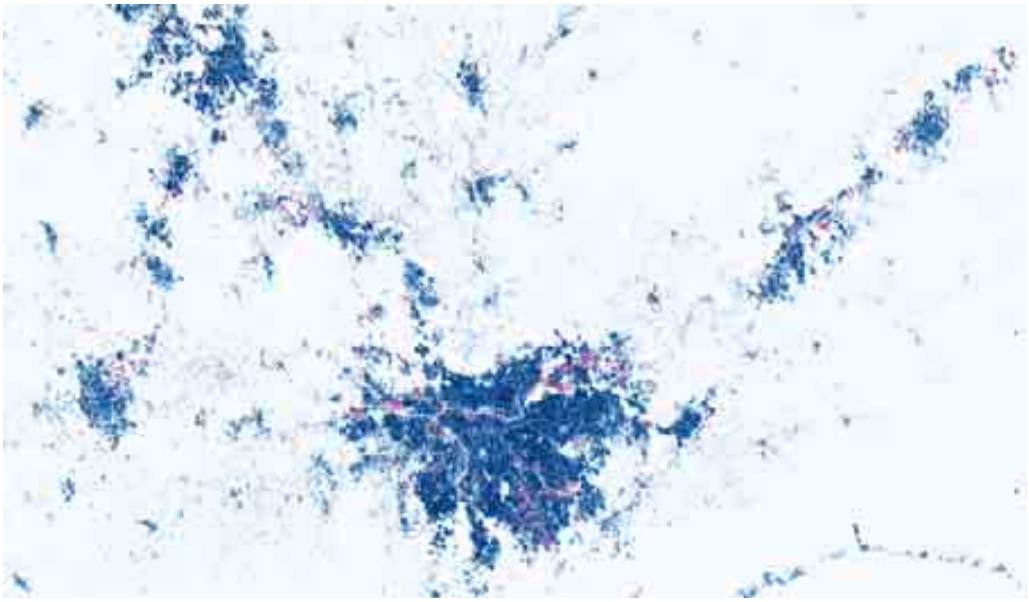

Figure 5 - Sao Paulo- Campinas - Sao Jose dos Campos (Brazil): residential (RES) and non-residential (NRES) components of the built-surfaces estimated for the GHSL 2020 epoch. RES and NRES are represented with blue and magenta, respectively.

Figure 6 - Detroit-Lansing-Flint (United States): residential (RES) and non-residential (NRES) components of the built-surfaces estimated for the GHSL 2020 epoch. RES and NRES are represented with blue and magenta, respectively.

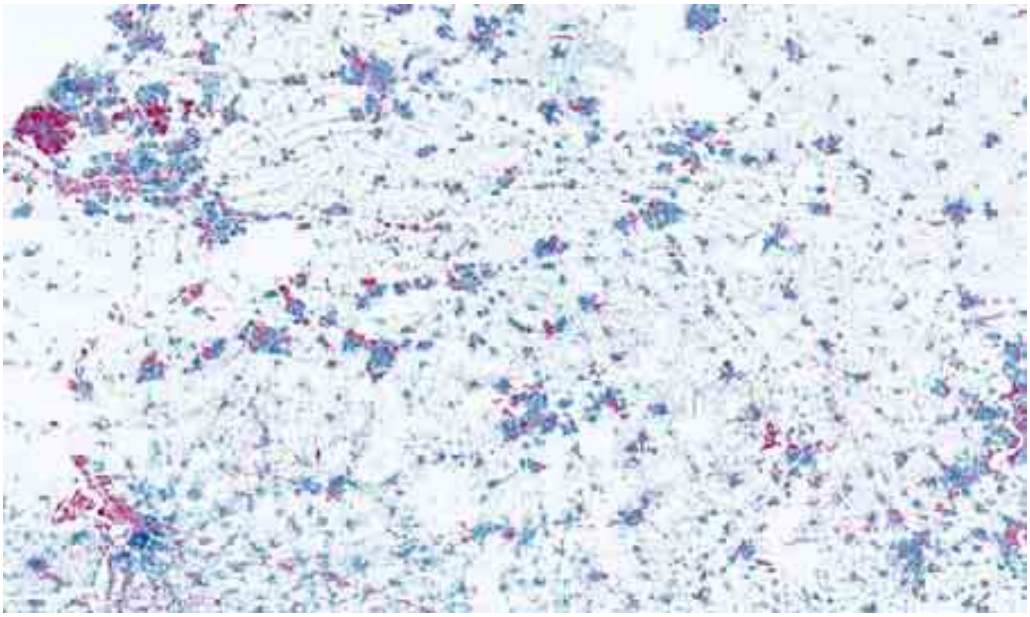

Figure 7 - The Hague - Rotterdam- Antwerp (The Netherlands): residential (RES) and non-residential (NRES) components of the built-surfaces estimated for the GHSL 2020 epoch. RES and NRES are represented with blue and magenta, respectively.

2.1.2 Expected Errors

The estimation of the GHS-BUILT-S errors is currently ongoing, and will be delivered in peer-reviewed publications targeting the different GHS-BUILT-S thematic aspects possibly in 2023-2024.

The ongoing error assessment it is solved by two complementary approaches a) comparing model predictions with human visual inspection of the same imagery data input, and b) comparing model predictions with other data of presumably higher accuracy after passing consistency and completeness checks. In the following, it is adopted a pragmatic approach by considering synonyms "accuracy metrics" or "agreement metrics", both measuring the agreement between the reference data and the model predictions, in both (a) and (b) approaches.

2.1.2.1 Errors in the 2018 predictions

On the error assessment approach (a) a total of 1 million sample points it is under assessment for the four Boolean classes listed below:

1. NBU_WATER : non built-up water surfaces
2. NBU_LAND: non built-up land surfaces
3. BU_RES: residential built-up surfaces
4. BU_NRES: non-residential built-up surfaces

A stratified uniform random sampling schema it is applied targeting an equalized number of samples for the four considered classes, uniformly distributed across the whole global landmass with the exception of Antarctica. Each random sample was visually inspected in three independent inspection campaigns (trials), performed by nine distinct professional photo-interpreters that were randomly assigned to each sample interpretation task. Each human labelling decision was accompanied by a high (H) vs. low (L) score of the confidence in the label assignment.

The results summarized here are aggregated from the first tranche of 250,000 validation points, which were available at the date of the report.

The class assignment from the different trials in each random sample is not necessarily the same, reflecting possibly different opinions of the human interpreters on the same sample. These uncertainties can be solved by two approaches: a) a voting schema and b) subsampling of the observations minimizing the human disagreement domain. This subsampling is done by using a metric called the Joint Agreement High Confidence (JAHC), which is defined as the domain where all the three independent human interpretations provide the same class answer, all of them with a high confidence score, so it represents the domain with the highest confidence of the human interpretation results. Among the 250,000 considered points, 155,649 (62.26%) are JAHC points, accordingly to the first tranche of 250,000 observed points discussed here. Observing the geographic distribution, the JAHC share ranges from 57.1% to 70.1% of Europe and Oceania, respectively (Table 1).

Table 1 – Total and JAHC samples per Region, sorted by decreasing Joint Agreement High Confidence share of the human interpretation of S2 data.

Region	N Samples (All)	N Samples JAHC (*)	JAHC share
Oceania	13,542	9,505	70.19%
America	78,061	50,094	64.17%
Africa	38,779	24,596	63.43%
ALL	250,000	155,649	62.26%
Asia	71,279	43,821	61.48%
Europe	48,339	27,633	57.17%

(*) Joint Agreement High Confidence (JAHC) samples

Table 2 and Table 3 show the confusion matrix and the accuracy metrics for the 4-class classification scenario, by solving the human uncertainty using the voting schema, and the subsampling to the JAHC universe, respectively. In both cases, per-class Precision and Recall scores (also called Producer and User accuracy) are summarized in the horizontal and vertical highlighted rows, respectively. The voting approach including both High and Low confidence scores of the human labelling, yields an overall accuracy of 82.6% on the 250,000 reference points (Table 2), while the subsampling on the high confidence high agreement of the JAHC domain yields an overall accuracy of 90.5% on the 155,649 reference points passing the agreement and confidence tests (Table 3).

Table 4 shows the summary of the 4-class agreement metrics by sub-regions of the world[8] ordered by decreasing accuracy scores, focusing the attention on the optimal JAHC human interpretation domain. Significative positive accuracy extrema are collected in South East Asia (94.8%) and Melanesia (95.2%), while the worst region is the Central Africa and Southern Africa, yielding an 86.4% and 83.8% accuracy, respectively.

Table 2 – Reference set by voting schema, any confidence level: confusion matrix and accuracy or agreement metrics

Label		**OBSERVED NBU_WATER**	**OBSERVED NBU_LAND**	**OBSERVED BU_RES**	**OBSERVED BU_NRES**	**Total predicted**	**Total predicted [%]**	**Metrics**	
		99.5%	*56.8%*	*88.7%*	*83.3%*				
PREDICTED NBU_WATER	*95.5%*	52,323	2,401	1	43	54,768	21.9%	Precision	82.1%
PREDICTED NBU_LAND	*98.6%*	215	30,575	75	159	31,024	12.4%	Recall	86.2%
PREDICTED BU_RES	*73.3%*	2	11,446	63,750	11,746	86,944	34.8%	Accuracy	82.6%
PREDICTED BU_NRES	*77.4%*	31	9,389	8,037	59,807	77,264	30.9%	Specificity	94.3%
Total observed		52,571	53,811	71,863	71,755	250,000		F1 score	82.5%
Total observed [%]		21.0%	21.5%	28.7%	28.7%				

Table 3 – Reference set by subsampling in the JAHC domain: confusion matrix and accuracy or agreement metrics

Label		**OBSERVED NBU_WATER**	**OBSERVED NBU_LAND**	**OBSERVED BU_RES**	**OBSERVED BU_NRES**	**Total predicted**	**Total predicted [%]**	**Metrics**	
		99.9%	*73.5%*	*95.0%*	*91.9%*				
PREDICTED NBU_WATER	*98.0%*	47,873	994	0	0	48,867	31.4%	Precision	90.1%
PREDICTED NBU_LAND	*99.8%*	60	28,439	5	0	28,504	18.3%	Recall	90.6%
PREDICTED BU_RES	*85.3%*	0	3,344	34,761	2,633	40,738	26.2%	Accuracy	90.5%
PREDICTED BU_NRES	*79.3%*	2	5,933	1,824	29,781	37,540	24.1%	Specificity	97.0%
Total observed		47,935	38,710	36,590	32,414	155,649		F1 score	89.6%
Total observed [%]		30.8%	24.9%	23.5%	20.8%				

[8] Sub-regions are drafted accordingly to the UN world population prospect 2022 definitions (https://population.un.org/wpp/)

Table 4 – Agreement metrics by sub-region in the JAHC domain, ordered by decreasing overall accuracy. Precision and Recall are also called "Producer Accuracy" and "User Accuracy", respectively.

Region	Precision	Recall	Overall Accuracy	Specificity	F1 Score	N Samples
Polynesia	91.9%	95.7%	98.0%	99.4%	93.5%	355
Micronesia	73.2%	70.5%	97.4%	99.2%	71.6%	352
Melanesia	88.5%	84.2%	95.2%	98.5%	85.6%	1,653
South Eastern Asia	94.1%	93.9%	94.8%	98.3%	93.7%	10,794
Seven seas open ocean	74.2%	62.5%	94.8%	97.8%	66.9%	193
Western Europe	91.5%	96.1%	94.3%	98.2%	93.3%	2,270
Middle Africa	92.0%	85.3%	93.8%	97.9%	87.1%	5,527
Central America	92.3%	94.6%	93.7'	98.0%	93.2%	4,514
Caribbean	92.2%	94.1%	93.7%	98.0%	92.76	1,487
Australia and New Zealand	91.7%	91.7%	93.5%	97.9%	91.4%	6,952
Eastern Europe	92.4%	90.3%	92.8%	97.7%	90.7%	18,060
Eastern	90.2%	92.3%	91.7%	97.4%	90.3%	15,631
Western Africa	92.1%	85.8%	91.5%	97.0%	87.2%	4,380
Southern Asia	90.3%	91.76	90.8%	96.9%	90.2%	9,009
ALL	90.1%	90.6%	90.5%	97.0%	89.6%	155,649
Northern Africa	88.7%	86.9%	89.5%	96.4%	87.4%	4,782
Northern Europe	87.4%	89.9%	89.3%	96.6%	87.8%	3,533
South America	88.0%	88.2%	89.1%	96.6%	86.8%	21,434
Eastern Africa	88.9%	34.5%	88.5%	96.3%	84.7%	6,989
Southern Europe	85.4%	91.8%	87.7%	96.3%	86.4%	3,770
Northern America	88.6%	84.0%	87.1%	95.8%	84.3%	22,659
Western Asia	87.0%	88.9%	86.6%	95.5%	87.3%	5,019
Central Asia	88.0%	83.6%	86.4%	95.7%	83.9%	3,368
Southern Africa	85.2%	85.7%	83.8%	94.9%	83.5%	2,918

Table 5 shows the 2-class agreement in model detection of (a) the BU vs. NBU and (b) the WATER vs. LAND abstraction semantics from the S2 data as compared with human visual inspection of the same image data in the high confidence JAHC domain. The performances are aggregated by regions in the world and ordered by accuracy value added of the R2023 vs. the previous R2019 GHSL release if evaluated using the same reference samples. Noticeable is the increase in the BU vs. NBU accuracy of the new release in Asia and Africa (+23.54%, + 22.69%, respectively), while the least increase in accuracy is yielded in Europe (+13.55%).

Table 5 - Accuracy performances in model detection as compared with human visual inspection of the same image data. Jaccard similarity (also called intersection-over-union), Overall Accuracy, Commission Error and Omission Error. (a) the BU vs. NBU and (b) the WATER vs. LAND abstraction semantics. (right) the new GHSL release R2023, (left) the previous GHSL release R2019 tested vs. the same reference data

		GHSL R2019 / LDS_SML2019				GHSL R2023 / S2_SML2023				
(a) BU vs NBU	**N samples**	**Jaccard similarity**	**Overall accuracy**	**Commission error**	**Omission error**	**Jaccard similarity**	**Overall accuracy**	**Commission error**	**Omission error**	**Increase of Overall Accuracy**
Asia	43,821	42.71%	70.51%	0.08744	0.55314	88.29%	94.05%	0.11702	0.00013	23.54%
Africa	24,596	34.77%	70.37%	0.10282	0.63797	86.53%	93.06%	0.13466	0.00000	22.69%
Oceania	9,505	19.54%	76.95%	0.58935	0.72184	91.43%	98.04%	0.08568	0.00000	21.09%
ALL	155,649	48.38%	75.47%	0.12162	0.48148	88.14%	94.04%	0.11854	0.00007	18.56%
America	50,094	56.83%	78.34%	0.09401	0.38766	88.87%	94.48%	0.11124	0.00003	16.14%
Europe	27,633	65.53%	80.53%	0.13341	0.27412	89.62%	94.09%	0.10379	0.00000	13.55%
(b) WATER vs LAND										
Oceania	9,505	74.33%	85.82%	0.05663	0.22095	99.00%	99.43%	0.00303	0.00700	13.61%
Africa	24,596	53.87%	88.01%	0.00353	0.46030	97.17%	99.21%	0.02826	0.00000	11.21%
Asia	43,821	62.55%	89.59%	0.00967	0.37088	98.78%	99.68%	0.01105	0.00117	10.10%
ALL	155,649	68.07%	90.00%	0.02445	0.30746	97.84%	99.32%	0.02034	0.00129	9.32%
America	50,094	77.78%	92.99%	0.02270	0.20552	98.30%	99.41%	0.01580	0.00123	6.42%
Europe	27,633	82.39%	94.83%	0.00272	0.17420	99.23%	99.73%	0.00698	0.00069	4.90%

On the error assessment approach (b) the preliminary error scores in prediction of the BUFRAC were estimated vs. observed built-up surfaces from building footprints available in vector format at scale 1:10K. The test set is made by the global collection of building footprints used for quality control during the GHSL production.

Table 6 shows the expected errors of the new GHS-BUILT-S R2023A release at 10m resolution stratified by class of the Copernicus Global Land cover at 100m resolution (Buchhorn et al., 2020).

Table 7 shows the expected errors of the new GHS-BUILT-S R2023A release at 100m resolution as compared to the previous GHSL release made from Boolean classification of Landsat data (GHS-BUILT R2018A) and as compared to the predictions included in the continuous "urban cover fraction" (UCF) of the Copernicus Global Land cover at 100m resolution (Buchhorn et al., 2020). The test data are subdivided in two geographical strata ("non-US" and "US") in order to control the performances of the model in different settlement pattern conditions.

The capacity of GHS-BUILT-S R2023A to predict the BUFRAC (quasi continuous, 64 levels) in a globally-representative set of almost 50,000 test cases of 80x80 meter size visually inspected with a Boolean interpretation schema at 10m resolution (See et al., 2022), yield a Pearson Correlation Coefficient of the linear least square regression equal to 0.81363. To be noticed that the correlation is systematically decreased by the fact that the reference data is not spatially aligned with the GHSL data, and by the fact that the GHSL uses a continuous classification schema of the 10m-res raster samples, while the reference data used in See et al. applies a human interpretation schema based on Boolean classification.

Table 8 shows the amount of total built-up surface and NRES built-up surface assessed by the GHS-BUILT-S R2023A data (epoch 2020) stratified by land use classes in United States (NLUD[9]; Theobald, 2014), and Europe

9 Land Use Classification and Map for the US http://csp-inc.org/public/NLUD2010_20140326.zip

(CLC[10]), ordered by the GHSL NRES surface share. The table shows the empirical association between the NRES class and the land use classes. To be noticed that the measured association is only indicative and systematically decreased by the fact that the GHSL data is derived from 10m-resolution imagery and consequently has a much higher spatial precision as compared to the land use data used as reference, which is defined with a minimal mapping unit in the order of hectares.

Table 6 – Expected errors of the new GHS-BUILT-S R2023A release at 10m resolution stratified by class of the Copernicus Global Land cover at 100m resolution (Buchhorn et al., 2020).

CODE	LABEL	RMSE	MAE	NSAMPLES
0	Not Classified	0.122	0.070	7,509
20	Shrubs	0.120	0.056	555,261,719
30	Herbaceous vegetation	0.136	0.060	1,310,171,720
40	Cultivated and managed vegetation agriculture cropland	0.071	0.020	914,235,915
50	Urban / built up	0.296	0.218	337,089,799
60	Bare soil or sparse vegetation	0.192	0.111	123,557,016
70	Snow and Ice	0.001	0.000	13,447,153
80	Permanent water bodies	0.028	0.005	116,262,034
90	Herbaceous wetland	0.062	0.019	63,299,127
100	Moss and lichen	0.006	0.001	347,875
111	Closed forest, evergreen needle leaf	0.071	0.025	573,176,106
112	Closed forest, evergreen, broad leaf	0.012	0.002	575,661,481
113	Closed forest, deciduous needle leaf	0.069	0.018	19,610
114	Closed forest, deciduous broad leaf	0.035	0.007	751,956,669
115	Closed forest, mixed	0.027	0.006	145,625,305
116	Closed forest, unknown	0.071	0.026	142,841,787
121	Open forest, evergreen needle leaf	0.101	0.042	111,872,259
122	Open forest, evergreen broad leaf	0.023	0.005	2,599,647
123	Open forest, deciduous needle leaf	0.000	0.000	181
124	Open forest, deciduous broad leaf	0.067	0.021	168,233,648
125	Open forest, mixed	0.048	0.012	1,299,363
126	Open forest, unknown	0.107	0.043	839,029,693
200	Open sea	0.072	0.025	144,012,925
	Total	0.075	0.034	6,890,008,541

Table 7 – Expected error scores in prediction of the built-up surface fraction (BUFRAC) at the aggregated 100m and 1km resolution.

	Non-US			US			Total		
Prediction at 100m resolution	**RMSE**	**MAE**	**N Samples**	**RMSE**	**MAE**	**N Samples**	**RMSE**	**MAE**	**N Samples**
GHS-BUILT-S R2022	0.062	0.035	32,418,503	0.062	0.035	35,886,736	0.062	0.035	68,305,239
GHS-BUILT R2018A	0.194	0.129	32,418,503	0.308	0.213	35,886,736	0.258	0.177	68,305,239
Copernicus Global Land Service UCF	0.255	0.173	32,418,503	0.285	0.195	35,886,736	0.272	0.186	68,305,239
Prediction at 1km resolution	**RMSE**	**MAE**	**N Samples**	**RMSE**	**MAE**	**N Samples**	**RMSE**	**MAE**	**N Samples**
GHS-BUILT-S R2022	0.036	0.026	301,940	0.038	0.030	329,121	0.037	0.028	631,061
GHS-BUILT R2018A	0.144	0.110	301,940	0.259	0.209	329,121	0.213	0.170	631,061
Copernicus Global Land Service UCF	0.195	0.148	301,940	0.242	0.193	329,121	0.223	0.175	631,061

10 Corine Land Cover European seamless 100m raster database (Version 20b2)

https://land.copernicus.eu/pan-european/corine-land-cover/clc2018?tab=metadata

Table 8 – The amount of total built-up surfaces, the NRES built-up surfaces assessed in the GHS-BUILT-S R2023A data and the NRES built-up surface share stratified by land use classes in United States (NLUD) and Europe (CLC), ordered by decreasing NRES surface share.

LUSOURC	LUCLASS	LABEL1	LABEL2	LABEL3	BUTOT m2	BUNRES m2	NRES_sha
NLUD	251	Built-up	Transportation	Airports (developed)	371014506	253357093	68.29%
CLC	121	Artificial surfaces	Industrial, commercial and transport units	Industrial or commercial units	7649839186	5056391057	66.10%
CLC	123	Artificial surfaces	Industrial, commercial and transport units	Port areas	327183968	215970214	66.01%
NLUD	231	Built-up	Industrial	Factory, plant	1619490370	1045588694	64.56%
NLUD	223	Built-up	Commercial	Entertainment (stadiums, amusement, etc.)	53632078	33581582	62.61%
NLUD	221	Built-up	Commercial	Office	1614791936	978439733	60.59%
NLUD	222	Built-up	Commercial	Retail/shopping centers	1742663387	1039508683	59.65%
CLC	124	Artificial surfaces	Industrial, commercial and transport units	Airports	170180711	98086496	57.64%
NLUD	233	Built-up	Industrial	Confined animal feeding	88703844	50398369	56.82%
NLUD	249	Built-up	Institutional	Prison/penitentiary	16152586	8924402	55.25%
NLUD	241	Built-up	Institutional	Schools (dev)	225846287	123608473	54.73%
NLUD	242	Built-up	Institutional	Schools (undeveloped)	598016804	326285150	54.56%
NLUD	243	Built-up	Institutional	Medical (hospitals, nursing home, etc.)	367951278	196239687	53.33%
NLUD	244	Built-up	Institutional	Government/public	86726621	44112906	50.86%
CLC	212	Agricultural areas	Arable land	Permanently irrigated land	622458355	268234717	43.09%
CLC	132	Artificial surfaces	Mine, dump and construction sites	Dump sites	40734981	12866115	31.58%
NLUD	261	Built-up	Transportation	Rural buildings, cemetery	1242246217	391906221	31.55%
NLUD	341	Production	Timber	Timber harvest	165945723	46319990	27.91%
CLC	122	Artificial surfaces	Industrial, commercial and transport units	Road and rail networks and associated land	224903444	57729502	25.67%
CLC	133	Artificial surfaces	Mine, dump and construction sites	Construction sites	94362978	22856256	24.22%
NLUD	255	Built-up	Transportation	Undeveloped	173761477	39555968	22.76%
CLC	213	Agricultural areas	Arable land	Rice fields	32328402	6089237	18.84%
CLC	131	Artificial surfaces	Mine, dump and construction sites	Mineral extraction sites	142861969	24980523	17.49%
NLUD	330	Production	Mining	Mining strip mines, quarries, gravel pits	8185807	1420084	17.35%
NLUD	245	Built-up	Institutional	Military/DOD/DOE (dev)	104554407	17490907	16.73%
NLUD	246	Built-up	Institutional	Military/DOD (training)	217411481	31391901	14.44%
CLC	211	Agricultural areas	Arable land	Non-irrigated arable land	7385826240	982409064	13.30%
CLC	141	Artificial surfaces	Artificial, non-agricultural vegetated areas	Green urban areas	165087965	20670163	12.52%
NLUD	252	Built-up	Transportation	Highways, railways	1042336050	121981507	11.70%
CLC	111	Artificial surfaces	Urban fabric	Continuous urban fabric	2160143750	194234820	8.99%
CLC	222	Agricultural areas	Permanent crops	Fruit trees and berry plantations	398088530	33459968	8.41%
NLUD	411	Recreation	Developed park	Urban park	202120970	16703263	8.26%
CLC	142	Artificial surfaces	Artificial, non-agricultural vegetated areas	Sport and leisure facilities	586211113	45131809	7.70%
CLC	231	Agricultural areas	Pastures	Pastures	4571276116	337410401	7.38%
NLUD	310	Production	General	General agricultural	158011938	11520002	7.29%
NLUD	313	Production	Cropland	Orchards	71453036	4800693	6.72%
CLC	112	Artificial surfaces	Urban fabric	Discontinuous urban fabric	30206040355	1990977548	6.59%
NLUD	314	Production	Cropland	Sod & switch grass	14887041	933461	6.27%
NLUD	311	Production	Cropland	Cropland/row crops	3107157034	190204508	6.12%
CLC	242	Agricultural areas	Heterogeneous agricultural areas	Complex cultivation patterns	4975196695	283937849	5.71%
NLUD	213	Built-up	Residential	Suburban (1–2.5 ac)	9335940633	520592660	5.58%
CLC	243	Agricultural areas	Heterogeneous agricultural areas	Land principally occupied by agriculture, with si	2121377478	109108914	5.14%
NLUD	321	Production	Rangeland	Grazed	3116664551	160096140	5.14%
CLC	221	Agricultural areas	Permanent crops	Vineyards	391114177	19693622	5.04%
NLUD	214	Built-up	Residential	Exurban (2.5–10 ac)	8778472493	408859021	4.66%
NLUD	211	Built-up	Residential	Dense urban (>0.1 ac)	1461479361	64398593	4.41%
CLC	223	Agricultural areas	Permanent crops	Olive groves	436848677	17939775	4.11%
NLUD	415	Recreation	Developed park	Resort/ski area	43570356	1739805	3.99%
NLUD	410	Recreation	Undifferentiated park	General park	201511693	7926588	3.93%
NLUD	421	Recreation	Natural park	Natural park	144612371	5448426	3.77%
NLUD	417	Recreation	Developed park	Campground/ranger station	20170842	710508	3.52%
CLC	244	Agricultural areas	Heterogeneous agricultural areas	Agro-forestry areas	44227686	1378654	3.12%
CLC	241	Agricultural areas	Heterogeneous agricultural areas	Annual crops associated with permanent crops	149655325	4474819	2.99%
NLUD	412	Recreation	Developed park	Golf course	202243104	5893655	2.91%
NLUD	312	Production	Cropland	Pastureland	3456950217	91632160	2.65%
NLUD	212	Built-up	Residential	Urban (0.1–1)	16435713363	399274360	2.43%
NLUD	215	Built-up	Residential	Rural (10–40 ac)	5350324389	111117959	2.08%
NLUD	422	Recreation	Natural park	Designated recreation area	203351	2066	1.02%

2.1.2.2 Errors in the multi-temporal predictions

Preliminary error assessment results in the multi-temporal domain include the comparison of model-predicted built-up surfaces vs. observed built-up surfaces as deduced from the rasterization of vector building footprint data including the year of construction that are available in France, Spain, the Netherlands, Switzerland and the US (Uhl & Leyk, 2022). Table 9 shows the number of valid samples used in the preliminary multi-temporal test by country and by degree of urbanization spatial raster dataset (GHS-SMOD 1km epoch 2020) URBAN vs. RURAL stratification.

Table 9 – number of valid samples used in the MT test

	RURAL	URBAN	ALL
France	2,862,465	658,771	**3,521,236**
Netherlands	207,186	309,022	**516,208**
Spain	929,967	550,710	**1,480,677**
Switzerland	6,244	17,600	**23,844**
USA	87,947	132,451	**220,398**
Grand Total	**4,093,809**	**1,668,554**	

Figure 8 shows the time series of agreement measures estimated by a generalized version of the Jaccard similarity index to the continuous numerical domain (Costa, 2022) of the built-up surface predictions at 100m of spatial resolution in the temporal domain 1975-2015 in 5 year intervals. Besides the aforementioned rural-urban stratification, some alternative products were used, assessing similar information as the GHS-BUILT-S. In particular, the BENCHMARK(R2022A) case is the GHSL data released in June 2022, the GHSL R2023A is the current release, the PRIOR(GHS_B_P2019) is the previous GHSL release (Corbane et al., 2019), the PRIOR(GISA) is the Global Impervious Surface Area (Huang et al., 2022), and the PRIOR(WSF_EVO) is the World Settlement Footprint Evolution product (Marconcini et al., 2021). According to these empirical evidences, the new R2023A discussed here i) improves the accuracy of the previous release R2022A in both the URBAN and RURAL application domains in all the predicted epochs, and ii) among the considered alternative sources, it is the best predictor of the built-up surfaces at 100m resolution in both the URBAN and RURAL application domains in all the predicted epochs.

Figure 9 and Figure 10 show the temporal evolution of the predicted and observed (REF) normalized built-up surfaces in the considered test areas, in the URBAN 2020 and RURAL 2020 stratum, respectively, in three different GHSL models: The R2019, the R2022A, and the current R2023A. The built-up surfaces are normalized by the respective average in all the considered epochs. According to these results i) the new R2023A release is more sensitive to the change in URBAN domain as compared to the previous releases R2019, R2022 (Figure 9) and ii) the new release R2023 fixes the issue on the unrealistic change rates in the RURAL domain observed in the R2022 (Figure 10).

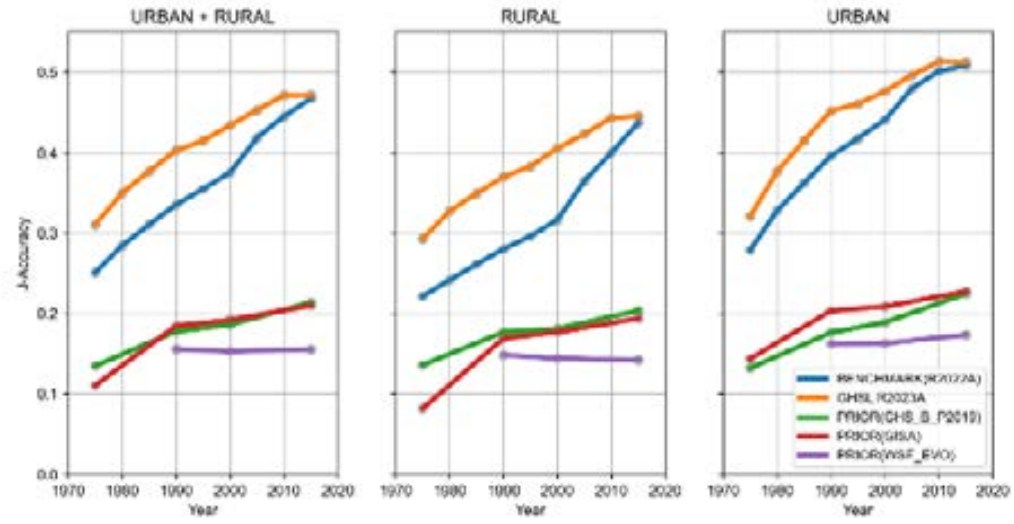

Figure 8 - Multi-temporal accuracy estimations in URBAN and RURAL domains, and across both domains. J-Accuracy is the generalized version of the Jaccard similarity index to the continuous numerical domain (Costa, 2022)

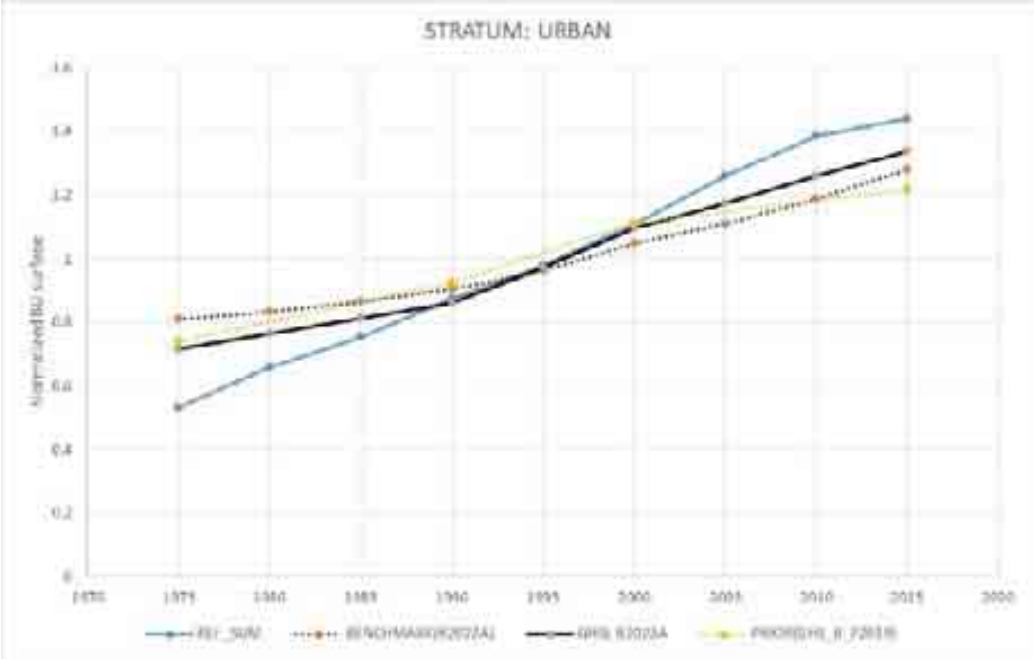

Figure 9 – Temporal evolution of the predicted and observed (REF_SUM) normalized built-up surfaces in the considered test areas, URBAN 2020 stratum.

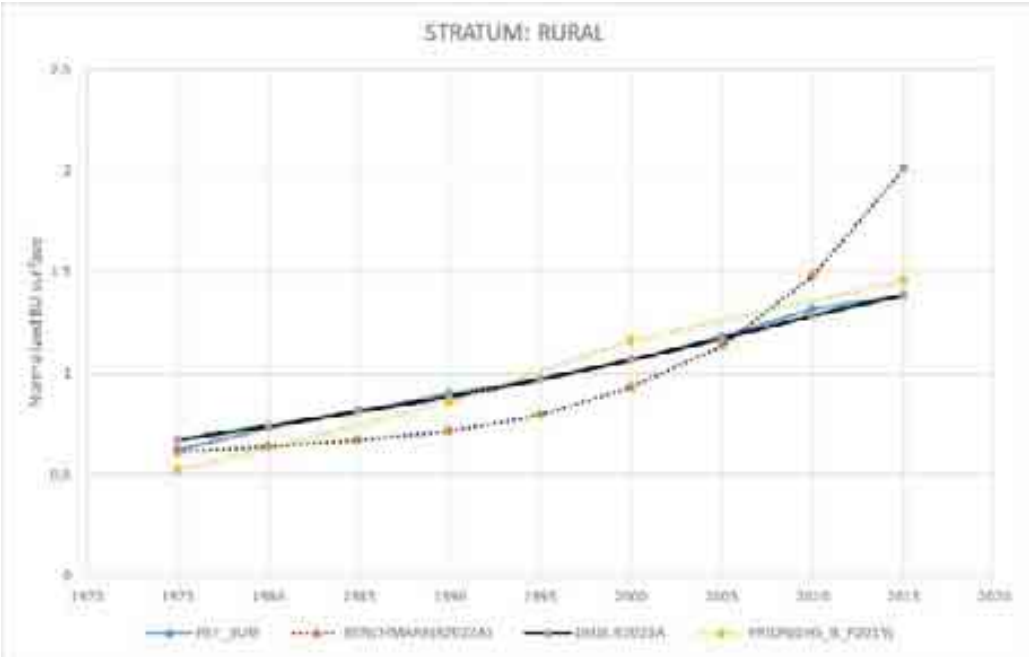

Figure 10- Temporal evolution of the predicted and observed (REF_SUM) normalized built-up surfaces in the considered test areas, RURAL 2020 stratum.

2.1.3 Improvements compared to the previous release

Improved input

Improved input data used in this release includes: improved satellite imagery and new prior knowledge and learning set.

Regarding improved satellite imagery we have included a more populated historical Landsat data series, a new epoch 2018 from Sentinel-2 with 10m spatial resolution vs. previous 30m Landsat data.

Regarding new prior knowledge and learning set we have included new BU labels as i) the GHS-BUILT-S2 R2020A[11] that is a probability to the BU class spatial raster dataset derived from Sentinel-2 global image composite for reference year 2018 using Convolutional Neural Networks (GHS-S2Net; Corbane, Syrris, et al., 2020) and ii) the buildings and the settlement delineation derived from VHR imagery by Microsoft[12] and Facebook[13] open efforts, the new BU change map as included in the GHS-BUILT R2018A, GHS built-up spatial raster dataset, derived from Landsat, multi-temporal (1975-1990-2000-2014), and Land Use and other information included in National land use data (US NLCD[14], EU CORINE[15]) and Volunteered geographical information by Open Street Map (OSM)[16] on LANDUSE, ROADS, RIVER, STREAMS.

List of Countries where Microsoft building footprint data were available during the GHSL production: USA, Canada, Australia, Uganda, Tanzania, Nigeria, Kenya, Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Guyana, Paraguay, Peru, Uruguay, Venezuela.

List of Countries where high resolution settlement layer (HRSL) data from Facebook were available during the GHSL production: Albania, Algeria, American Samoa, Andorra, Angola, Anguilla, Antigua and Barbuda, Argentina, Aruba, Australia, Austria, Bahamas, Bahrain, Bangladesh, Barbados, Belarus, Belgium, Belize, Benin, Bhutan, Bolivia, Bosnia and Herzegovina, Botswana, Brazil, British Virgin Islands, Brunei, Bulgaria, Burkina Faso, Burundi, Cabo Verde, Cambodia, Cameroon, Cayman Islands, Central African Republic, Chad, Chile, Colombia, Comoros, Congo, Cook Island, Costa Rica, Cote d'Ivoire, Croatia, Czechia, Democratic Republic of the Congo, Djibouti, Dominica, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea, Eritrea, Estonia, Federated States of Micronesia, Fiji, France, French Guiana, French Polynesia, Gabon, Gambia, Gibraltar, Georgia, Germany, Ghana, Greece, Grenada, Guadeloupe, Guam, Guatemala, Guinea, Guinea Bissau, Guyana, Haiti, Honduras, China Hong Kong Special Administrative Region, Hungary, Iceland, Indonesia, Iraq, Ireland, Italy, Jamaica, Japan, Jordan, Kazakhstan, Kenya, Kingdom of Eswatini, Kiribati, Kuwait, Kyrgyzstan, Laos, Latvia, Lebanon, Lesotho, Liberia, Libya, Liechtenstein, Lithuania, Luxemburg, China Macao Special Administrative Region, Madagascar, Malawi, Malaysia, Maldives, Mali, Malta, Marshall Islands, Mauritania, Mauritius, Mayotte, Mexico, Moldova, Monaco, Mongolia, Montserrat, Mozambique, Namibia, Naura, Nepal, Netherlands, New Caledonia, New Zealand, Nicaragua, Niger, Nigeria, The former Yugoslav Republic of Macedonia, Northern Mariana Islands, Oman, Palau, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Poland, Portugal, Puerto Rico, Qatar, Reunion, Romania, Rwanda, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Samoa, San Marino, Sao Tome and Principe, Saudi Arabia, Senegal, Serbia, Seychelles, Sierra Leone, Singapore, Slovakia, Slovenia, Solomon Islands, South Africa, South Korea, Spain, Sri Lanka, Suriname, Switzerland, Taiwan, Tajikistan, Thailand, Timor Leste, Togo, Tonga, Trinidad and Tobago, Tunisia, Turkmenistan, Turks and Caicos, Tuvalu, Uganda, United Arab Emirates, United Kingdom, United Republic of Tanzania, Uruguay, US Virgin Islands, United States of America, Uzbekistan, Vanuatu, Vietnam, Wallis and Futuna Islands, Zambia, Zimbabwe

Improved Output

Several improvements regarding the assessment of the built-up surfaces are included in this new release, as compared to the previous GHSL data. They may be summarized in the following points:

---

[11] https://ghsl.jrc.ec.europa.eu/download.php?ds=buS2

[12] https://github.com/microsoft/USBuildingFootprints

[13] https://research.fb.com/downloads/high-resolution-settlement-layer-hrsl/

[14] https://www.mrlc.gov/data

[15] https://land.copernicus.eu/pan-european/corine-land-cover/clc2018

[16] https://www.openstreetmap.org/

The accuracy of the Boolean prediction of BU vs. NBU surfaces has improved from the 75.47% of the previous R2019 to the 94.04% of the current R2022/R2023 using the same reference data. Accuracy improvements are remarkable in Asia and Africa (Table 5). The WATER vs. LAND Boolean discrimination has also improved significantly as compared to the previous release R2019 (Table 5).

The prediction of the continuous built-up surface improved considerably with a MAE at 100m-resolution that drops from 0.117 of the previous release to the 0.035 of the current release (Table 7).

The new GHSL release includes a new classification of the non-residential (NRES) built-up surfaces that was not available in the previous releases. This feature will improve the usability of the new GHSL data in applications requiring a functional classification of the built environment.

The new GHSL release produces data at equal intervals in the time period 1975-2030 using a spatial-temporal interpolation process, while previous GHSL releases were reporting data in arbitrary points in time where satellite images were available. This will further enhance the usability of the new GHSL data in trend and projection analysis requiring consistent time intervals (see section 2.1).

The accuracy of the multi-temporal built-up surface prediction of the new release R2023A has improved with respect to all previous releases (R2022, R2019) in all epochs (Figure 8). The unrealistic built-up surface change rate in RURAL domain noticed in the R2022 has been fixed in the new release R2023 (Figure 10).

Table 10 - Summary of the characteristics of the new GHS-BUILT data vs. the previous releases

Data Characteristics	GHS-BUILT R2018A	GHS-BUILT-S R2023A
Definition of the built-up class abstraction	INSPIRE "BUILDING" roofed structure above ground for any use	INSPIRE "BUILDING" roofed structure above ground for any use
Definition of the NRES vs. RES class abstraction	Not available	Derived from INSPIRE "residential" use definition
Built-up surface : class	Boolean	Continuous
Built-up surface : RES vs. NRES class	Not available	Boolean
Built-up surface : spatial resolution	30m	10m
Built-up surface : observed epochs	4 (1975,1990,2000,2014)	5 (1975,1990,2000,2014,2018)
Built-up surface : change map	Pixel-based	Segment-based
Building height : measurement	Not available	Continuous
Building height : spatial resolution	Not available	100m
Building height : observed epoch	Not available	2018
Built-up surface : spatial resolution of the generalized grids	250m 1km	100m 1km
Built-up volume : spatial resolution of the generalized grids	Not available	100m 1km
Equal-time-interval spatial-temporal interpolated grids of built-up surfaces and volume	Not available	12 epochs (1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015, 2020, 2025, 2030)
Future grids projections	Not available	2025 and 2030

2.1.4 Input Data

2.1.4.1 Remotely sensed image data

The remotely sensed image data supporting this GHSL release are collected by the Landsat and the Sentinel platforms, organized in five epochs: 1975, 1990, 2000, 2014, and 2018.

The Landsat data used in input include 35479 individual scenes organized in four epochs 1975, 1990, 2000, and 2014. The average absolute time tolerance of the image data collection time vs. the nominal time barycentre of the epoch is 2.0, 2.2, 1.2, and 0.8 years for the 1975, 1990, 2000, and 2014 epochs, respectively. The aggregated time precision of all the data in the four epochs is of 1.5 years. The empirical time barycentre for the epochs 1975, 1990, 2000, and 2014 is the year 1975.1, 1989.4, 2000.8, and 2013.2, respectively.

Table 11 - Summary of the Landsat Image data used in input

Row Labels	Count of YEAR	Average of YEAR	Average of ABS_TimeTolerance
1975	**7355**	**1975.1**	**2.0**
L1	4403	1973.4	1.6
L2	1880	1976.8	1.8
L3	1072	1979.3	4.3
1990	**8011**	**1989.4**	**2.2**
L4	258	1983.5	6.5
L5	7475	1989.4	2.0
L6	278	1994.3	4.3
2000	**9774**	**2000.8**	**1.2**
L5	459	2004.7	5.7
L6	8827	2000.4	0.7
L7	488	2005.5	5.5
2014	**10339**	**2013.2**	**0.8**
L5	638	2009.6	4.4
L7	259	2009.5	4.5
L8	9442	2013.5	0.5
Grand Total	**35479**	**1996.5**	**1.5**

The epoch 2018 is derived from the GHS_composite_S2_L1C_2017-2018_GLOBE_R2020A[17] that represents a global, cloud-free pixel-based composite created from the Sentinel-2 data archive (Level L1C) available in Google Earth Engine[18] for the period January 2017 - December 2018.

2.1.4.2 High-level semantic abstraction data

Several high-level semantic datasets are used in the process with the function of prior knowledge supporting the various phases of data classification to obtain BU surface fraction estimates, the historical change detection in the BU surfaces, and the land use classification of RES vs. NRES BU surfaces.

BU class abstraction labels: i) the GHS-BUILT-S2 R2020A[19] that is a probability to the BU class spatial raster dataset derived from Sentinel-2 global image composite for reference year 2018 using Convolutional Neural Networks (GHS-S2Net) and ii) the buildings derived from VHR imagery by Microsoft[20] and the settlement delineation from Facebook[21] open efforts

[17] https://ghsl.jrc.ec.europa.eu/download.php?ds=compositeS2

[18] https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2

[19] https://ghsl.jrc.ec.europa.eu/download.php?ds=buS2

[20] https://github.com/microsoft/USBuildingFootprints

[21] https://research.fb.com/downloads/high-resolution-settlement-layer-hrsl/

Multi-temporal assessments: i) the BU change map GHS built-up spatial raster dataset, derived from Landsat, multi-temporal (1975-1990-2000-2014) of the GHSL R2019 (Corbane et al., 2019), ii) GISA the Global Impervious Surface Area (Huang et al., 2022), and WSF_EVO the World Settlement Footprint Evolution product (Marconcini et al., 2021).

Land Use and other: information included in National land use data (US NLUD[22], EU CORINE[23]) and Volunteered geographical information by Open Street Map (OSM)[24] on LANDUSE, ROADS, RIVER, STREAMS

2.1.5 Technical Details

Author: Pesaresi, Martino; Politis Panagiotis

Product name: GHS-BUILT-S_GLOBE_R2023A

Spatial extent: Global

Temporal extent: from 1975 to 2030, 5 years interval

Coordinate Systems: World Mollweide (ESRI:54009), WGS84 (EPSG:4326)

Spatial resolution available: 10 m, 100 m, and 1 km, 3ss, 30ss

Encoding: integer (Byte, UInt16, UInt32), unit: built square meters in the grid cell

Data organisation: GeoTIFF file (10 m, 100m, 1km, 3 ss, 30 ss) with overview images (OVR). Data tiles of 100x100 km size in GeoTIFF format (10 m, 100 m, 1 km, 3 ss, 30 ss). Tile schema in shapefile format

Table 12 outlines the technical characteristics of the datasets released in this data package.

Disclaimer: the re-projection of the World Mollweide version of the GHS-BUILT-S_GLOBE_R2023A to coordinate systems requires specific technical knowledge. No responsibility is taken for workflows developed independently by users.

Table 12 - Technical details of the datasets in GHS-BUILT-S_GLOBE_R2023A

GHS-BUILT-S_GLOBE_R2023A		
ID	**Description**	**Resolution (projection)**
GHS_BUILT_S_E<epoch>_GLOBE_R2023A_<proj>_<res>_V1_0	BU surface <epoch> 1975-2030; <proj> 54009, 4326; <res> 100, 1000, 3ss, 30ss Encoding: UInt16 (100 m), UInt32 (1 km) Values range: 0-10000 (100 m, 3 ss), 0-1000000 (1 km, 30 ss) NoData: 65535 (100 m), empty (3 ss), 4294967295 (1000 m), empty (30 ss)	100 m, 1000 m World Mollweide (ESRI:54009) 3 ss, 30 ss WGS84 (EPSG:4326)
GHS_BUILT_S_E2018_GLOBE_R2023A_54009_10_V1_0	BUFRAC at 10m spatial resolution for E2018 Encoding: Byte Values range: 0-100 NoData: 255	10 m World Mollweide (ESRI:54009)

[22] https://www.mrlc.gov/data

[23] https://land.copernicus.eu/pan-european/corine-land-cover/clc2018

[24] https://www.openstreetmap.org/

GHS-BUILT-S_NRES_GLOBE_R2023A		
ID	**Description**	**Resolution (projection)**
GHS_BUILT_S_NRES_E<epoch>_GLOBE_R2023A_<proj>_<res>_V1_0	Non-residential BU surface <epoch> 1975-2030; <proj> 54009, 4326; <res> 100m, 1000, 3ss, 30ss Encoding: UInt16 (100 m), UInt32 (1 km) Values range: 0-10000 (100 m, 3 ss), 0- 1000000 (1 km, 30 ss) NoData: 65535 (100 m), empty (3 ss), 4294967295 (1000 m), empty (30 ss),	100 m, 1000 m World Mollweide (ESRI:54009) 3 ss, 30 ss WGS84 (EPSG:4326)
GHS_BUILT_S_NRES_E2018_GLOBE_R2023A_54009_10_V1_0	NRES 10m (boolean) at 10m spatial resolution for E2018 Encoding: Byte Values range: 0:non-NRES ; 1:NRES NoData: 255	10 m World Mollweide (ESRI:54009)

2.1.6 Summary statistics

Table 13 - Summary statistics of predicted surface (square meters) of built-up total (BUTOT) and the built-up non-residential (BUNRES) component, per years of prediction

YEAR	BUTOT	BUNRES	NRES SHARE
1975	173,579,688,013	14,851,659,436	8.56%
1980	194,287,591,836	15,786,940,488	8.13%
1985	220,083,865,086	16,824,775,584	7.64%
1990	249,833,100,538	17,952,331,271	7.19%
1995	280,112,957,342	20,040,927,733	7.15%
2000	315,676,641,605	22,572,549,796	7.15%
2005	345,785,691,160	24,500,768,168	7.09%
2010	382,248,138,896	26,910,963,662	7.04%
2015	422,815,687,372	29,654,448,253	7.01%
2020	464,586,276,041	32,175,346,993	6.93%
2025	491,637,335,418	34,010,908,125	6.92%
2030	509,277,184,552	35,391,231,420	6.95%

2.1.7 How to cite

Dataset:

Pesaresi, Martino; Politis, Panagiotis (2023): GHS-BUILT-S R2023A - GHS built-up surface grid, derived from Sentinel2 composite and Landsat, multitemporal (1975-2030). European Commission, Joint Research Centre (JRC) [Dataset] doi: 10.2905/9F06F36F-4B11-47EC-ABB0-4F8B7B1D72EA PID: http://data.europa.eu/89h/9f06f36f-4b11-47ec-abb0-4f8b7b1d72ea

Concept & Methodology:

European Commission, GHSL Data Package 2023, Publications Office of the European Union, Luxembourg, 2023, ISBN 978-92-68-02341-9, doi:10.2760/098587, JRC133256

Essential methodological background:

Pesaresi M, Corban C, Julea A, Florczyk A, Syrris V, Soille P. Assessment of the Added-Value of Sentinel-2 for Detecting Built-up Areas. Remote Sensing 8 (4); 2016. p. 299. JRC99996

Pesaresi M; Syrris V; Julea A. A New Method for Earth Observation Data Analytics Based on Symbolic Machine Learning. Remote Sensing 8 (5); 2016. p. 399. JRC99747

Pesaresi M. Global fine-scale information layers: the need of a paradigm shift. In: Soille P, Marchetti PG, editors. Proceedings of the 2014 conference on Big Data from Space (BiDS`14). Luxembourg (Luxembourg): Publications Office of the European Union; 2014. p. 8-11. JRC92345

Pesaresi M, Ouzounis G, Gueguen L. A new compact representation of morphological profiles: report on first massive VHR image processing at the JRC. In Conference Proceedings: Sylvia S. Shen, Paul E. Lewis, editors. Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII. Vol. 8390. SPIE; 2012. JRC70542

Gueguen L, Soille P, Pesaresi M. A New Built-Up Presence Index Based On Density of Corners. In Conference Proceedings: Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International. Piscataway (USA): IEEE; 2012. p. 5398-5401. JRC68582

Ouzounis G, Pesaresi M, Soille P. Differential area profiles: decomposition properties and efficient computation. IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (8); 2012. p. 1533-1548. JRC59388

Pesaresi M, Gerhardinger A, Kayitakire F. A Robust Built-up Area Presence Index by Anisotropic Rotation-invariant Textural Measure. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 1 (3); 2008. p. 180-192. JRC37845

Pesaresi M, Benediktsson J. A New Approach for the Morphological Segmentation of High-Resolution Satellite Imagery.. IEEE Transactions on Geoscience and RS 39 (2); 2001. JRC19264